BRAIN-INSPIRED AUDIO-VISUAL INFORMATION PROCESSING USING SPIKING NEURAL NETWORKS

ANNE WENDT

A THESIS SUBMITTED TO AUCKLAND UNIVERSITY OF TECHNOLOGY IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

2021

SCHOOL OF ENGINEERING, COMPUTER AND MATHEMATICAL SCIENCES

To everyone who believed in me

ABSTRACT

Artificial neural networks are one of the most popular and promising approaches to modern machine learning applications. They are based on a mathematical abstraction of the intricate processing mechanisms in the human brain, remaining sufficiently simple for efficient processing in conventional computers. Despite efforts to mimic the capabilities of the brain, however, they are limited in their contextual understanding of concepts and behaviours. With the aim to explore ways to overcome these limitations, this thesis endeavours to investigate alternatives that are closer to the original biological systems, with a focus on processing auditory and visual signals. Inspired by the functioning of human hearing and vision and by the brain's capabilities to dynamically integrate newly perceived information with previous experiences and knowledge, this thesis presents the hypothesis that mimicking these processes more closely could lead to an enhanced analysis of such signals.

The framework that was developed to investigate this hypothesis consisted of three separate but connected projects that looked into biologically inspired computational processing of auditory, visual, and combined audio-visual signals, respectively. One aim of designing the framework was to largely preserve the spectral, spatial, and temporal characteristics of the original signals through tonotopic and retinotopic mapping. For the auditory processing system, an encoding and mapping method was developed that could transform sound signals into electrical impulses ("spikes") by simulating the human cochlea, which were then fed into a brain-shaped three-dimensional spiking neural network at the location of the auditory cortices. For the visual system, the method was developed analogously, simulating the human retina and feeding the resulting spikes into the location of the visual cortex. A key advantage of this approach was that it facilitated a straightforward brain-like combination of input signals for the analysis of audio-visual stimuli during the third project.

The approach was tested on two existing benchmark datasets and on one newly created New Zealand Sign Language dataset to explore its capabilities. While the sound processing system achieved good classification results on the chosen speech recognition dataset (91%) compared to existing methods in the same domain, the video processing system, which was tested on a gesture recognition dataset, did not perform as well (51%). The classification results for the combined audio-visual processing model were between those for the individual models (76.7%), and unique spike patterns for the five classes could be observed.

Even though the models created in this work did not exceed the statistical achievements of conventional machine learning methods, they demonstrated that systems inspired by biological and neural mechanisms are a promising pathway to investigate audio-visual data in computational systems. Increasing the biological plausibility of the models is expected to lead to better performance and could form a pathway to a more intuitive understanding of such data. To broaden the applicability of the model, it is suggested that future work include the addition of other sensory modalities or signals acquired through different brain recording and imaging methods and to perform further theoretical and statistical analysis of the relationship between model parameters and classification performance.

TABLE OF CONTENTS

LIST OF FIGURES[1]

[1] Where figures displayed in this thesis have not been explicitly attributed to a source, they were created by the thesis author herself. Copies of permission letters for any reused figures can be found in Appendix C.

LIST OF TABLES

ATTESTATION OF AUTHORSHIP

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Anne Wendt

CO-AUTHORED WORKS

Apart from several research presentations at events within and outside of Auckland University of Technology (for example, at the AUT Research Project Fairs in 2017 and 2018, at the 3-Minute Thesis competition as a participant in 2017 and finalist in 2018, and at the Auckland AI and Machine Learning Meetup in their October 2018 event), two peer-reviewed publications were submitted during the candidate's PhD studies, as cited later where appropriate.

Paulun, Wendt, and Kasabov (2018)

Journal paper published in Frontiers of Computational Neuroscience.

Title: A Retinotopic Spiking Neural Network System for Accurate Recognition of Moving Objects Using NeuCube and Dynamic Vision Sensors.

Authors: Lukas Paulun, Anne Wendt, Nikola Kasabov.

Year of publication: 2018.

Abstract: This paper introduces a new system for dynamic visual recognition that combines bio-inspired hardware with a brain-like spiking neural network. The system is designed to take data from a dynamic vision sensor (DVS) that simulates the functioning of the human retina by producing an address event output (spike trains) based on the movement of objects. The system then convolutes the spike trains and feeds them into a brain-like spiking neural network, called NeuCube, which is organized in a three-dimensional manner, representing the organization of the primary visual cortex. Spatio-temporal patterns of the data are learned during a deep unsupervised learning stage, using spike-timing-dependent plasticity. In the second stage, supervised learning is performed to train the network for classification tasks. The convolution algorithm and the mapping into the network mimic the function of retinal ganglion cells and the retinotopic organization of the visual cortex. The NeuCube architecture can be used to visualize the deep connectivity inside the network before, during, and after training and thereby allows for a better understanding of the learning processes. The method was tested on the benchmark MNIST-DVS dataset and achieved a classification accuracy of 92.90%. The paper discusses advantages and limitations of the new method and concludes that it is worth exploring further on different datasets, aiming for advances in dynamic computer vision and multimodal systems that integrate visual, aural, tactile, and other kinds of information in a biologically plausible way.

Abbott, Sengupta, and Kasabov (2016)

Conference paper presented at the International Joint Conference on Neural Networks (IJCNN) in Vancouver, Canada.

Title: Which method to use for optimal structure and function representation of large spiking neural networks: A case study on the NeuCube architecture.

Authors: Anne Abbott (nee Wendt), Neelava Sengupta, Nikola Kasabov.

Year of publication: 2016.

Abstract: This study analyses different representations of large spiking neural network (SNN) structures for conventional computers and uses the NeuCube SNN architecture as a case study. The representation includes neuronal connectivity and network's and neurons' states during the learning process. Three different structure types, namely adjacency matrix, adjacency list, and edge-weight table, were compared in terms of their storage needs and execution time performance of a learning algorithm, for varying numbers of neurons in the network. Comparative analysis shows that the adjacency list, combined with a backwards indexing mechanism, scales up most efficiently both in terms of performance and of storage requirements. The optimal algorithm was further used to simulate a large scale NeuCube system with 241,606 spiking neurons in a 3D space for prediction and analysis of benchmark spatio-temporal data.

Comment: This paper received a WCCI 2016 Outstanding Paper Travel Grant. The first and second author contributed equal parts.

ACKNOWLEDGEMENTS

The completion of this thesis has probably taken longer than anyone, including me, expected. I would like to use this opportunity to say thank you to everyone who was part of my journey and especially to those who stuck around until the very end. You are amazing people and I am forever grateful to have met you.

Associate Professor Jacqueline Whalley – for being wonderfully supportive, encouraging, understanding, and hands-on problem-solving. Your feedback on my work and my writing constantly lifted the quality of my thesis. Your moral support and advice got me to where I am today. Thank you for sharing your strength with me.

Dr Krassie Petrova – for your encouragement, your compassion, and for stepping in when I needed you. Your insight and guidance shaped my work invaluably. Thank you for walking with me to the finish line again.

Reverend Linda Teleo Hope – my mentor and shepherd who found me when I was lost and guided me back to the right path. You showed me that I can climb the mountain even if it looks too high to conquer. Thank you for liberating my spirit.

Ms Joyce D'Mello – for your kindness, compassion, and motherly advice. You have instilled your beautiful spirit into our working environment, and into me, you have instilled my will to fight.

Without the four of you, there would be no thesis.

I would also like to thank AUT's School of Computer and Mathematical Sciences for awarding me the school's PhD scholarship. Lessening the financial burden was immensely helpful to focus on my work.

I am grateful to have received academic support from the following people (in order of appearance):

- Dr Grace Wang for helping me with the EEG data collection for Chapter 2;
- Dr Neelava Sengupta for the very enjoyable collaboration on the WCCI paper and his friendship;

- Dr Josafath Israel Espinosa Ramos for helping me understand the NeuCube architecture and for giving me access to his JNeuCube implementation, for his help with interpreting results and understanding reviewers' expectations, and for his delicious cooking and supportive friendship;
- Giovanni Saraceno for his initial prototyping of the auditory system;
- Dr Mithila Durai for her help and advice on the theory of audiology;
- Lukas Paulun for his initial design of the retinal encoding, the fruitful collaboration on the initial visual system, and for his friendship and support;
- Susie Ovens and Pam Croskery for introducing me to the world of the Deaf community and New Zealand Sign Language; and
- Dr Stefan Marks for his help with visualising the tonotopic mapping and for his guidance.

I am grateful to have received moral support from the following people (in order of appearance):

- The Northcote Glenfield hockey team for giving me a good start into my week;
- The many colleagues and friends I met at AUT – Bálint, Brody, Dhvani, Helena, Kaushalya, Mahima, Nathan, Vinita, Vivienne, and Yoshiki – for their camaraderie and helpful discussions;
- Chloe for being a wonderful friend, becoming my second family and giving me a home away from home;
- Dr Rosser Johnson and Associate Professor Dave Parry for demonstrating AUT's values of tika, pono, and aroha in everything they do;
- Robert for listening, motivating, and providing additional computing power.

Additional organisations and people that helped with completing this thesis were the AUT library and ICT services teams, the archive.org community, Alexandra Elbakyan, TeamViewer, the Dilmah and Greenfield tea companies, Bombay Chinese takeaways, and Megumi and her team at the AUT Refuel café.

I would further like to thank Professor Nikola Kasabov for teaching me resilience and valuable life lessons about the principles of research integrity and ethical behaviour, and for giving me the opportunity to become stronger than I ever thought I could be.

And last but certainly not least, I would like to express my sincerest gratitude to my family: Mutti, Papa, and Juliane – for believing in my ability to undertake and finish this work and for supporting me all the way through good times and bad decisions. I could write another thesis about my appreciation for your presence in my life.

For the Lord is good; his steadfast love endures forever, and his faithfulness to all generations. *Psalm 100:5*

1 INTRODUCTION

"I consider that a man's brain originally is like a little empty attic, and you have to stock it with such furniture as you choose."

– Sherlock Holmes in A Study in Scarlet by Sir Arthur Conan Doyle

1.1 BACKGROUND AND SCOPE

This thesis presents a computational model that attempted to process audio-visual information in a brain-inspired way. It structures and verbalises several years of research building and exploring the capabilities of a novel computational architecture that was based on the functioning of the human brain. The term **brain-inspired** in this context means that certain aspects of how the brain processes information were simulated in a computational model. This includes the translation of multimodal input stimuli into electrical signals and the principles of signal transmission, as well as the spatial arrangement and layout of the model, which facilitates anatomically plausible signal mapping. While this computational model is inspired by the functioning of the human brain, it is by no means trying to copy neurological processes entirely and should not be interpreted as such. Instead, it showcases and explores the capabilities of brain-inspired computational methods when applied to real-world problems.

The research described in this thesis is based on the premise that the human brain is the most sophisticated "computer" in existence. The brain effortlessly consolidates a multitude of environmental stimuli, identifies and systematises central meaningful elements, and extracts knowledge to make decisions in real-time. Its shape and structure evolved over millions of years before it became the highly functional control organ that it is now. Containing over 86 billion neurons (Azevedo et al., 2009), its computing power is immense, and it has anecdotally been described as the only organ that studies itself.[2]

The brain is exceptionally good at interpreting ambiguous behaviour and situations, making sense out of its unpredictable surroundings, and deriving a sufficiently appropriate response. Computers, on the other hand, excel at analysing numerical data, precisely storing and recalling information, and performing complex calculations. Both systems have been created and grown to perform different groups of tasks, and, hence, take a different approach to solve problems. While computers focus on precision, i.e., preserving all details of information exactly as they were presented, brains take a more pragmatic approach that focuses on understanding the meaning of information and its implications as a whole.

Despite their differing functional paradigms, there have always been efforts to make computers solve human problems by creating some kind of **Artificial Intelligence** (AI), most famously first described by Marvin Minsky in 1961. Minsky formulated five problems that computers needed to be able to solve if they were to be considered intelligent: search, pattern recognition, learning, planning, and induction. Interestingly, he also noted that "in the long run, we must be prepared to discover profitable lines of heuristic programming which do not deliberately imitate human characteristics" (Minsky, 1961, p. 26). He argued that biologically inspired approaches such as the perceptron (Rosenblatt, 1958) were limited in their abilities and would thus not be required for the development of AI.

Now, more than half a century later, tremendous improvements in transaction speed and computing power, as well as the development of new biologically inspired algorithms to solve computational problems, have seen the idea of using **Neural Networks** revived. Multi-layered Artificial Neural Networks and Deep Learning architectures (LeCun, Bengio, & Hinton, 2015; Schmidhuber, 2015), that are based on how information is processed in the brain through neurons communicating with each other, have gained immense popularity due to their outstanding results in pattern recognition and machine learning competitions within the last decade (Schmidhuber, 2015). Deep learning approaches typically contain a multitude

[2] This statement is, of course, quite exaggerated, taking into consideration that all organs in the human body have to study themselves to monitor their state and try to uphold a certain level of functionality. The remarkable thing about the brain is, however, that it does so not just for operational purposes, but for enjoyment.

of interconnected layers of neurons arranged in a sequential layout (LeCun et al., 2015). In contrast to the von-Neumann architecture prevalent in conventional computers, procedures and data are not held separately but instead stored together in the connections between the neurons in the network. When signals are propagated through the network, the weights of the connections are modified based on certain learning rules which create distinctive, recognisable patterns that can be used to classify data or predict unknown relationships. Every change in the input signal causes specific changes in the network. An artificial neural network can, thus, be seen as a specialised way of data representation and computation.

Despite their slow start caused by the shortcomings of early hardware, neural networks and deep learning are nowadays recognised as powerful tools that can be applied to almost every problem for which data exist. A few examples where the use of such models have attracted attention outside the AI community are:

- large-scale image recognition such as the ImageNet dataset with over 14 million images (J. Deng et al., 2009), for which a Convolutional Neural Network achieved considerably better results than conventional methods (Krizhevsky, Sutskever, & Hinton, 2012);
- recognising spoken words from a large corpus of sentences called TIMIT (Garofolo et al., 1993), where a Deep Recurrent Neural Network achieved the lowest error rate so far (Graves, Mohamed, & Hinton, 2013);
- structuring vast amounts of general trivia knowledge to create a sophisticated question answering system called IBM Watson (Ferrucci, 2010); and
- mastering complex board and strategy games such as Go and StarCraft II using deep reinforcement learning and repeatedly beating world-class players in real-life tournaments (Čertický, Churchill, Kim, Čertický, & Kelly, 2019; Silver et al., 2016; Silver et al., 2017).

All these case studies are specialised applications that show that neural networks can perform exceedingly well in their respective domain. However, they might not work equally well when applied to other "normal" human tasks, such as holding a coherent conversation. The algorithms could be too specialised to perform well on other or more general problems.

This phenomenon was first described as the **No Free Lunch Theorem** by Wolpert and Macready (1997). Although the term was initially coined for optimisation problems, it applies to the majority of AI systems as well. The theorem states that increased performance on one type of problem is always counteracted by decreased performance on all other types of problems. This also implies that a generalised approach would be just as good as a specialised one because on average, all algorithms perform equally well according to the theorem.

The research presented in this thesis tries to advance existing efforts towards generalisable, brain-inspired AI. It employs a brain-inspired type of artificial neural networks called **Spiking Neural Networks** (SNN) and explores its applicability to audio-visual data. Since the brain comprehends such multimodal data with ease, the idea of mimicking some of the involved neurological processes and combining them into a computational model warrants investigation. This research is exploratory and focuses on the investigation of these ideas in breadth rather than in depth. It includes case studies on auditory data, visual data, and combined audio-visual data to assess the capabilities of the brain-inspired SNN framework. The thesis tracks the pathway from the idea to the final model.

The remainder of this chapter explains the objectives of the research, derives the research questions, outlines the structure of the thesis, and discusses the main contributions of this work.

1.2 MOTIVATION AND OBJECTIVES

Initially, this research was intended to investigate language processing in the brain, driven by a keen personal interest in foreign languages and observing the variations to which degree semantic concepts can or cannot be expressed in them. While phonetic components, grammar, sentence structure, and vocabulary of two languages can have no relation at all, their speakers are still able to express similar concepts. For example, the English phrase "My name is …" translates to "Ko … tōku ingoa" in Te Reo Māori, but neither the English verb *to be* nor the Māori particle *ko* has an equivalent in the other language. On the other end of the spectrum, languages that are closely related linguistically may not be able to convey the same meaning and might need to borrow words from each other, such as the German word Doppelgänger used in English to describe someone looking almost identical to a stranger. All languages solve the problem of communication, but their large number shows that it can be solved in many different ways.

A recent study by Y. Yang, Wang, Bailer, Cherkassky, and Just (2016) suggests that concepts communicated in different languages by the same individual are stored in and retrieved from the same regions of the brain. In another study, it was found that distinct brain areas respond to certain high-level concepts such as *relationship* or *shelter* when presented with words from these categories (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016). The authors concluded that it can be assumed that there is an underlying conceptual framework in the brain across individuals that allows us to connect, memorise, and communicate our experiences (Huth et al., 2016). In order to create a general-purpose AI as described in the previous section, this mental framework would have to be analysed and encoded in a suitable computer architecture that is capable of structuring and connecting semantic concepts.

After an initial literature review, it soon became apparent that creating a brain-inspired computational model of multi-lingual comparative semantics was rather ambitious for the timeframe of a doctoral dissertation. Neither have the biological processes of how the brain comprehends and generates language been fully understood by neuroscientists, nor had research in the area of Natural Language Processing yielded any fruitful outcomes that were able to deal with pragmatics (Cambria & White, 2014). It was, therefore, found to be a more realistic approach to start with the "mechanics" of communication, like perceiving and processing audio-visual information, using a brain-inspired system that would subsequently be able to integrate more biological features of the involved processes such as new findings on semantics and memory as they become available.

The **primary objective** of this study was to create a computational model that could process audio-visual information in a brain-inspired way. The notion of being brain-inspired or brain-like implied a **secondary objective**, namely, using a computational architecture that facilitates future integration of new findings about the workings of the brain to extend the model's functionality. This required an understanding of neuro-biological concepts of information processing in the brain, as well as knowledge about similar previous computational approaches and their advantages and limitations. Applying the model to two different input modalities (auditory and visual) and finding a method to combine them to gain insights about the data emerged as a central component of the research, and the natural way of integrating these modalities would be to use a brain template to define the structure and shape of the network.

Subsequently, another **secondary objective** was then to explore the capabilities of this model and its applicability to real-world scenarios and problems. Consistent with the focus on auditory and visual modalities, it was established that there were three components required to investigate the applicability of the model in breadth rather than in depth: The model should be able to process sound data, video data, and combined audio-visual data, and preferably these data should be dynamic, i.e., the signals change over time, to better simulate the kind of data that is handled by the brain. Sound waves are dynamic in their very nature of travelling through the air; however, for the visual part, it was determined that video data should be used instead of static images. For the analysis of combined audio-visual data, using language data was preferred, maintaining the original intent of the study.

The final **secondary objective** was related to the implementation of the model and what characteristics it should possess. To support its applicability to different case studies and facilitate the exploratory nature of the research, the model was expected to be able to handle multiple standard audio and video formats such as WAV, MP3, AVI, and MP4 for the input data. From a computational point of view, possible future deployment on different hardware systems was considered in the development of the software architecture. For example, specialised biologically-inspired hardware platforms were available or under active development by other research groups that were envisioned to be potentially used later to transform the sound and video signals. The transformed signals could then be fed into an implementation of the neural network that was running on either a standard von-Neumann computer or on neuromorphic hardware. The transformation methods developed here were expected to interface easily with the existing and anticipated hardware. As a final consideration, the model was required to produce an interpretable form of output – contrary to the majority of existing deep learning algorithms, it should not be a black box.

Revisiting the scope as explained in the previous section, four items were **not considered objectives** of this study. They are briefly mentioned here to clarify expectations. Firstly, this study did not aim to discover new methods to understand the functionality of the brain. Instead, it shows a new method of analysing brain-related data in a brain-like way. Secondly, and related to that, it did not want to generate new insights into the areas of neuroscience or biology but rather utilise knowledge from those areas for the creation of the model. Thirdly, it was not expected to be able to form semantic concepts based on the audio-visual input data like the brain of an infant would do. This was due to the variety of other senses such as olfactory or haptic perception that were not considered here but are necessary to develop a full understanding of one's environment. Adding these senses could form the basis for future investigation though, as suggested in Section 9.4. Lastly, the goal of the case studies was to illustrate the method, not to replace existing techniques that were based on mathematical models and specialised in a particular application area. Compared with those methods, it was expected that the more generalised approach presented here would achieve a lower accuracy on typical classification benchmarking datasets. However, as explained above, more general insights into the capabilities of the model were expected.

1.3 RESEARCH QUESTIONS

Based on the objectives of this study as described in the previous section, one central research question had been developed along with three sub-questions that allowed a systematic solution approach. The overarching question for this research was:

How can a computational model of audio-visual information processing be created that uses brain-inspired mechanisms to analyse those data, and what can be learned from such a model?

Several sub-questions were derived that related to different aspects of the research question. These questions are revisited in Section 9.2 of the thesis.

(1) Biological inspiration
  a. How can the biological background of audio-visual information perception and processing inform the design of an audio-visual computational model?
  b. Can neurological pathways of audio-visual information that are observed in the brain also emerge in a brain-inspired computational model?
  c. What aspects of the human audio-visual processing system can enhance the analysis of audio and video data?

(2) Design of an audio-visual spiking neural network
  a. How can audio-visual data be transformed (encoded) into electrical impulses for use in a spiking neural network?
  b. What is a biologically plausible way to input (to map) sound and visual stimuli into the model?
  c. How can the use of both auditory and visual data in one combined model be facilitated in a biologically plausible, yet computationally feasible way?

(3) System evaluation
  a. How does the brain-inspired model perform on sound and video benchmark datasets compared to conventional approaches?
  b. What are the advantages and disadvantages of using biologically plausible encoding and mapping approaches for processing audio-visual data?
  c. Does the size of the neural network influence the learning processes and performance of the model?

1.4 STRUCTURE OF THE THESIS

The very first chapter in this thesis following the introduction describes a pilot study that was conducted at the beginning of the candidate's research journey. In this study, the candidate recorded her brain activity while viewing images of objects and listening to the objects' names. She then built a model that tried to classify the data into object categories and to observe semantic concept formation, which was the original motivation and objective for the research (see Section 1.2). While not very fruitful in terms of research output, this study did provide valuable insights into the limitations of the model and largely informed and directed the work that followed. **Chapter 2** explains the setup and outcome of this study in detail.

After the learnings from the pilot study led to an adjustment in the research direction, the work carried out for this thesis was organised into three projects that looked at processing auditory, visual, and audio-visual signals, respectively. Both the auditory and the visual systems consist of separate models including encoding and mapping methods, while for the audio-visual system, their functionalities were combined into one model. All three systems were then tested on domain-specific datasets and the results were compared to existing methods found in the literature. Following the established procedure of presenting research in Computer Science, this thesis contains chapters on literature background (Chapters 3 and 4), method and system design (Chapter 5), case studies, experiments, results, and discussion (Chapters 6, 7, and 8), and a conclusion (Chapter 9). The three projects are thematically organised across these chapters as shown in Figure 1-1.

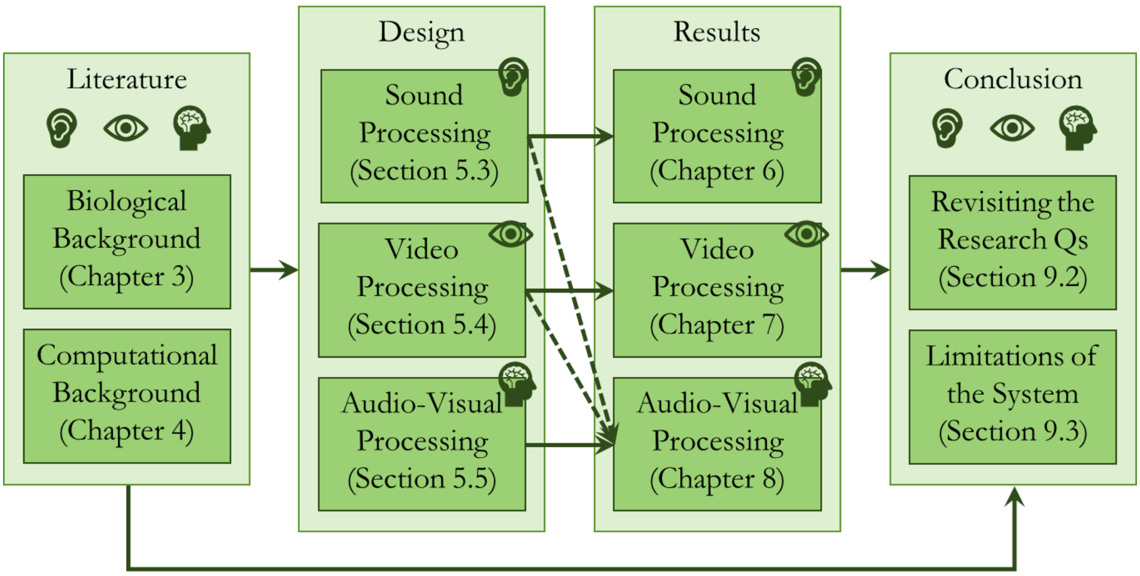


FIGURE 1-1: THEMATIC ORGANISATION OF THE THREE PROJECTS (AUDITORY, VISUAL, AUDIO-VISUAL) PRESENTED IN THIS THESIS.

Chapter 3 begins by exploring the biological background of human hearing and vision, and how multimodal information is processed in the brain. It describes what sound is, how the ear and more specifically the cochlea transform sound waves into electrical signals, and how these signals travel through the auditory pathway into the auditory cortices to be processed further. It analogously explains the mechanics of vision, how the eye and more specifically the retina and the ganglion cells transform and converge photons into electrical signals, and how these signals travel through the visual pathway into the visual cortex to be processed further. The chapter closes by summarising what is known about the brain's capability to combine these two modalities and make sense of the received information.

Chapter 4 gives an overview of existing computational methods for auditory, visual, and audio-visual information processing. It first outlines established machine learning methods for sound and image analysis, followed by a summary of brain-inspired methods that have been applied to audio-visual information processing tasks. The focus of this chapter is on algorithms that claim biological inspiration or plausibility.

Chapter 5 describes the computational approach that was created to answer the majority of the formulated research questions. Following a general overview of the system architecture, three separate paradigms of the framework are introduced that relate to sound processing, video processing, and combining modalities, respectively. The section on sound processing describes the cochlear encoding and tonotopic mapping approaches with a focus on a newly developed data compression algorithm. The section on video processing describes the retinal encoding and retinotopic mapping approaches with a focus on a novel mechanism for realising colour recognition and modelling receptive fields. The final section of the chapter describes how both sound and video data were integrated into one model, how the brain-shaped network enabled this integration, and how the signal times were synchronised.

The following three chapters then report on the experimental results that were achieved with these three software frameworks.

Chapter 6 presents a case study conducted with the sound processing system using a benchmark dataset in the domain of spoken digit recognition. The chapter highlights the peculiarities of the data, explains the experimental setup and parameters, and states both the qualitative and the quantitative results. It also discusses how the sound processing framework compared to other models using the described benchmark dataset.

Chapter 7 presents a case study in which the video processing system was applied to a benchmark dataset for classifying a set of gestures from videos. This study followed the same outline as the sound processing study, describing the dataset, explaining the experimental setup, reporting on the achieved results, and comparing it to previous work in the same domain.

Chapter 8 presents the final case study, an application for the integrated audio-visual processing system that is based on New Zealand Sign Language. In this study, the model was used to learn the characteristics of five signs and their spoken English equivalents. Like the previous two chapters, this chapter includes a dataset description, information about the experimental setup, as well as results and conclusions about the knowledge gained from the experiment.

Chapter 9 concludes the thesis by summarising its content and revisiting the research questions. It reflects on limitations that arose from model simplifications and generalisations. In the end, the chapter draws attention to possible future directions such as integrating brain data or other senses like olfactory and haptic perception.

In the appendices, this thesis also includes a detailed glossary and list of abbreviations, as well as exemplary code snippets, more detailed results for the experiments, and copyright licenses for all third-party figures.

1.5 MAIN CONTRIBUTIONS OF THIS RESEARCH

The work presented in this thesis introduces a novel method of audio-visual data processing that was inspired by the functioning of the human auditory and visual systems. As such, it is the first of its kind that combines biologically inspired data transformation (encoding) with tonotopic and retinotopic signal mapping into a brain-shaped neural network. The main contributions of this work are:

- The conceptual design of the model that offers a new approach for data analysis with a wide range of application areas;
- A retina-inspired encoding algorithm for visual data that includes colour vision capabilities and a moving focal area;
- A network layout that facilitates straightforward, biologically plausible bimodal signal integration with the option to add more modalities in the future, in 16 different sizes;
- Tonotopic and retinotopic signal mapping algorithms that can determine the optimum number and location of input neurons in the network;
- Compression algorithms for auditory and visual data; and
- Results of the initial exploration of the model's capabilities and limitations.

2 DECODING OBJECTS FROM BRAIN DATA – A PILOT STUDY

"It is a capital mistake to theorise before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts."

– Sherlock Holmes in A Scandal in Bohemia

2.1 BACKGROUND AND MOTIVATION

This chapter describes a preliminary study in which the author of this thesis attempted to classify electroencephalography (EEG) data recorded while a subject was presented with images and spoken and written representations of ten items. It was a first **proof-of-concept study** to gauge the possibilities and limitations of an originally proposed framework[3] related to decoding the underlying meaning of perceived images and words across multiple languages. The results did not provide support for the study's hypothesis that objects can be identified by analysing EEG data using a spiking neural network architecture. However, these findings informed the further design of the research presented in this thesis and provided an

[3] This framework is described in detail in this chapter, in Section 2.2.

important basis for all of the work included in the subsequent chapters. Therefore, a summary of this study is included here as an introductory chapter to serve as additional context and give an explanation of the inspiration for and development of the work presented in this thesis.

The **aim of this project** was to investigate how language and concept formation, in particular, are processed in the brain, leading to the question as to whether concept formation can be detected by analysing brain data. Set in the area of **Natural Language Processing** (NLP), which deals with the computational analysis and generation of human languages, the project was motivated by a review paper that described the three different stages through which NLP research is currently evolving (Cambria & White, 2014). The authors of this review paper argue that while research into syntactic and also increasingly into semantic processing of words and word sequences by machines has made progress in finding feasible explanations of the underlying workings of language understanding, the pragmatics or contextual embedding of language particles is still poorly understood.

Large, annotated word corpora such as the Penn Treebank (Marcus, Marcinkiewicz, & Santorini, 1993) have advanced through and with the **syntactic analysis** of language by providing ways to systematise the grammatical function of sentence components. The outcomes from this area form the foundation for **semantic analysis**, which looks at the meaning of words by putting them in relationship with each other, normally using a graph data structure. One of the earliest approaches to forming such a "relationship graph" between words is the WordNet database (Miller & Fellbaum, 2007), which at the time of writing contained more than 150,000 words and their semantic associations (Princeton University, n.d.). Cambria and White (2014) concluded that this development will eventually lead to a **pragmatic, contextual understanding** of written and spoken content that also takes factors like sensory, emotional and social knowledge into consideration and hence will be able to facilitate "natural" communication between humans and machines.

In relation to the conclusions drawn by Cambria and White (2014), the research project presented in this chapter is positioned between the semantic and the pragmatic aspects of language analysis. By studying brain data to uncover the "semantic footprint" of a small set of words, the hope was to automatically incorporate pragmatic knowledge from the brain's existing associations and acquired knowledge that would aid the classification and categorisation of objects. The hypothesis of this study was inspired by the, at that time recently published, first semantic brain atlas based on functional brain data (Huth et al., 2016). This atlas let users interactively explore a three-dimensional model of the human brain and the semantic concepts most likely associated with specific brain regions.

Based on the findings by Huth et al., this study **hypothesised** that the meaning of objects can be retrieved from EEG data recorded while these objects were presented to a study subject. The following sections describe the data analysis framework and the experimental setup that were used to test this hypothesis, followed by the unfortunately rather dissatisfying results. This is followed by a discussion of why the outcomes were not as good as initially thought and what other research groups have done differently to achieve better results. The conclusion summarises the main findings of this work and discusses how it directed the research that encompasses this thesis.

2.2 DATA ANALYSIS FRAMEWORK

The data analysis framework for this project was based on that of previous studies by KEDRI, the research group of which the author of this thesis was part during her PhD. It follows a novel technique for analysing EEG data reported by Kasabov and Capecci (2015) and by Doborjeh, Wang, Kasabov, Kydd, and Russell (2016). In this technique, EEG data are entered into a brain-shaped three-dimensional spiking neural network architecture called NeuCube (Kasabov, 2014). The EEG signals are fed into those neurons in the network that are spatially closest to their recording electrodes' respective location on a human head. This preserves the spatio-temporal information in the data and enables an easily understandable visualisation of activity clusters. The NeuCube then learns the patterns in the data using the unsupervised spike-timing-dependent plasticity algorithm, which modifies the connections between the neurons as the signals are propagated through the network. After this training stage, the neural activity that is observed in the network when being presented with single samples is quantified using the supervised dynamic evolving SNN algorithm. Finally, these quantified "summaries" of the network activity are labelled using the k-nearest neighbour algorithm, so that samples that exhibit similar activity in the network are grouped together. A more detailed description of the NeuCube architecture can be found in Section 5.2 of this thesis.

The EEG data were collected from the thesis author's brain with the help of a senior lecturer from the university's psychology department. Since this was a proof-of-concept study, one subject was considered sufficient to obtain a first insight into possible outcomes. The support of a senior staff member with a background in psychology who is familiar with the process of collecting EEG data was invaluable to the thesis author. Firstly, her support in designing the experiment minimised issues with the method of stimuli presentation. Secondly, during data collection, the expert ensured that the signals would be as free of interferences as possible by providing a sound-proof recording room that was free of electrical noise and by making sure the impedance levels of the electrodes were kept at a minimum level.

For the presentation of stimuli to the subject, a software tool called Presentation (NeuroBehavioralSystems, n.d.) was used that can be connected to the data collection software and also to the MATLAB package that was used for pre-processing the data, EEGLAB (Delorme & Makeig, 2004). This meant that the start and finishing points of stimuli could be timed precisely and later be linked back together with the corresponding sections of the EEG signals to create epochs.

As **stimuli**, ten everyday objects and common living beings were presented to the study subject: ball, book, child, clock, dog, horse, house, shoe, table, and tree. In the five modes of the data collection process, these items were presented first as coloured images, then as spoken words in English and German, and finally as written words in English and German. In each mode, each of the ten items was presented ten times to the subject in a randomised order. For the mode with images, this included ten different images for each item, while for the spoken and written words, the same recording or text display of a word was used repeatedly throughout the respective segments. Since both the English and the German spoken words were downloaded from an online dictionary to provide a standardised pronunciation, there were two speakers, a male and a female, between whom each set of words was divided. The written text was shown in the Arial font with a font size of 48 in white colour on a black background.

Between the **visual stimuli**, i.e., images and written words, a black screen was shown to create a clear segmentation between the items and, thus, also between the brain signals that were measured for them. The visual stimulus itself was shown for 300 milliseconds, followed by one second of the black screen. This was similar to the setup of a previous study that reported promising results (Simanova, van Gerven, Oostenveld, & Hagoort, 2010). The duration of 300 milliseconds was also chosen because findings from neuroscience literature suggest that the so-called P300 signal plays an important role in semantic categorisation (Azizian, Freitas, Watson, & Squires, 2006; Houlihan, Campbell, & Stelmack, 1994; Mecklinger & Ullsperger, 1993). The P300 component in the signals is evoked about 300 milliseconds after stimulus presentation. The thesis author's intention behind stopping the stimulus presentation after exactly that time was that noise from subsequent thoughts associated with the presented item would be minimised.

For the **spoken words**, a white crosshair was presented during the recording that served as a visual anchoring point for the participant and, hence, helped with minimising the influence of artefacts created by eye movement. The sound files were cropped to 0.5 seconds in length and played over stereo headphones at a sampling rate of 44,100 Hertz. In contrast to the presentation time of visual stimuli for 300 milliseconds, a length of 500 milliseconds was chosen for the sound files because shortening them further would have made the endings of some words unintelligible. Between the sound stimuli, there was one second of silence to provide a clear break between signals, similar to the visual stimulus presentation.

A cap with 64 gel-based electrodes was used to collect the **EEG data**. The baseline reference channel was the electrode at position Cz and the sampling rate was 1,000 Hertz. After re-referencing the dataset, the signals were split into epochs corresponding to the stimuli presentation times. They were then entered into 64 hand-selected input neurons of the neural network, which contained 1,471 neurons in total. The locations of the input neurons were chosen based on the locations of the EEG electrodes on the scalp. The parameters of the network were then adjusted based on the goals of the performed experiments as described in Section 2.3.

2.3 EXPERIMENTS AND RESULTS

Several experiments were carried out with the pre-processed dataset, beginning with an analysis of the signals collected while the subject was presented with the **images**. For the **first experiment**, the 100 data samples were split into two groups of 50 randomly selected samples each, with one group being used as the training dataset and the other group being used for testing. The training samples were passed into the network model to modify the weights of the connections between neurons in the network and thus make the model "learn" the patterns that are specific to a particular category of items. The samples from the testing dataset were then classified into those categories using the trained model.

All results presented in this section were based on the default **system parameters** that at the time were deemed most appropriate by the developers of the NeuCube software.[4] For the encoding of EEG data into spikes, a threshold representation algorithm with a threshold of 0.5 was used; for the initialisation of the network connections, the small-world radius was set to 25; the spike-timing-dependent plasticity learning was performed with a learning rate of 0.01, a potential leak rate of 0.002, a firing threshold of 0.5, and a refractory time of 6; and for the dynamic evolving spiking neural network classifier, the parameters were set to a modulation factor of 0.8, a drift of 0.005, a K of 3 and a sigma of 1.

Overall, only six of the testing samples were classified correctly (i.e., 12% classification accuracy), with half of the classes not having any correct matches at all. Figure 2-1 shows the detailed classification results. The red triangles indicate the expected class of a sample and the blue asterisks represent the class that was assigned to the sample by the model. Apart from three houses being misclassified as trees, there is no clear crossover between classes; rather, the classification seems somewhat random. Noticeably, only very few samples have been misclassified as shoe, child, or horse (only two samples each), while comparatively many have been misclassified as tree (nine samples), book (eight samples), or clock (six samples).

[4] A detailed explanation of the meaning of these parameters as well as the NeuCube architecture and its components is included in this thesis in Section 5.2.

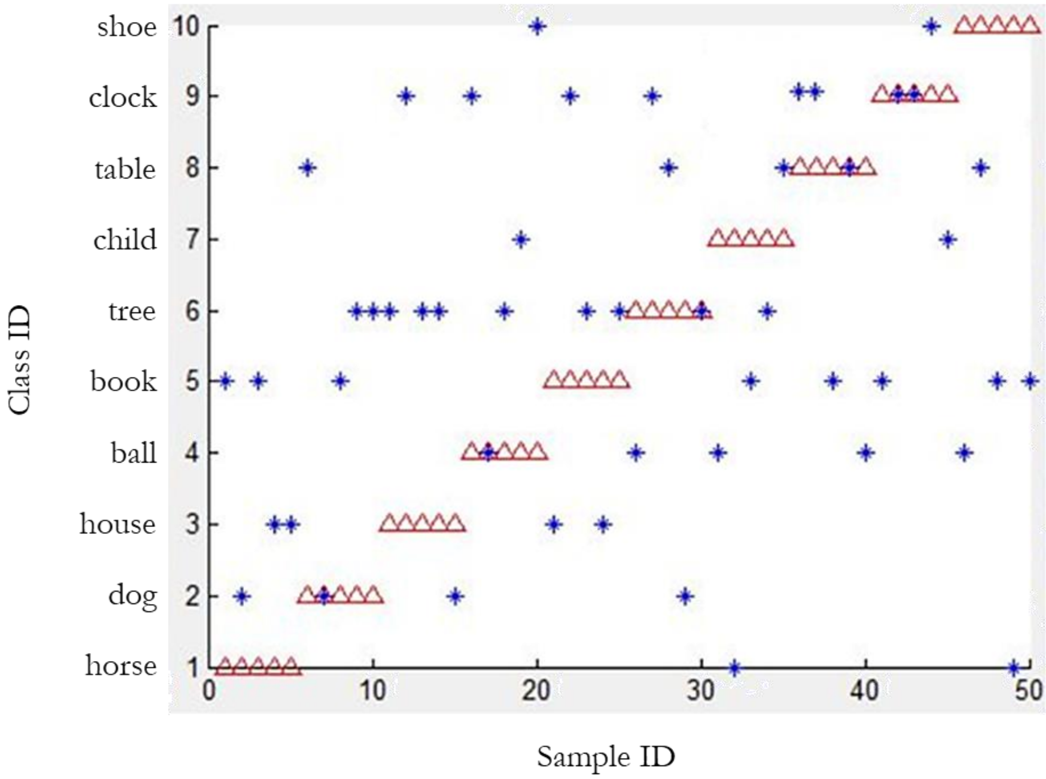


FIGURE 2-1: OVERALL CLASSIFICATION RESULT FOR THE EEG DATA COLLECTED WHILE PRESENTING IMAGES. 50% OF THE SAMPLES WERE USED FOR TRAINING AND 50% FOR TESTING.

In an attempt to reduce any potential misclassification introduced by noise in the signals, the pre-processing stage was re-visited before conducting the **second experiment** and 20 of the 100 samples that seemed to contain the most irregular features were removed based on visual inspection. The remaining 80 samples were then split into a training and a testing dataset, although unlike the first experiment, 65% of the samples (i.e., 52 randomly selected samples) were used for training the network and the remaining 35% (i.e., 28 samples) were used for testing. This slightly improved the classification accuracy and four out of 28 samples were correctly identified (i.e., 14.29%) as shown in Figure 2-2.

Similar to the first experiment, the classifier seemed to behave somewhat randomly. By far the highest number of samples were classified as ball, which also made this the class with the highest accuracy and, at the same time, the highest number of misclassifications. No samples were classified as shoe or book, even though there was a large allocation of training samples for both these classes, as indicated by the small number of samples in the testing set (red triangles in Figure 2-2). Consequently, the learning process in the network should have been best for these classes.

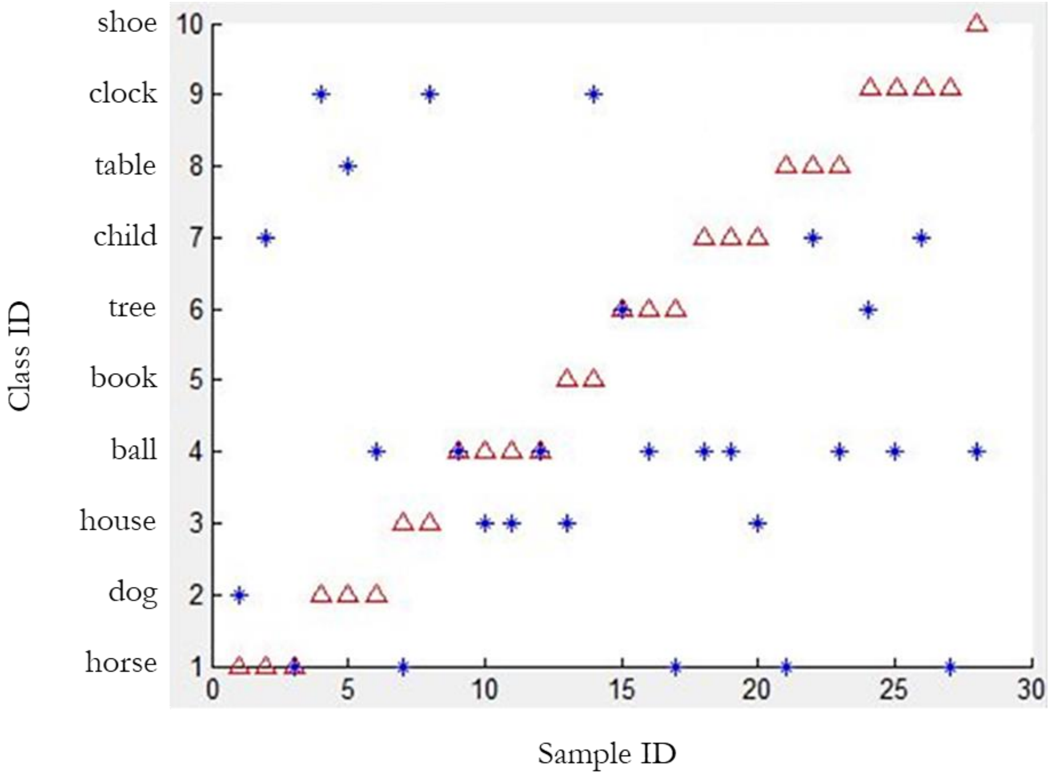


FIGURE 2-2: OVERALL CLASSIFICATION RESULT FOR THE REDUCED EEG DATA COLLECTED WHILE PRESENTING IMAGES. 65% OF THE SAMPLES WERE USED FOR TRAINING AND 35% FOR TESTING.

Based on these results, the **third experiment** looked at comparing only two classes, horse and table, to investigate if the network can discriminate between classes at all. The chosen classes had very little overlap in the first two experiments, so the hope was that they would be reasonably distinguishable. For this experiment, all 20 samples for horse and table from the original full dataset were used to avoid having too few samples when using the reduced dataset. Like in the first experiment, they were split into two groups of the same size, with samples randomly allocated to either the training or the testing group. The overall classification accuracy for this experiment was 50% with 20% of the horse samples and 80% of the table samples correctly classified. The majority of samples were classified as table, which means that the network did not learn to properly discriminate between the two classes. The detailed results for this experiment are shown in Figure 2-3.

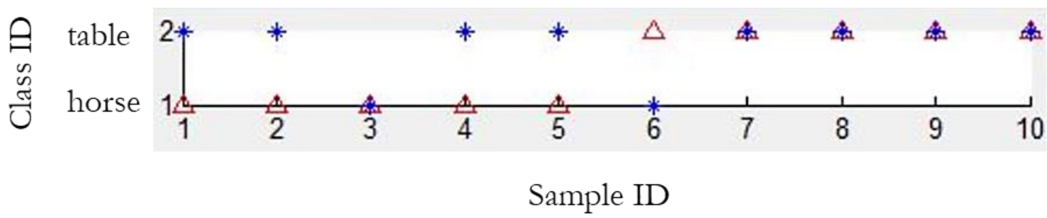


FIGURE 2-3: OVERALL CLASSIFICATION RESULT FOR TWO CLASSES OF THE EEG DATA COLLECTED WHILE PRESENTING IMAGES. 50% OF THE SAMPLES WERE USED FOR TRAINING AND 50% FOR TESTING.

The **fourth** and final **experiment** that was performed on the EEG data for the images looked at categorising the items into living beings and non-living objects. For this experiment, the reduced dataset with 80 samples was used because it had yielded better classification results in the second experiment and with only two target classes, "living beings" and "non-living objects", the impact of missing samples was expected to only have a minor influence. The four items categorised as living beings were child, dog, horse, and tree, while ball, book, clock, house, shoe, and table were categorised as non-living objects. This selection led to a slightly imbalanced dataset with 33 samples for the living beings category and 47 samples for the non-living objects category. The samples were again split in half by randomly assigning them to either a test or training dataset. Due to a bug in the software implementation of the NeuCube, when calculating half of both numbers to split the training and test sets, the numbers for the training dataset were rounded down, leaving 16 samples for the living beings and 23 samples for the non-living objects (39 training samples), and, hence, a total of 41 samples in the test dataset instead of the expected 40.

The overall classification accuracy for this experiment was 65.85% with 23.5% of the living beings and 95.8% of the non-living objects correctly classified. As shown in Figure 2-4, 35 of the 41 samples were classified as non-living objects, which explains the large disparity between the classification results of the categories. As in the third experiment comparing two classes, the network did also not learn to discriminate between the two categories.

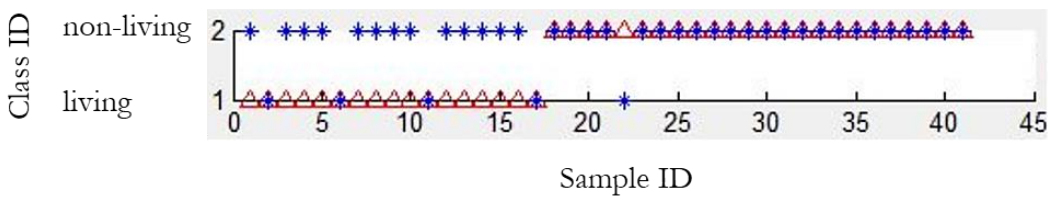


FIGURE 2-4: OVERALL CLASSIFICATION RESULT FOR TWO CATEGORIES OF EEG DATA COLLECTED WHILE PRESENTING IMAGES. 50% OF THE SAMPLES WERE USED FOR TRAINING AND 50% FOR TESTING.

Despite the bad classification results for this experiment, it was still considered to be a useful first attempt at the semantic categorisation of the data. While it was expected that the two categories were sufficiently different to be discriminated, the model's classifier did not manage to distinguish between them. This could, however, have been an issue with the classifier's algorithm rather than being caused by the network not learning those semantic differences. Therefore, taking a deeper look at the connections created in the network seemed like a reasonable next step to gain an understanding of the data's spatio-temporal characteristics.

The trained network was **visualised** by displaying only the strongest 10% of the connections between the neurons. Figure 2-5 shows the network after being trained on all samples for the living beings, while Figure 2-6 shows the same for the non-living objects. The display angle for both figures is from the left rear side of the brain, which means the viewer looks towards the left temporal lobe and the occipital lobe. The blue lines mark the connections between neurons and the line thickness indicates the connection weight. The colours of the neurons correspond to their activation level, i.e., how many times they spiked during the training process: the lighter a neuron is, the more spikes it has emitted, while black neurons have emitted less than five spikes during the whole training process.

The connections in both networks looked relatively similar, especially in the temporal lobes on both sides of the brain. The temporal lobe is generally associated with auditory processing (Amunts, Morosan, Hilbig, & Zilles, 2012), but has also been found to be involved in memory processing (Squire & Zola-Morgan, 1991) and in visual object recognition (Lech & Suchan, 2014; Rolls, 1996). In both networks, there is a noticeable absence of connections in the centre of the occipital lobe, which is generally known as the visual processing centre (Wandell, 1999), even though the task performed by the subject involved high visual attention and processing.

Perhaps the two most prominent differences between the networks can be seen in the left superior temporal lobe and in the right lateral occipital lobe, where the number of connections for the non-living objects is larger than that for the living beings. In return, the network for the living beings exhibits connections in the right frontal lobe that are absent in the network for the non-living objects. In general, the connections in the network for the living beings seem to be more spread out than those for the non-living objects, which appear more clustered.

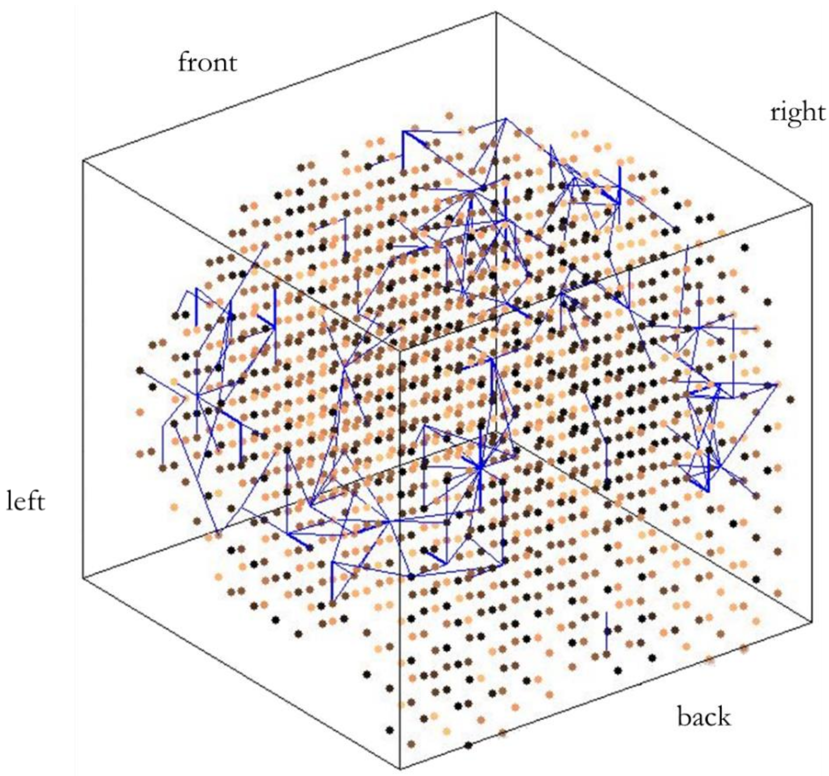


FIGURE 2-5: VISUALISATION OF THE STRONGEST 10% OF THE CONNECTIONS IN A NEURAL NETWORK TRAINED ON EEG DATA RECORDED WHILE SEEING IMAGES OF LIVING BEINGS.

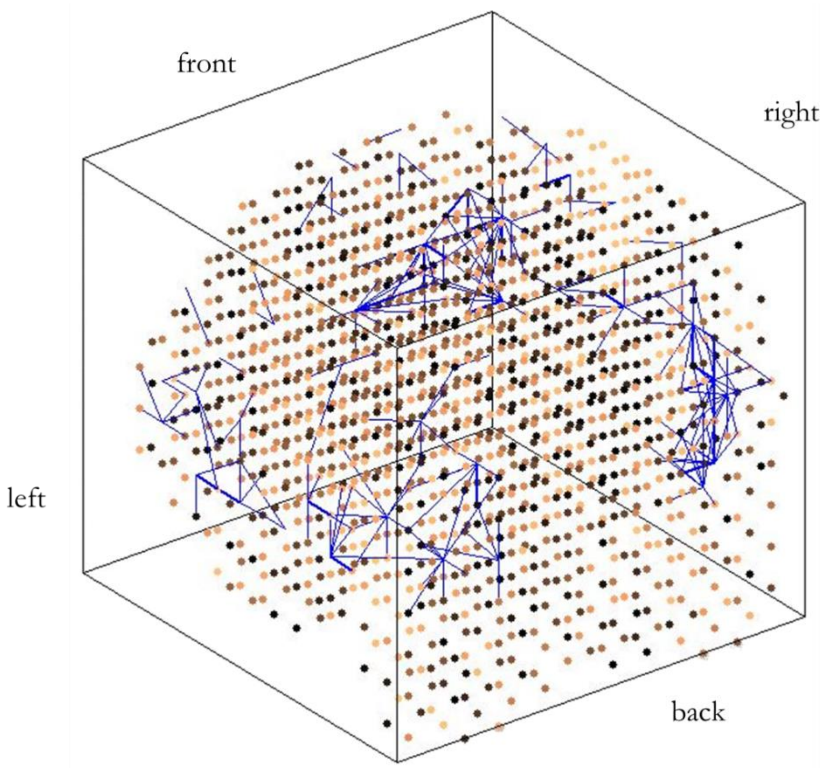


FIGURE 2-6: VISUALISATION OF THE STRONGEST 10% OF THE CONNECTIONS IN A NEURAL NETWORK TRAINED ON EEG DATA RECORDED WHILE SEEING IMAGES OF NON-LIVING OBJECTS.

After analysing the EEG data that were collected while the subject was presented with images, the next set of experiments was performed on the EEG data collected while the subject listened to the **German words**. In this instance, the German words were chosen first over the English words because German was the subject's mother tongue and she was more proficient in it.

Similar to the first experiment on the EEG dataset for images, the German spoken words dataset consisted of 100 samples that were collected while the ten stimuli were presented to the subject, in this case by listening to the spoken words while seeing a white crosshair on the otherwise black screen. In contrast to the image stimuli, however, there was only one recording for each of the ten classes, which was presented ten times. The resulting 100 samples were again split into a training and a test dataset that each contained 50 randomly selected samples. The training dataset was then used to train the network, while the test dataset was used to assess how well the trained network would classify the unseen samples. The overall classification accuracy was 10.0%, which is equivalent to assigning the samples to classes by chance. A detailed overview of the classification results is shown in Figure 2-7.

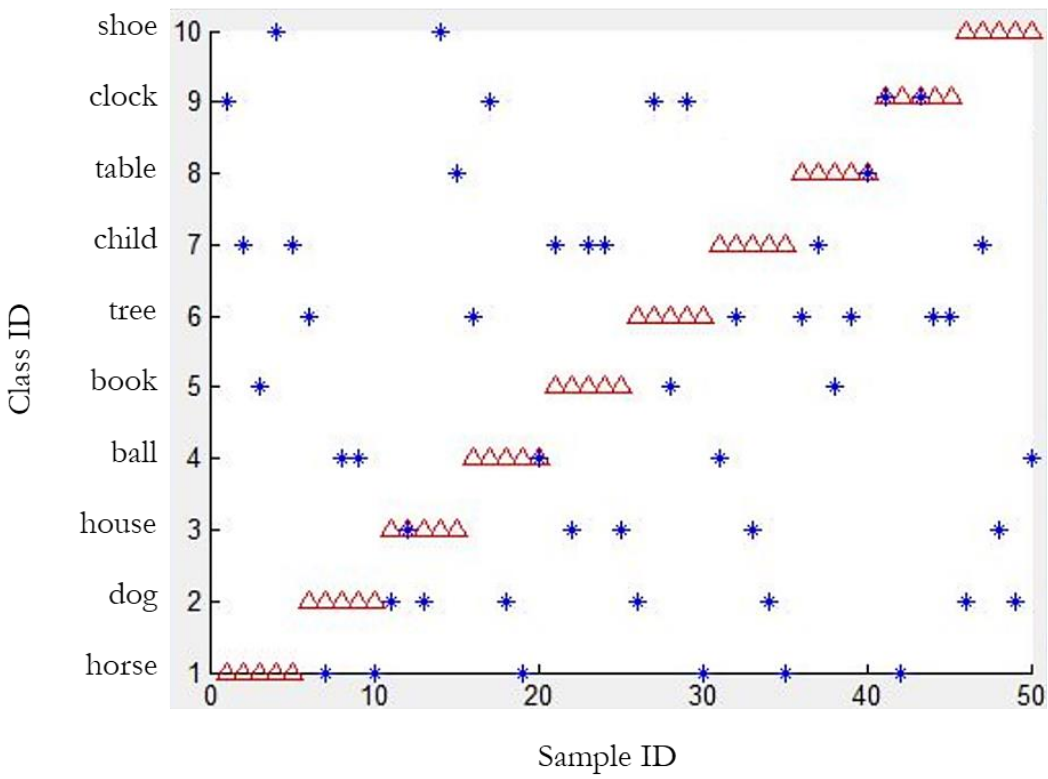


FIGURE 2-7: OVERALL CLASSIFICATION RESULT FOR THE EEG DATA COLLECTED WHILE PRESENTING SPOKEN GERMAN WORDS. HALF OF THE SAMPLES WERE USED FOR TRAINING AND THE OTHER HALF FOR TESTING.

Overall, only five samples were classified correctly, which is likely not due to actual learning in the network but rather caused by the seemingly random assignment of samples to classes that matched by chance in these few cases. There is no clear crossover between classes and the samples are fairly evenly distributed across them. Noticeably, only a few samples were classified as a shoe or table (two samples each) or as a book (three samples), while comparatively many have been misclassified as child, tree, or dog (seven samples each).

No further experiments were conducted with the EEG data for the spoken English or the written German and English words. Based on the results for the first two presentation modes, it was believed that the outcomes for the remaining datasets would not have been significantly better. Moreover, contrary to image and sound processing, identifying written words is a relatively recent cultural addition that the human brain had to handle in its evolution, which means that neural pathways are likely to be embedded into other processing stations and have to be explicitly formed through learning by each person instead of being acquired naturally based on existing neural pathways (Dehaene & Cohen, 2011; I. Y. Liberman, Shankweiler, & Liberman, 1989). This means that the resulting semantics of the written words will be even harder to detect. Studies on orthographic processing using EEG data have so far yielded insights into the temporal order in which the involved neural processing clusters perform visual word recognition (Carreiras, Armstrong, Perea, & Frost, 2014). However, none have reported a semantic identification of the perceived words (Carreiras et al., 2014).

A further planned experiment to analyse possible interactions across all five data acquisition modes was abandoned based on the results obtained from the first two sets of experiments. Even though such an investigation was the original motivation and goal for this study, pursuing this direction further did not seem fruitful enough to invest more time or resources.

Another path that was explored in the experimentation stage was **parameter optimisation** as better parameters may have yielded improved results. A few attempts were made at using the software's in-built parameter optimisation feature that was based on Genetic Algorithms. However, this did not lead to any usable outcomes because the system would run for several days only to be forcefully interrupted by memory errors.

In summary, the network could not discriminate between all ten classes, between a selection of two classes, or between categories. This was consistent across the tested stimulus presentation modes. Overall, the results of this study strongly suggested an adjustment of the framework and research direction. The following section will discuss the possible causes of errors and compare the results to similar published studies.

2.4 ERROR ANALYSIS AND DISCUSSION

This section analyses the results of the EEG experiments from a neuroscientific perspective and contextualises them in relation to other studies that have tried to achieve similar goals using different data analysis frameworks. While in hindsight, some issues with the presented method and experimental setup seem to have been avoidable at the time, it remains questionable if a significant improvement could have been achieved. And after all, the lessons learned from this preliminary study inspired the creative solution that became the foundation of the new data analysis framework presented in this thesis.

It is likely that **process errors** influenced the unfavourable outcome of this study. Firstly, from a methodological point of view, the epoching of the signals at 300 milliseconds after stimulus onset was probably too short, despite the literature cited in Section 2.2 suggesting otherwise. More evidence from studies on event-related potentials (ERP) of EEG data consistently shows another distinctive feature in the signals at about 400 milliseconds after stimulus onset (Kutas & Federmeier, 2011). This so-called N400 component can usually be observed in tasks related to semantic processing regardless of the stimulus type and is, therefore, considered an important feature of any research related to decoding the meaning of words (Kutas & Federmeier, 2011). Secondly, another possible cause of error in the experimental setup was the use of a male and a female speaker for the recordings of the spoken English and German words. The difference in pitch between the genders (Pernet & Belin, 2012) might have led to slightly different processing locations due to the tonotopic organisation of the auditory cortex (Saenz & Langers, 2014), and, hence, facilitated a classification by speaker rather than by meaning. Lastly, there is a possibility that errors were introduced by the analysis method with NeuCube. The software implementation that was available to the thesis author at the time was based on a MATLAB prototype that was not fully tested and therefore, likely contained programming bugs or possibly even algorithmic errors (Scott, 2015, pp. 118-119, 132-139). This became apparent when trying to run an automated parameter optimisation based on Genetic Algorithms, which in all attempts forcefully ended in memory errors or other software crashes.

Another possible source of errors was the thesis author's limited familiarity with established methods of **EEG data pre-processing and analysis** at the time. This meant that she relied on the advice of an expert in the field. Literature on retrieving category knowledge of objects from participants' EEG data had shown encouraging results. However, these could not be replicated with the proposed NeuCube framework. The following paragraphs look at previous studies from literature, their methodology, and results.

For example, Simanova et al. (2010) achieved up to 79% classification accuracy using a similar experimental framework to the one presented in this chapter. While Simanova et al. also looked at classifying EEG data that were collected while the participants saw images (in their case, line drawings) and read and heard word representations of their eight chosen stimuli, a significant difference in their data collection procedure was that they asked their participants to determine the semantic category of the stimulus, i.e., animals or tools. This evoked active thinking on the part of the participants, contrary to the passive presentation of stimuli described in the study presented in this chapter. Simanova et al. then applied Bayesian logistic regression with a multivariate Laplace prior to classify the data into the two categories. They employed five-fold cross-validation (instead of the 50-50 holdout method used in the study presented in this chapter) to validate their classifier and achieved 79% accuracy for the image stimuli, 61% for the spoken words and a modest 56% for the written words across all their 20 study subjects.

Similarly, a set of experiments described by B. Murphy et al. (2011) employed an experimental framework where participants were asked to silently name the animal or the tool that was shown to them as a greyscale image while their EEG data were being recorded. Murphy et al. then applied a time/frequency window search algorithm and a Support Vector Machine for classifying the data with five-fold cross-validation. Their classifier managed to discriminate between the two categories with on average 72% accuracy for single participants. Interestingly, Murphy et al. mention in their discussion that they tried to classify the objects that were presented to the participants, which were 30 land mammals and 30 work tools, but achieved no significant results. They concluded that more advanced data mining techniques needed to be developed to achieve such fine-grained discrimination.

In a third study, this time on EEG data recorded while listening to spoken words in English and Dutch, a research group from the Netherlands looked at which time intervals of the signals after stimulus presentation are most significant when distinguishing between the words (Correia, Jansma, Hausfeld, Kikkert, & Bonte, 2015). Correia et al. presented their participants with ten spoken words, four animals and six inanimate objects, both in English and in Dutch, pronounced by three female speakers. The participants were asked to press a button whenever they heard an inanimate object, which sustained attention during the recording and evoked semantic categorisation of the words. Correia et al. then employed multivariate pattern analysis with a linear Support Vector Machine to discriminate between different words in the same language and generalise semantically similar words across languages. For the intra-language discrimination task, they performed a binary classification between word pairs, similar to the third experiment described in Section 2.3. Their training data consisted of those two-thirds of the data that corresponded to two of the three speakers, while the test data were comprised of the data related to the third speaker. They also made use of a temporal-windows approach to segment the continuous EEG data into intervals that were then analysed separately. On average across all word pairs and time windows, Correia et al. achieved 53% accuracy for the inter-language discrimination of the words. For the intra-language generalisation, Correia et al. split the dataset in half, training on one language and testing on the other. This approach specifically assessed semantic concept formation and overlap across the two languages. On average across all world pairs and time windows, the classification accuracy for this task was 51%.

Besides these approaches using EEG, the task of retrieving semantic meaning from brain signals has also been attempted using magnetoencephalography (MEG) (Sudre et al., 2012; Vartiainen, Parviainen, & Salmelin, 2009), a combination of EEG and MEG (Chan, Halgren, Marinkovic, & Cash, 2011; Hagoort, 2008; B. Murphy & Poesio, 2010), positron emission tomography (Martin, Wiggs, Ungerleider, & Haxby, 1996), and functional magnetic resonance imaging (fMRI) (Buchweitz, Shinkareva, Mason, Mitchell, & Just, 2012; Correia et al., 2014; S. J. Hanson, Matsuka, & Haxby, 2004; Haxby et al., 2001; Huth et al., 2016; Mitchell et al., 2008; Y. Yang et al., 2016), with varying but generally promising results.

The study that was most intriguing to the author of this thesis and inspired her to pursue this research direction was an **fMRI-based semantic brain atlas** developed by Huth et al. (2016). While not strictly being a classification study, the researchers' goal was to investigate semantic selectivity of brain regions across individuals. Huth et al. collected their participants' brain data through fMRI for over two hours while the participants listened to stories. They then mapped the activity of the voxels to the meaning of the words in the stories, resulting in a semantically organised map across the whole brain that was remarkably consistent across the study participants. The map showed that semantically similar concepts elicit activity in spatially close areas in the brain. In a follow-up study by the same research group (Deniz, Nunez-Elizalde, Huth, & Gallant, 2019), the researchers confirmed that these findings hold for different modes of stimulus presentation, concretely, for both listening and reading.

The semantic information that can therefore clearly be extracted by analysing brain data, was undetectable by the research framework employed by the study presented in this chapter. The great disparity between what was observed in literature and what was found here led to the conclusion that a radically different data analysis framework had to be developed to explore semantic concept formation in the brain. The following section briefly describes this new approach, which is the core of this thesis, and its motivation.

2.5 LEARNINGS AND CONCLUSIONS

The purpose of the study reported in this chapter was to investigate if the meaning of objects could be retrieved by analysing EEG data that were collected while ten objects were presented to a participant in different stimulus modes. The approach that was chosen was largely based on conventional methods for EEG pre-processing and on a novel spiking neural network architecture called NeuCube (Kasabov, 2014), which performed very poorly on this task. However, while the model could not *detect* the semantic representations of words from brain data, the author of this thesis hypothesised that it would maybe be able to organically *create* such concepts if trained with the right data. These data would first have to be encoded into spikes and then mapped into the network in a biologically plausible way, which could help to facilitate the multi-sensory integration of data in the model and facilitate concept formation. Fascinated and inspired by the sophisticated processing mechanisms in the human auditory and visual pathways and the seemingly easy integration of audio-visual data in the brain to form these elusive semantic concepts, the thesis author decided to try and copy the key characteristics of those processes and use them to build a **brain-inspired audio-visual information processing system**.

Instead of retrieving the semantic information from brain data, this new approach would start at the stage of stimulus perception and then let the model form its own conceptual representation of the data. This approach of bio-inspired perception and mapping differs from traditional methods of audio-visual data analysis in that it employs more biological processes in place of purely mathematical algorithms. The auditory processing pipeline should follow the auditory pathway from the sound transformation in the cochlea to the tonotopic mapping into the auditory cortex, while the visual processing pipeline should be based on the transformation of light in the retina and the retinotopic mapping into the visual cortex. Furthermore, the attempt to combine these two modalities should take inspiration from the integration of those two modalities in the brain. The biological background of these three processes is described in detail in Chapter 3. Choosing the location into which the signals will be mapped based on the stimulus mode and the actual location of the respective processing regions in the brain will facilitate a "natural" means of combining the data in the network. Since the locations of the auditory and visual cortices are known, they can be replicated in a brain-shaped neural network that employs mechanisms of brain-like neural communications. This novel method is presented in Chapter 5, and the results of a set of experiments that were performed to explore the capabilities of the new framework are reported in Chapters 6, 7, and 8.

Overall, it can be concluded that this temporary setback was unexpectedly beneficial for the thesis project as it motivated the development of an exciting and completely new approach to a currently popular aspect of data processing in neural networks.

3 BIOLOGICAL BACKGROUND

"Always approach a case with an absolutely blank mind. It is always an advantage. Form no theories, just simply observe and draw inferences from your observations."

– Sherlock Holmes in The Adventure of the Cardboard Box

3.1 CHAPTER OVERVIEW

The human body is a remarkable piece of natural engineering. Billions of cells work together as little building bricks that have unique functionalities and are able to communicate their respective needs to one another in order to function as a whole. Our senses are no exception to that. They enable us to gather information about noteworthy characteristics of our environment, to understand their significance, and to draw conclusions about possible reactions to external stimuli. Planning one's movements based on seeing possible obstacles, listening and responding in a conversation, smelling and tasting food and deciding if it is likeable – nearly every situation in our daily lives involves sensory processing. This literature overview focuses on two of those senses, namely hearing and vision, and their integration in the brain, in line with the scope and motivation presented in Chapter 1.

Both sensory modalities, hearing and vision, are largely concerned with identifying objects and their respective locations. The ears and the auditory pathway in the brain do so by analysing the composition of air pressure waves at certain frequencies, while the eyes and the visual pathway can perceive and interpret electromagnetic radiation in a defined spectral range. Sound and light can be created or reflected by objects in one's surroundings. The characteristics of these objects, when in focus, are captured in the auditory and visual signals as they interact physically and chemically with each other. These physiological and chemical reactions influence the properties of the stimuli, such as the pitch of a sound or the intensity of light, that reach the ears and eyes. Once perceived, the brain uses neural mechanisms to decode and comprehend these stimuli to create an appropriate response.

This chapter sets the scene for the work described in this thesis by explaining the biological background of how humans perceive and process audio-visual information. It introduces the components and mechanisms that are involved in the auditory and visual perception of our surroundings, specifically the cochlea and the retina, and the auditory and the visual cortices. It starts with an overview of the hearing process and the auditory pathway, followed by the vision process and the visual pathway. A focus has been placed on the spectro- and spatio-temporal aspects of sound and vision perception and processing because these play an essential role in the architecture of the computational model developed as part of this research. The chapter closes with a synopsis of what is known to date about how the perceived auditory and visual signals are combined and interpreted in the brain in order to extract meaningful information and learn from past experiences.

The processes described in this chapter form the basis of the framework described in Chapter 5. The literature to which this chapter refers is predominantly sourced from biology and neuroscience textbooks (Amunts et al., 2012; Bruce, Georgeson, & Green, 2003; Goebel, Muckli, & Kim, 2012; Schnupp, Honey, & Willmore, 2013; Schnupp, Nelken, & King, 2011; Swanston & Wade, 2013) since these provide a comprehensive overview of all the major processes that need to be considered for the construction and design of the computational model. Rapid advancement in technologies for medical imaging and data analysis means that a wide range of active research is conducted in these areas; this research is mentioned where relevant to elaborate on and support the discussion. Consequently, this also means that the processes described here are based on the current understanding of the anatomy and physiology of the ears and eyes and the auditory and visual pathways. Since the focus of the work presented in this thesis lies on the development of computational models and not on the investigation of the underlying biological processes, this current understanding is assumed to be true and final for the purpose of this research.

3.2 THE HUMAN HEARING PROCESS

Hearing is defined as "the process, function, or power of perceiving sound; specifically: the special sense by which noises and tones are received as stimuli" (Merriam-Webster, n.d.-a). In humans, this process is performed by the two ears, and it constitutes one of the five primary senses that allow us to perceive our environment.

Sound is a combination of pressure waves created by the motion or the vibration of objects that is propagated through a medium, usually air, at different frequencies. Sound waves can be reflected or absorbed by surfaces in their path creating an auditory scene with a multitude of information (Schnupp et al., 2011, pp. 2-3). When sound waves enter the ear, they are transformed into electrical signals by the cochlea and sent to the brain, where they are processed to extract information about the characteristics and locations of sound sources. Our auditory apparatus has evolved over millions of years to decode and untangle these sound stimuli and is capable of gaining detailed knowledge from this process.

This section describes the two main aspects of the hearing and understanding process, namely the perception of sound by the ear, and the processing of signals by the auditory cortex in the brain. The first subsection discusses the architecture and functioning of the inner ear and its intricate mechanical properties that can transform sound waves into electrical signals. The second subsection then discusses what happens to these signals along the auditory pathway and how they are further processed in and beyond the auditory cortex. The processes described here form the basis for the development of the sound processing system presented in Section 5.3.

3.2.1 COCHLEA, BASILAR MEMBRANE, AND HAIR CELLS

Our ears, and in particular the inner ear, perform the task of translating the physical properties of sound waves into mechanical and further into electrical energy to be processed by the brain. The level of detail that is preserved through this transformation is quite significant with regards to frequency composition, loudness, and spatial arrangement of sound sources. This is made possible by the specialised structure and physiognomy of the ear and its components as illustrated in Figure 3-1. From left to right, the diagram shows a cross-section of the outer and middle ears as well as the inner ear and the nerves that are connected to the brain. The dashed green line shows the pathway of the sound stimuli.

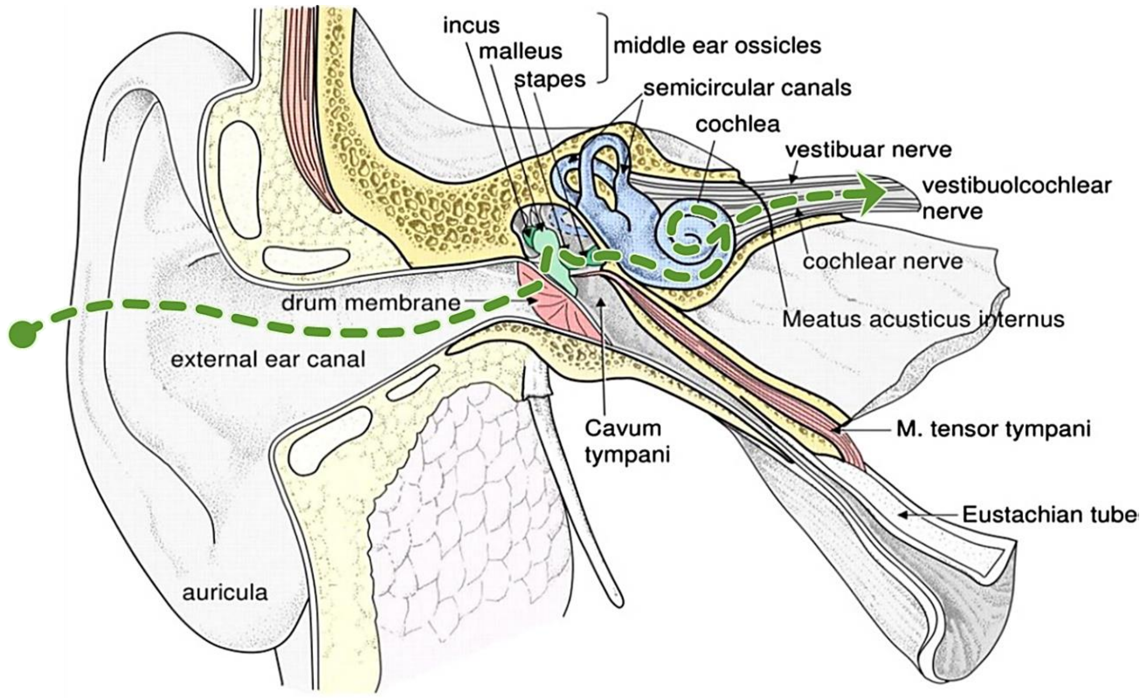


FIGURE 3-1. CROSS-SECTION OF THE HUMAN EAR AND ITS COMPONENTS.[5]

When sound waves reach the eardrum, they create vibrations that activate three bones in the middle ear called the incus, malleus, and stapes after the Latin terms for their appearance as anvil, hammer, and stirrup, respectively (Schnupp et al., 2011, p. 51). These bones are connected to each other and pass on the incoming movements through the middle ear to stimulate the oval window, the entrance of the **cochlea** through which the vibrations are transferred into the cochlea's lymphatic liquids. This process amplifies the sounds waves coming from the relatively sparse molecular arrangement of the air so that they can penetrate the relatively dense molecular arrangement of liquid in the cochlea.

The processing steps inside the cochlea can be described as a bio-mechanical version of a Fourier transformation in that the sound waves are separated based on their frequency composition and sound intensity (Schnupp et al., 2011, pp. 14-15). This is made possible by the unique structure of the cochlea and its components as shown in Figure 3-2.

The cochlea is a spiral-shaped organ that contains three tubes filled with lymphatic liquids. Between the two larger of these tubes, the scala vestibuli and the scala tympani, runs a bone-like structure called the **basilar membrane**. This membrane is relatively narrow and stiff at the base of the cochlea and relatively wide and flexible towards its apex at the top (Schnupp et al., 2011, p. 55).

[5] Adapted from *Chapter 36 – Auditory System* by Amunts et al. (2012, p. 1271). Reproduced with permission.

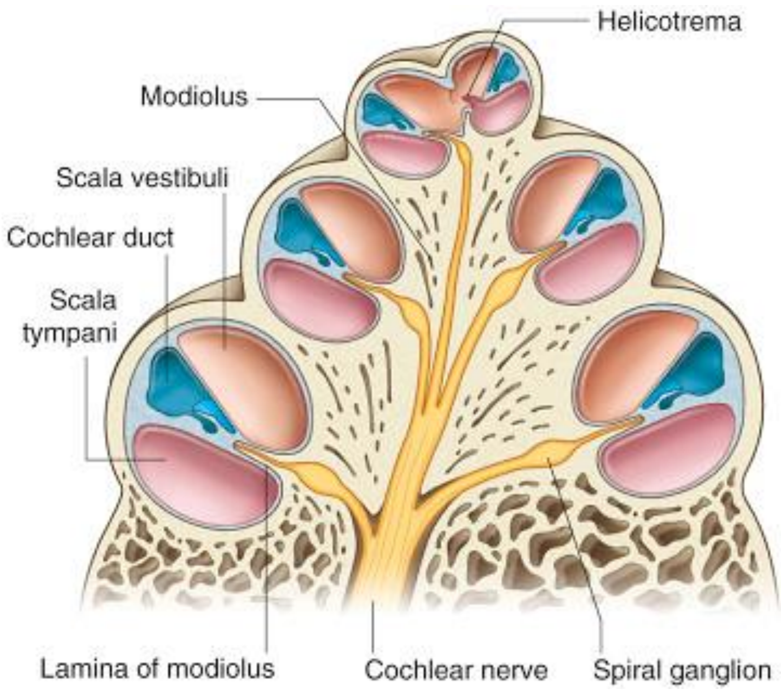


FIGURE 3-2: CROSS-SECTION OF THE COCHLEA AND THE COCHLEAR NERVE.[6]

On top and along the whole length of the basilar membrane sits the organ of Corti, which is responsible for transforming the vibrations of the sound waves into electrochemical signals. The electrical signals created by the cochlea are then sent to the brain through a large number of auditory nerve fibres, which together form the cochlear nerve. The process by which the organ of Corti does this transformation is made possible by the physical structure of the basilar membrane and the inertia of the lymphatic fluid that surrounds it. As one force influencing the equation, the basilar membrane is stiffer at the base of the cochlea than at its apex, which creates a certain amount of resistance in the membrane that is dependent on the distance from the cochlea's base. The other force that is at play during the process of sound transformation is related to the amount of fluid that needs to be moved by the vibrations created by the sound waves, the force required increases towards the apex. The fluid also creates a certain amount of resistance that is dependent on the distance to the base of the cochlea, however, the gradient of the resistance is reversed when compared to that of the stiffness of the basilar membrane. Figure 3-3 shows this relationship schematically in a diagram. Increasing darkness levels of the green colour indicate increasing inertia of the lymphatic liquids, while lighter blue colour indicates less stiffness of the basilar membrane.

[6] From *Gray's Anatomy for Students* by Drake, Vogl, and Mitchell (2020, Fig. 8.130). Reproduced with permission.

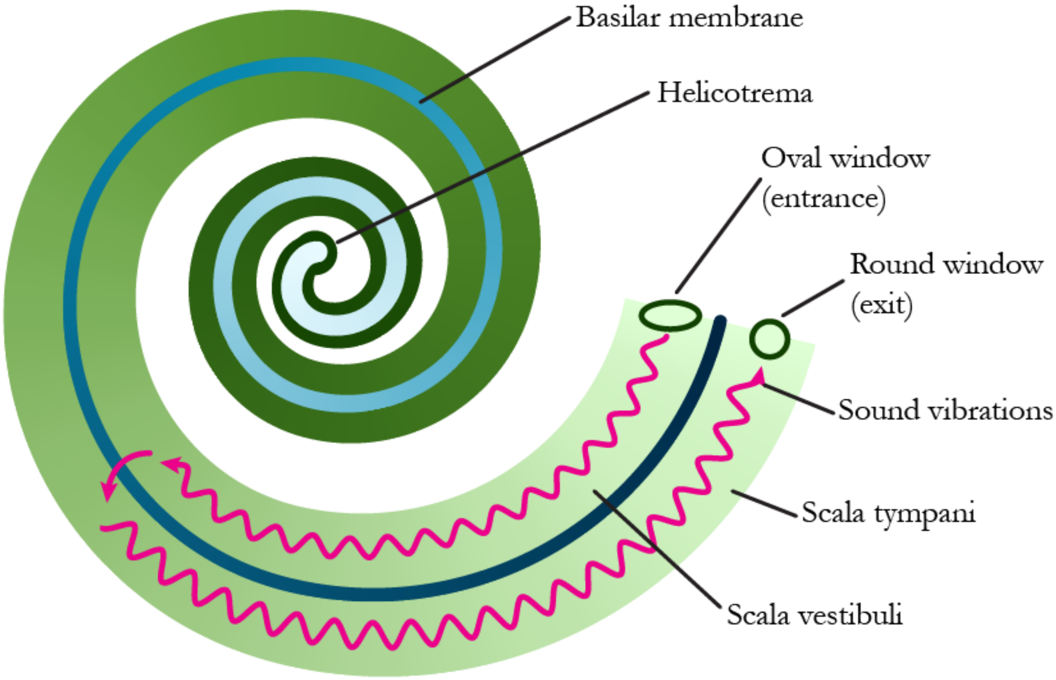


FIGURE 3-3: DIAGRAM OF SOUND WAVES TRAVELLING THROUGH THE COCHLEA AND BASILAR MEMBRANE. COLOUR GRADIENTS INDICATE FLUID INERTIA AND MEMBRANE STIFFNESS, RESPECTIVELY.

The vibrational energy (pink line in Figure 3-3) entering the cochlea through the oval window must first travel through the fluid in the scala vestibuli, then cross through the basilar membrane at a particular point, then travel back to the base of the cochlea through the scala tympani, and finally leave the cochlea through the round window. The point at which the vibration crosses the basilar membrane is determined by the frequency of the sound: higher sound frequencies travel through a point that is closer to the base of the cochlea, while lower sound frequencies travel through a point that is closer to its apex.

The reason for this kind of frequency-based mapping or **tonotopic organisation** is the relative amount of energy that is required to overcome the inertia of the fluid and the stiffness of the basilar membrane. The energy from the sound vibrations sets the lymphatic fluid in the scala vestibuli in motion – the higher the sound frequency, the faster the fluid needs to move. However, the fluid's inertia creates a resistance, which means that vibrations with higher frequencies will not be able to travel very far through the scala vestibuli and rather cross the basilar membrane at an earlier stage. Low-frequency vibrations, on the other hand, will travel further down the tube until there is a point where they can physically overcome the resistance caused by the stiffness of the basilar membrane. At the lower end of the audible spectrum around and below 20 Hertz, the vibrations will not pass through the basilar membrane at all. Instead, these very low-frequency vibrations only pass through the helicotrema in the apex of the cochlea, making this the lower cut-off frequency for the human hearing spectrum (Schnupp et al., 2011, pp. 55-56).

Interestingly, but not surprising given the physical processes involved, lower-frequency sounds have the ability to *mask* higher-frequency sounds of the same intensity. As the lower-frequency sound waves travel through the cochlea, their vibrations also excite earlier parts of the basilar membrane which interferes with vibrations from higher-frequency sounds perceived at the same time. This phenomenon is commonly known as the upward spread of masking and was first reported almost 100 years ago (Wegel & Lane, 1924).

Once the signals have found their way through the cochlea they trigger an electrical impulse that can be sent to the brain. The part of the cochlea responsible for this so-called process of transduction is the organ of **Corti**. It is located on top of the basilar membrane in the cochlear duct and contains two sets of hair cells, different supportive cells, and auditory nerve fibres. Figure 3-4 shows a simplified schematic cross-section of the cochlear duct with the organ of Corti and its main components.

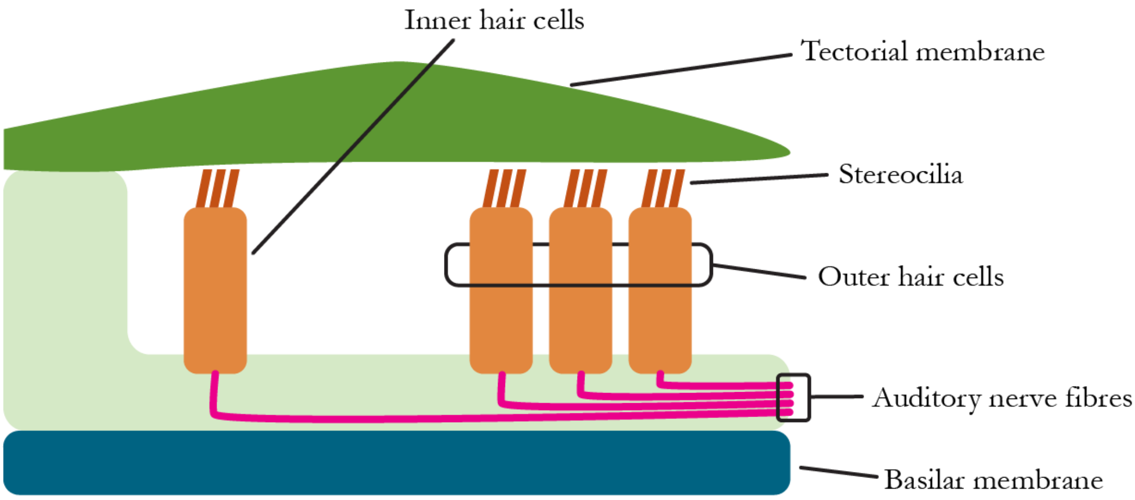


FIGURE 3-4: SIMPLIFIED SCHEMATIC CROSS-SECTION OF THE COCHLEAR DUCT WITH THE ORGAN OF CORTI AND ITS MAIN COMPONENTS.

The **hair cells** are arranged in several rows along the length of the basilar membrane. Their name is derived from the 50 to 200 stereocilia located on the top of each of these cells that resemble hairs. When the basilar membrane starts to vibrate from the energy of the sound waves, the hair cells are moved up and down as well, which pushes the stereocilia into the tectorial membrane located above the organ of Corti. This opens up cation-selective ion channels in the hair cell's membrane, leading to rapid depolarisation of the hair cell and thus creating an electrical impulse that is sent through the connected auditory nerve fibres to the brain (Schnupp et al., 2011, pp. 64-69).

There are about 3,500 inner hair cells running along the length of the cochlea in a single row, and approximately 8,500 outer hair cells forming three rows (Wright, Davis, Bredberg, Ulehlova, & Spencer, 1987). There are, equally, two types of **auditory nerve fibres**. Each inner hair cell is connected to several so-called Type I fibres, while several outer hair cells converge into single Type II fibres (Spoendlin & Schrott, 1989). It is generally assumed that the approximately 30,000 Type I fibres are more important for the transmission of auditory signals than the approximately 3,000 Type II fibres due to their comparatively larger number as well as being myelinated and thus having a faster transmission rate (Schnupp et al., 2011, pp. 75-76; Spoendlin & Schrott, 1989). The functionality of Type I fibres as the main transmission path for auditory signals has further been shown in several audiology studies (Goutman, Elgoyhen, & Gómez-Casati, 2015). Type II fibres, on the other hand, do not show sufficient firing activity to support the encoding of auditory signals (Weisz, Glowatzki, & Fuchs, 2009). While it is believed that they could be involved in detecting noise-induced damage (Flores et al., 2015), research is still ongoing to identify their exact functionality (Goutman et al., 2015; Heil & Peterson, 2015).

3.2.2 AUDITORY PATHWAY AND AUDITORY CORTEX

This section describes the stations along the primary auditory pathway from the cochleae to the auditory cortices. The electrical impulses that are created by the sound waves in the inner ear are sent along the **primary auditory pathway** to the auditory cortices located in the left and right temporal lobes of the brain, specifically on the transverse temporal gyri, also known as Heschl's gyri (Da Costa et al., 2011). The primary auditory pathway involves a variety of brain regions that perform subcortical functions such as sound source recognition and localisation. Its most significant processing stations are shown in Figure 3-5. All neural processing units exist bilaterally with multiple cross-overs between the right and left hemispheres, but generally contralateral projection from the cochleae to the auditory cortices, i.e., what is heard by the cochlea in the right hemisphere will eventually be processed by the auditory cortex in the left hemisphere, and vice versa (Langers, van Dijk, & Backes, 2005).

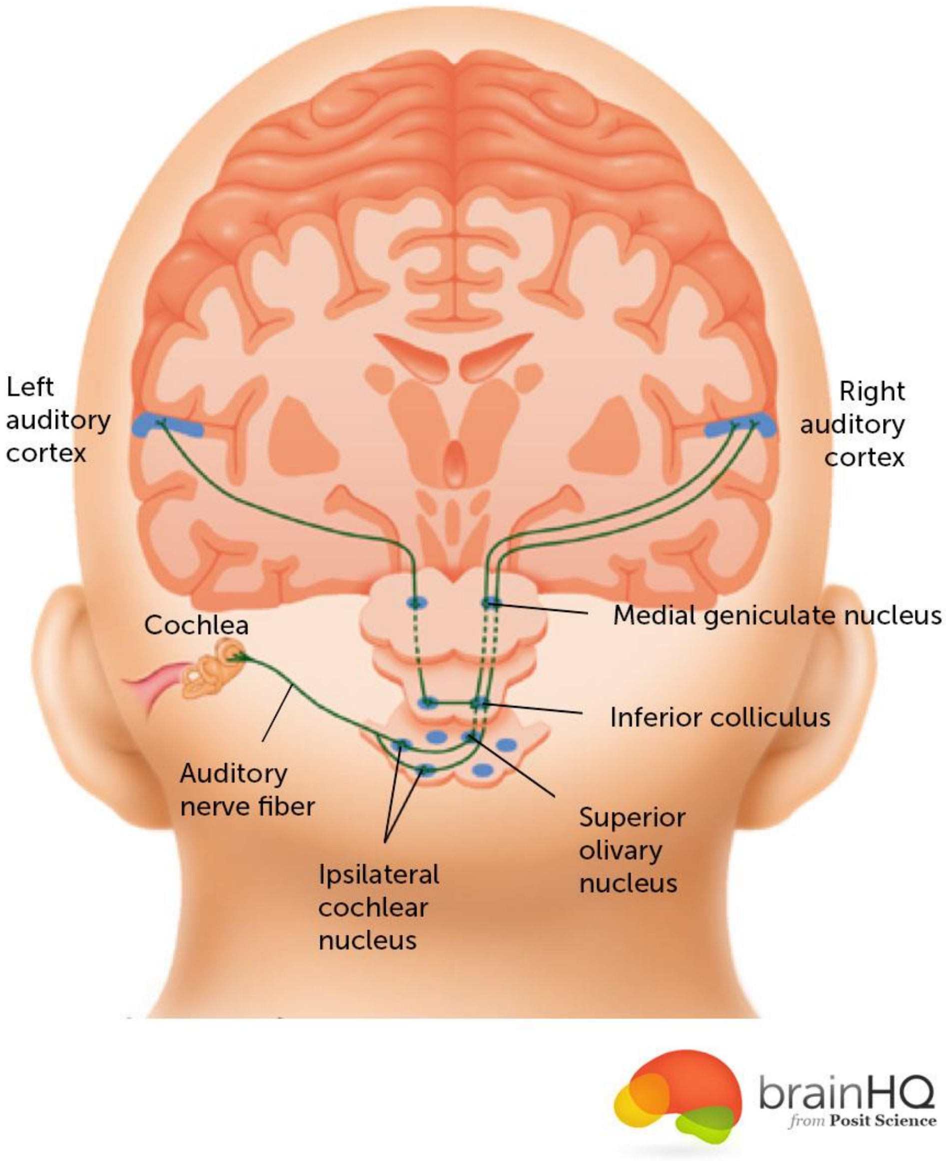


FIGURE 3-5: THE PRIMARY AUDITORY PATHWAY AND ITS PROCESSING STATIONS.[7]

The auditory nerve fibres originating in the left and right cochleae project through the vestibulocochlear nerves into their respective ipsilateral **cochlear nuclei**, where the signals are processed by several different types of neurons that are specialised in certain aspects of sound processing such as extracting particular frequencies or determining exact signal timing (Schnupp et al., 2011, pp. 86-88). Depending on this specialisation, some signals are then transmitted directly to the inferior colliculus, while others pass through the superior olivary complex first before also reaching the inferior colliculus (Amunts et al., 2012, pp. 1273-1275).

This step through the **superior olivary complex** is essential for sound localisation, as signals from both ears cross over for the first time (Brugge & Geisler, 1978). There are two methods

[7] From Posit Science brainHQ (https://www.brainhq.com/wp-content/uploads/2018/12/auditory-pathways.jpg). Reproduced with permission.

by which the location of a sound source can be identified as shown in Figure 3-6. The *lateral* superior olives detect the interaural *level* difference of the primarily higher-frequency sound that is caused by the head casting a “sound shadow”. This means that sound waves reaching the ear that is closer to the sound source will perceive the sound as louder than the other ear. The neurons in the lateral superior olive can measure this difference in perceived loudness because they are excited by ipsilateral input and inhibited by contralateral input from the cochlear nuclei, which creates an algorithmically simple calculation between positive and negative voltage values (Grothe, Pecka, & McAlpine, 2010). On the other hand, the *medial* superior olives detect the interaural *time* difference of mainly lower-frequency sound that is caused by the distance between the ears and sound waves taking slightly longer to reach the ear that is further away from the sound source. The neurons in the medial superior olives are excited by both ipsilateral and contralateral input, and they are specialised in using this input to detect minuscule time differences in signal arrival times (Grothe et al., 2010).

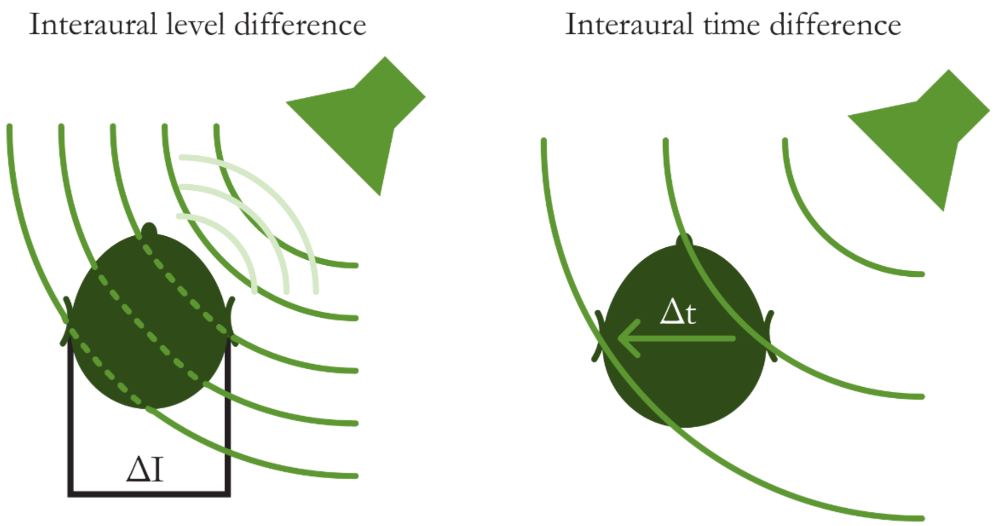


FIGURE 3-6: INTERAURAL LEVEL DIFFERENCE AND INTERAURAL TIME DIFFERENCE FOR SOUND LOCALISATION.[8]

The next step on the primary auditory pathway is through the **inferior colliculi**. These neural clusters are located in the midbrain and receive the majority of their incoming connections through a nerve bundle called the lateral lemniscus. The lateral lemniscus mainly serves as a connection point for ascending neural fibres and as a gateway to the inferior colliculi (Amunts et al., 2012, pp. 1278-1279). The signals arriving at the inferior colliculi are routed either directly from the cochlear nuclei or indirectly via the superior olivary complex as described before. Both the right and left inferior colliculi receive bilateral input, and they are densely interconnected with each other (Amunts et al., 2012, p. 1280). Although not fully

[8] Adapted from *Mechanisms of Sound Localization in Mammals* by Grothe et al. (2010). Reproduced with permission.

understood, they are probably involved in sound localisation (Grothe et al., 2010; Litovsky, Fligor, & Tramo, 2002) and integrating visual and other multisensory information with the sound signals on the subcortical level to enhance and complete sound information (Gruters & Groh, 2012).[9]

The signals from the inferior colliculi are then sent to the **medial geniculate body**, which is located in the thalamus (Amunts et al., 2012, p. 1280). It has three distinct parts called ventral, dorsal, and medial division based on their locations within the medial geniculate body. The three divisions mainly differ in their prevalent cell types, and only the ventral division is tonotopically organised (Amunts et al., 2012, pp. 1281-1282). Both the inferior colliculus and the medial geniculate body are probably involved in improving speech recognition by integrating visual cues from facial movements (von Kriegstein, Patterson, & Griffiths, 2008). The medial geniculate body is also believed to trigger subcortical emotional responses due to its connection to the amygdala (Sander & Scheich, 2001; Schnupp et al., 2011, p. 90) and to be responsive to somatosensory input and pain intensity (Amunts et al., 2012, p. 1282).

Along the auditory pathway, auditory signals from both the left and right cochleae cross over at several stages (Amunts et al., 2012, pp. 1289-1290) before finally reaching their respective contralateral **primary auditory cortices** in the right and left temporal lobes (Langers et al., 2005). These small but highly specialised processing centres are located in the transverse temporal gyri, also known as Heschl's gyri, and are generally associated with Brodmann area 41 (Brodmann, 1909, pp. 144-145).[10] Based on cytoarchitectonic observations, the auditory cortex can be divided into three areas called Te1.0, Te1.1, and Te1.2, which are composed of different cell types and densities and are, therefore, assumed to fulfil different aspects of the hearing process (Morosan et al., 2001). Although the locations of specific functionalities are not yet satisfactorily identified, these different regions in the auditory cortex seem to be specialised in fulfilling different tasks of the hearing process (Read, Winer, & Schreiner, 2002; Semple & Scott, 2003).

In addition to the discussed ascending auditory pathway from the cochleae to the auditory cortices, the auditory system has also developed **descending pathways** between several processing areas (Amunts et al., 2012, pp. 1290-1291). Although poorly understood, there is evidence that these enable higher-level processing areas to modify the behaviour of lower-

[9] This is the first known biological cross-over of auditory and visual signals in the brain. With regards to the main topic of the thesis, the implications of this finding are discussed in detail in Section 3.4.

[10] Interestingly, however, Brodmann himself believed that "it is totally unthinkable that such an important cortical function like the hearing should be limited to such a small part of the whole cortex." (Brodmann, 1909, p. 315), dismissing previous studies that had suggested otherwise (Flechsig, 1908). A detailed discussion of the size and location of the auditory cortices can be found in Section 5.3.2.

level areas. For example, the auditory cortex can influence how the neurons of the inferior colliculus respond to sound frequency, intensity, and location (King & Bajo, 2013).

For higher-level signal interpretation, the auditory cortex has "where" and "what" **processing streams** to localise sound sources and to recognise objects and patterns, respectively (Rauschecker, 2013; Rauschecker & Tian, 2000). These pathways are shown in Figure 3-7. While the dorsal stream projecting through the parietal lobe helps to resolve spatial references and determine object movement, the ventral stream projecting through the temporal lobe is involved in object identification and spoken word recognition (DeWitt & Rauschecker, 2012).

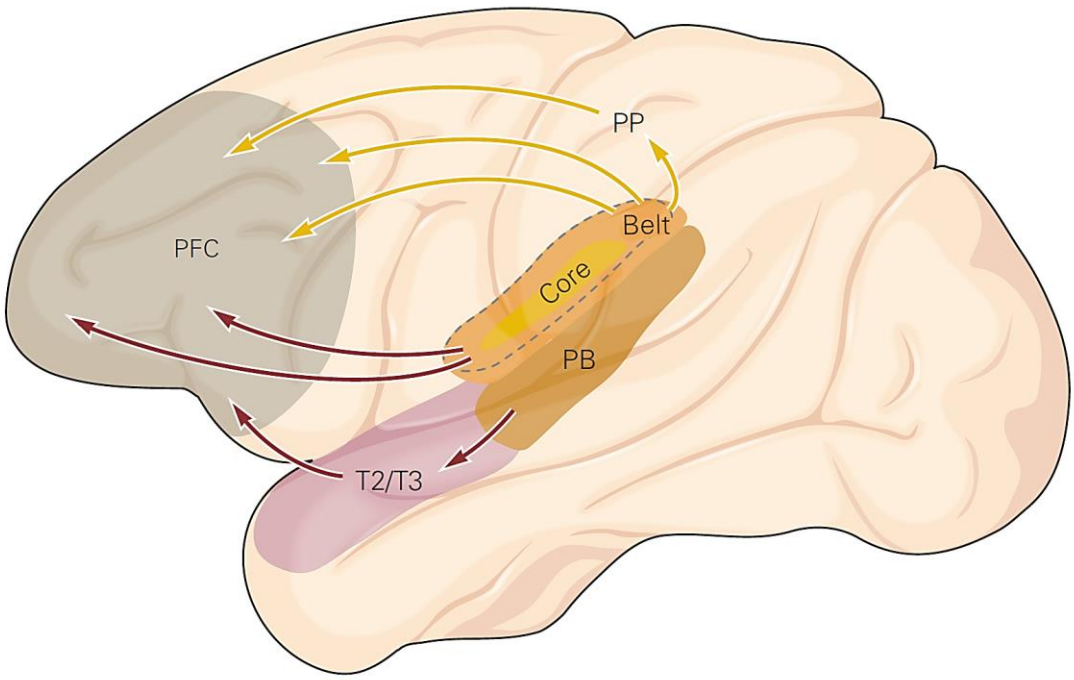


FIGURE 3-7: THE "WHAT" (VENTRAL, RED ARROWS) AND "WHERE" (DORSAL, YELLOW ARROWS) PROCESSING STREAMS IN THE AUDITORY SYSTEM.[11]

The auditory cortex and the areas to which it projects can identify **auditory objects** and their locations as sources of the perceived sounds based on characteristics such as frequency composition, loudness, and timing of the sound waves (Schnupp et al., 2013). Sounds created by auditory objects contain information about the object's properties such as the size, weight, shape, material, movement, or mode of sound creation that are used in conjunction with the other senses to draw conclusions about the listener's surroundings (Griffiths, Micheyl, & Overath, 2012; Kubovy & van Valkenburg, 2001; Schnupp et al., 2013). Several studies have also shown that emotional and contextual cues contribute to the perception and

[11] From *Chapter 31 – The Auditory Central Nervous System* by Oertel and Doupe (2013, p. 704). Reproduced with permission. **PFC**, prefrontal cortex; **PP**, posterior parietal cortex; **PB**, parabelt cortex; **T2/T3**, areas of temporal cortex.

interpretation of sounds (Asutay & Västfjäll, 2012; Pourtois, Schettino, & Vuilleumier, 2013; Stefanucci, Gagnon, & Lessard, 2011; Västfjäll, 2002). Furthermore, the auditory processing system has developed the ability to focus on only a subset of sound streams that are related to particular objects of interest (Schnupp et al., 2013), which is for example used to filter out speech in noisy environments (Mesgarani & Chang, 2012; Shinn-Cunningham, 2008).

But how does the brain develop all these processing steps? Not much is known about the **plasticity** of the auditory cortex (Shepard, Kilgard, & Liu, 2013), except that diverse forms of Hebbian plasticity, such as spike-timing-dependent plasticity, could be observed (Tzounopoulos & Leão, 2012). There is considerable evidence from studies on rodents that the majority of neural responses to acoustic stimuli are shaped during infancy (de Villers-Sidani, Chang, Bao, & Merzenich, 2007), although it has also been found that the auditory cortex of adult monkeys and humans can still learn or alter neural responses to sounds, depending on the significance of the stimuli as indicated by other sensory processing units (Dahmen & King, 2007).

The functioning of the auditory system that was explored in this section informed the design of the sound processing model described in Section 5.3. In particular, the functioning of the hair cells and the auditory nerve fibres was considered highly relevant for the first step in the model, a cochlea-inspired sound signal transformation. This is described in Section 5.3.1. The transformed signals were then intended to be mapped into a neural network in a biologically plausible way. For this process, the size and location of the auditory cortices were determined as described in Sections 5.3.2 and 5.3.3 so that insights from tonotopy studies could be applied. The auditory pathway as a signal relay and integration station was not included in the sound processing model due to its many uncertain interactions with other areas of the brain that were not part of the model at this stage. However, a novel process of signal compression was developed that was intended to compensate for this omission. This process is described in Section 5.3.4.

3.3 THE HUMAN VISION PROCESS

Vision is defined as "the special sense by which the qualities of an object [...] constituting its appearance are perceived through a process in which light rays entering the eye are transformed by the retina into electrical signals that are transmitted to the brain via the optic nerve" (Merriam-Webster, n.d.-b). This summary encompasses all the essentials of the important sensory function that is described in more detail in the following sections.

Within the context of this thesis, light is defined as electromagnetic radiation in the spectrum that can be perceived by the human eye, where it has a wavelength between about 390 and 700 nanometres ranging from violet over blue, green, yellow and orange to red, respectively (Bruce et al., 2003, p. 4). Light can be created both naturally and artificially by a variety of means. For example, chemical reactions, like the sun burning millions of tons of gas, or other ways of releasing energy from atoms, like high temperatures (e.g., lava) or electricity for electric lamps, can create light varying in brightness and colour. This light then scatters through the air and can be reflected, absorbed, refracted or diffracted by objects that it meets or through which it passes (Bruce et al., 2003, p. 5). Thus, the light rays contain a multitude of information about the objects in one's surroundings like their location and physical characteristics. This information is then perceived and decoded by the eye and the brain, which have evolved over millions of years to excel at this task and gain detailed knowledge from this process to elicit an appropriate response.

This section is divided into two parts introducing two main aspects of the human vision process, namely the perception of light rays by the eyes and the processing of signals by the visual cortex in the brain. The first subsection discusses the architecture and functioning of the retina and its characteristic photoreceptor cells that can transform the energy from the light rays into electrical signals. The second subsection then explains what happens with these signals along the primary visual pathway and how they are further processed in the visual cortices. The aspects described here form the basis for the development of the video processing system described in Section 5.4.

3.3.1 RETINA, PHOTORECEPTORS, AND BIPOLAR CELLS

The ubiquity of light has caused almost all living beings to develop some sort of ability to perceive or make otherwise use of it. In bacteria, for example, light can trigger chemical reactions required for locomotion or nutritional processes, while plants use light energy to conduct photosynthesis, a major component and the base of their life cycle (Bruce et al., 2003, pp. 7-8). Many more applications exist in which light plays a central role in a being's development and behaviours (Wolken, 1975).

The focus of this section lies on vision and, in particular, on the human eye and how its components allow it to encode what it sees in a way that is fast, reliable and preserves as much information as possible. A cross-section overview of the eyeball and its most important parts is shown in Figure 3-8.

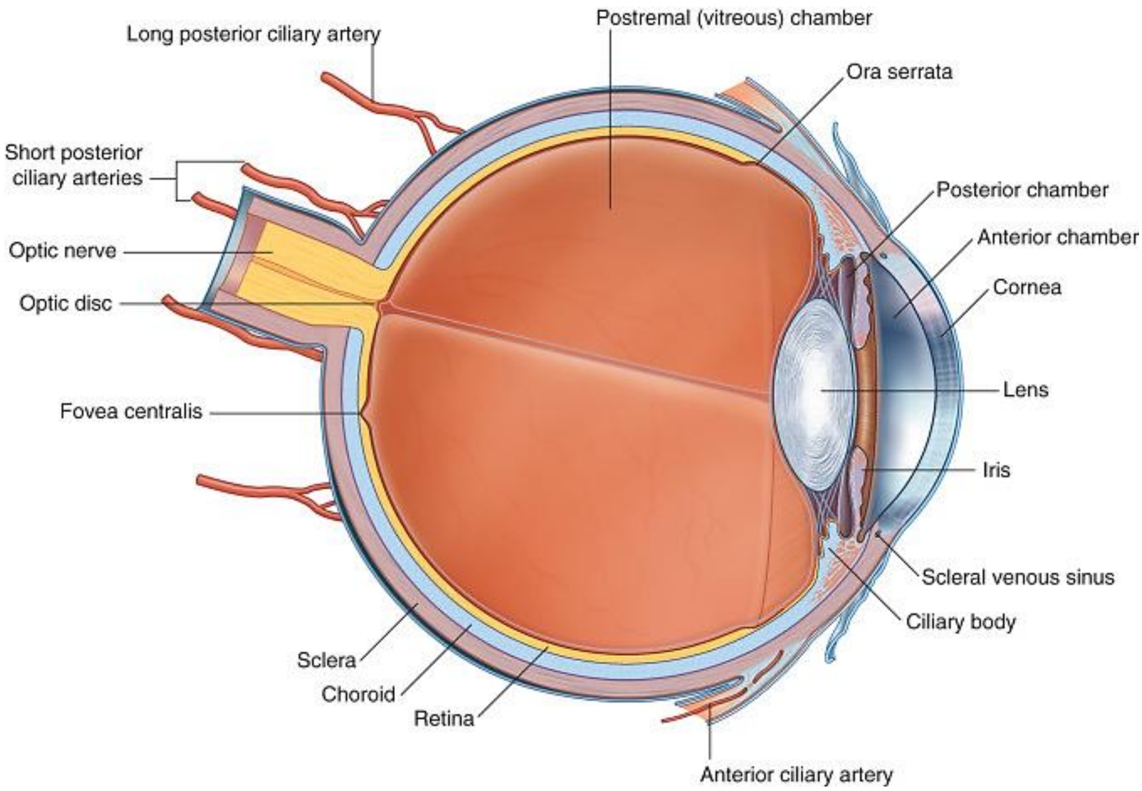


FIGURE 3-8: CROSS-SECTION OF THE HUMAN EYE AND ITS COMPONENTS.[12]

Before light reaches the retinal photoreceptor cells, it first passes through and is refracted by the cornea and the lens. This process adjusts the focal point of the light to be around 17 millimetres behind the lens, which is about the diameter of the eyeball (Goebel et al., 2012, p. 1302). While the cornea is responsible for most of the light refraction, the lens, due to its flexible shape, is able to finely adjust the focus of an image. Its surrounding ciliary muscles can move the lens's annular ligaments and hence alter its shape and refractional properties (Goebel et al., 2012, pp. 1302-1303; Swanston & Wade, 2013, pp. 116-117). This way, light rays from objects on which the eye wants to focus can be directed to the most sensitive area of the retina, the fovea centralis, independent of the object's distance from the viewer. Another set of muscles involved in pre-processing the image for optimal perception is located in the iris that sits between the cornea and the lens. The muscles in the iris control

[12] From *Gray's Anatomy for Students* by Drake et al. (2020, Fig. 8.108). Reproduced with permission.

the brightness of an image by influencing the amount of light that can enter the pupil through dilation and constriction, supporting the brightness adaptation of the photoreceptor cells (Goebel et al., 2012, p. 1303; Swanston & Wade, 2013, p. 19).

After the light has thus travelled through the pupil and the eyeball, it will reach the **retina** where it is transformed into neural signals. Figure 3-9 shows a cross-section of the retina's cellular organisation. Unlike most other sensory organs, the retina is considered to be a part of the central nervous system, due to its synaptic structure and its formation during fetal development (Goebel et al., 2012, p. 1303). It contains several distinct neural layers that are involved in signal creation, amplification, and suppression based on incoming light with the final layer directly projecting into the mid-brain through long neural axons. The retina performs a certain amount of pre-processing of the incoming stimuli and their features, before passing the information on through the numerous fibres of the optic nerve (Goebel et al., 2012, p. 1303). The most important aspects of these pre-processing steps are described in the following paragraphs.

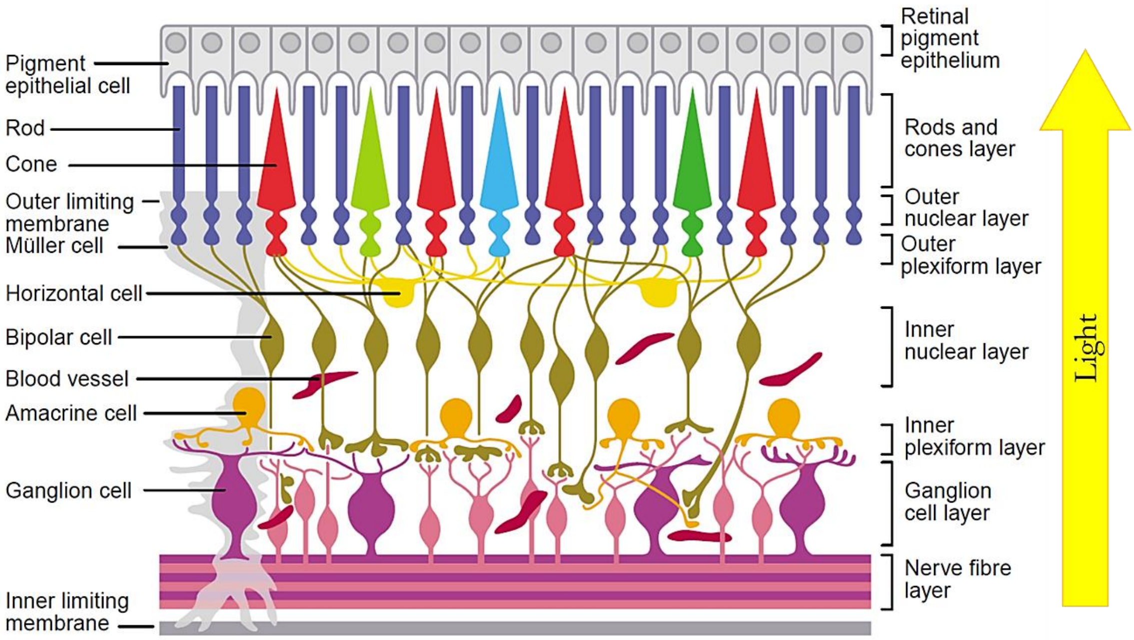


FIGURE 3-9: THE CELLULAR ORGANISATION OF THE RETINA. LIGHT TRAVELS FROM THE BOTTOM OF THE DIAGRAM TO THE TOP.[13]

The aim of the refraction and focusing process conducted by the cornea and the lens is to direct the light rays of the object of interest towards the **fovea** in the centre of the retina. The fovea is the most sensitive part of the retina with the highest density of cone-shaped, colour-perceiving photoreceptor cells, while the more peripheral areas of the retina largely

[13] Adapted from *Diabetes and retinal vascular disorders: role of the renin–angiotensin system* by Wilkinson-Berka (2004, p. 6). Reproduced with permission.

contain rod-shaped, brightness-perceiving photoreceptor cells (Swanston & Wade, 2013, p. 133). The fovea is also the only area of the retina that contains just photoreceptors. In most parts of the retina, the light rays have to travel through several layers of supporting nerve cells before reaching the light-sensitive pigments of the photoreceptors as shown in Figure 3-9. However, the neural connections in the fovea are placed so that the nerve cells can be located *around* the photoreceptors instead of directly on top of them. This removes unwanted scattering of light rays and minimises information loss caused by the nerve cells accidentally absorbing photons (Swanston & Wade, 2013, pp. 132-133).

The two types of photoreceptor cells are usually called **rods and cones** for short, based on the appearance of their light-sensitive outer segments as shown in Figure 3-10.

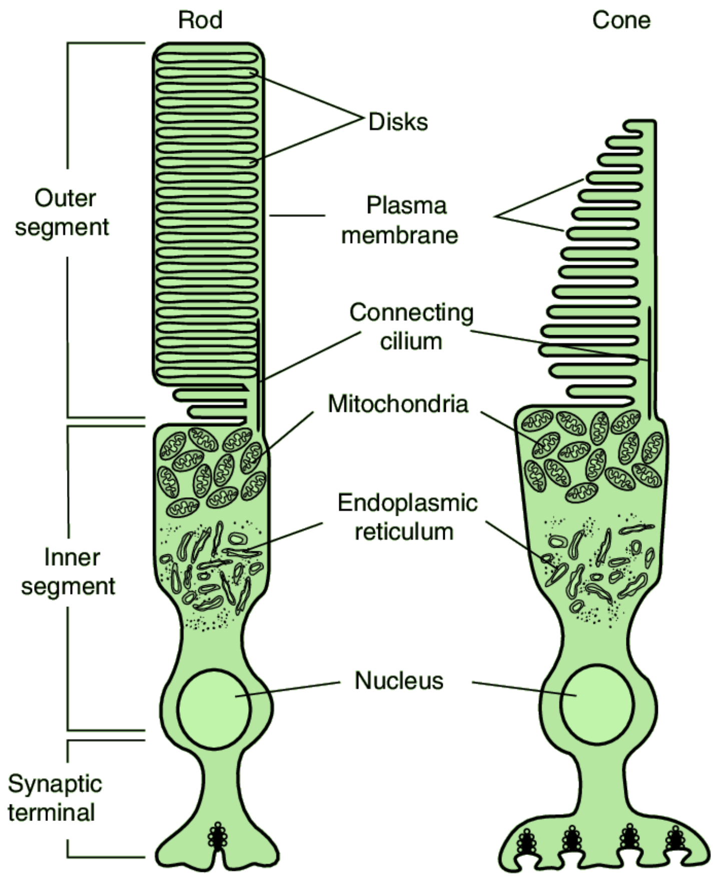


FIGURE 3-10: SCHEMATIC DIAGRAM OF ROD AND CONE PHOTORECEPTORS.[14]

The rods are very sensitive to low levels of brightness – absorbing only one photon of light might be enough to create an electrical impulse (Goebel et al., 2012, p. 1304). The cones, on the other hand, need quite bright light to function but can distinguish between colours. There are three types of cones in the human retina that respond to different wavelengths of light: S-cones ("S" standing for "short-wavelength") have their peak absorption rate of photons at

[14] From *Photoreceptor Phosphodiesterase (PDE6): A G-Protein-Activated PDE Regulating Visual Excitation in Rod and Cone Photoreceptor Cells* by Cote (2006, p. 167). Reproduced with permission.

a wavelength of around 420 nanometres, which corresponds to blue- and purple-coloured light; M-cones ("medium-wavelength") have their peak absorption rate at around 530 nanometres, which corresponds to green-blue light; and L-cones ("long-wavelength") absorb most photons at a wavelength of around 560 nanometres, which corresponds to yellow-green light but also covers colours in the red-light spectrum (Bruce et al., 2003, pp. 21-22; Swanston & Wade, 2013, p. 134). Combining the neural signals from the cones, the brain is then able to determine which colour was perceived. When all three colours are perceived in the same location with the same intensity, the brain combines them to white light, which effectively means that the human visual system uses an additive colour mixing model, albeit arranged across several neural processing stages (Bruce et al., 2003, pp. 21-23; Swanston & Wade, 2013, p. 136).

The transformation of photons from light rays into electrical signals is done by both rods and cones performing a process called **phototransduction**. The outer segments (see Figure 3-10) of the photoreceptors contain a visual pigment, either rhodopsin or one of three cone-opsins depending on the cell type, that consists of a protein called opsin and a photon-sensitive molecule called retinal. In darkness, a molecule called cyclic guanosine monophosphate (cGMP) ensures that sodium channels in the photoreceptor's membrane stay open, causing the cell to be in a resting state of depolarisation at around -40 millivolts. However, as soon as light photons reach the visual pigment, the retinal molecule changes its structure, which causes the opsin to trigger a series of chemical reactions that lead to a decrease of cGMP in the cell and, hence, to a closure of the sodium channels in the cell's membrane. With the sodium channels closed but potassium channels in other parts of the photoreceptor's membrane still open, the cell hyperpolarises – the first step to creating an electric impulse that can be sent to the brain (Bruce et al., 2003, pp. 12-13; Goebel et al., 2012, p. 1304; Pugh & Cobbs, 1986). By binding to the retinal molecules one at a time, the number of perceived photons directly influences the number of closed sodium channels, which means that the intensity of the light is encoded gradually in the strength of the polarisation of the photoreceptor cell (Pugh & Cobbs, 1986).

This process and, more importantly, its reversal to reset the photoreceptor to a state in which it can again react to photons, requires a high level of physiological maintenance, which is one of the reasons why the photoreceptors have to be attached directly to the epithelium and the retina appears to be structured backwards (Strauss, 2005). Minimising the negative impact of this inverted setup by for example unwanted scattering of light, special glial cells called Müller cells have been discovered to "guide" the photons to the visual pigments of the photoreceptors (Reichenbach, Agte, Francke, & Franze, 2014).

As shown in Figure 3-9, the photoreceptor cells are mainly connected to a layer of **bipolar cells**, as well as to a small number of horizontal cells. As long as the photoreceptors are depolarised during darkness, they release an inhibitory neurotransmitter called glutamate to this next neural layer. With increasing brightness and thus stronger polarisation of the photoreceptor cell, the release of glutamate is reduced proportionally. There are an estimated 12 types of bipolar cells in the human retina (Masland, 2012). Two of those, called "on-centre" and "off-centre" cells based on their polarity, have been found to have the ability to either turn the neurotransmitter signal from the photoreceptors around or pass it on directly to the next neural layer of ganglion cells, respectively. In consequence, this means that the "on-centre" bipolar cells emit more glutamate neurotransmitter to the ganglion cells when their photoreceptor cells are illuminated, while the "off-centre" bipolar cells are most excited when their photoreceptors are in darkness. While this mechanism is helpful to encode absolute responses to light intensities, its true significance becomes apparent when the bipolar cells are combined: when arranged next to each other in special formations called receptive fields, this provides the retina with a method to detect edges. The **horizontal cells** support this process by providing a way of lateral inhibition between nearby photoreceptors and their corresponding bipolar cells. This feedback loop helps to amplify and suppress signals and hence increase the contrast between more and less illuminated parts of the perceived image (Goebel et al., 2012, p. 1304).

Rods and cones not only differ in their functionality and distribution across the retina but also in their **connectivity** to the optic nerve. In each eye, there are about 120 million rods and six million cones (Goebel et al., 2012, p. 1303). While each bipolar cell is typically connected to only one or two cones, many rods provide input into the same bipolar cell. This means that cones can provide a very high spatial resolution, while the rods' ability to detect minimal amounts of light is supported by combining their signals (Bruce et al., 2003, p. 29). The bipolar cells are in turn connected to a total of about one million **ganglion cells** whose axons form the optic nerve for each eye. This kind of complex signal convergence through the neural layers of the retina is a characteristic feature of the human visual system (Goebel et al., 2012, p. 1305; Swanston & Wade, 2013, pp. 137-138).

3.3.2 VISUAL PATHWAY AND VISUAL CORTEX

The phototransduction and signal convergence in the retina create an abundant array of data representing the perceived visual stimuli. These data are then passed on to processing areas in the mid-brain through the axons of the ganglion cells that form the optic nerves. The first major processing station on the **primary visual pathway** is the optic chiasm. Here, visual information from both eyes crosses over as shown in Figure 3-11.

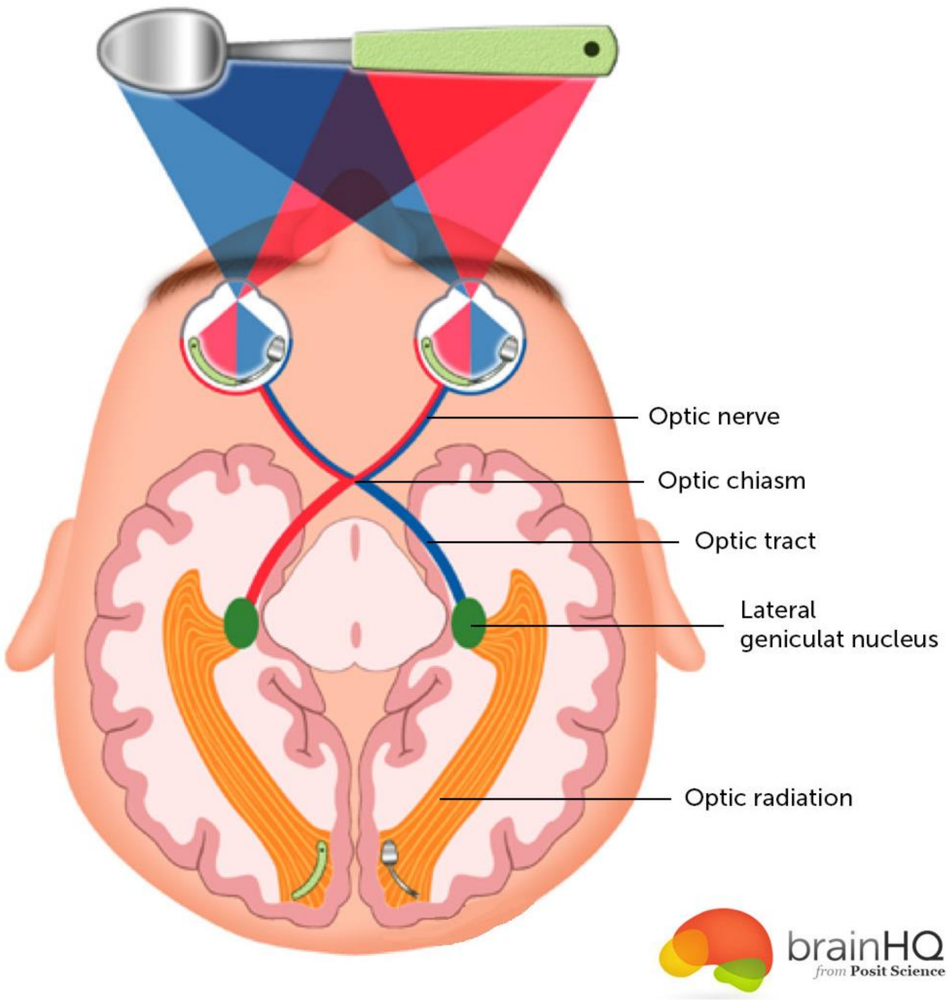


FIGURE 3-11: THE PRIMARY VISUAL PATHWAY AND ITS PROCESSING STATIONS.[15]

When light is refracted upon entering the eye, the light rays are turned upside down and flipped horizontally, resulting in a mirrored inverted projection on the retina (Swanston & Wade, 2013, pp. 111-112). In the **optic chiasm**, the ganglial axons originating in the left nasal retina then cross over towards the right optic tract, while the axons originating in the right nasal retina cross over towards the left optic tract. In contrast, all axons from the temporal areas of the retina stay on their respective sides. As a result of combining the initial refraction in the eyeball and this cross-over mechanism, everything located on one side of the visual field will be contralaterally projected to the other side of the brain through the optic tract and towards the respective lateral geniculate nucleus. The large number of ganglial axons and their prominent placement in the brain facilitated an early discovery of this cross-over mechanism almost 300 years ago (Swanston & Wade, 2013, p. 143).

[15] From Posit Science brainHQ (https://www.brainhq.com/wp-content/uploads/2018/12/visual-pathway.jpg). Reproduced with permission.

The axons of the retinal ganglion cells then synapse onto cells in the two **lateral geniculate nuclei**. Each lateral geniculate nucleus has six well-defined layers that are organised very methodically – layers 1, 4, and 6 are connected to fibres from the contralateral eye, while layers 2, 3, and 5 are connected to fibres from the ipsilateral eye (Swanston & Wade, 2013, p. 144). Furthermore, layers 1 and 2 have distinctively larger cells than layers 3 to 6, and these so-called magno- and parvocellular layers, respectively, are connected to only certain types of ganglion cells (Bruce et al., 2003, pp. 45-47). The axons of the neurons in the lateral geniculate nuclei then form the optic radiation, of which the majority directly project to the primary visual cortex (Goebel et al., 2012, p. 1307). The lateral geniculate nuclei thus serve as a major relay station on the primary visual pathway. However, they have also been found to provide straightforward access to the now neatly organised visual information to other neural processing areas such as those related to attention (Weyand, 2016). Furthermore, neurons in the lateral geniculate nuclei are involved in integrating information about eye movements and eye position, thus supporting the visual cortex in resolving egocentric spatial references (Weyand, 2016).

The final station of the primary visual pathway and the destination of the neural projections from the lateral geniculate nucleus is the **primary visual cortex**. It is located in the calcarine fissure in the occipital lobe of the brain and also known as **V1** or Brodmann area 17 (Brodmann, 1909, pp. 140-142; Goebel et al., 2012, p. 1309). This is the location where, for the first time after being separated at the beginning of the visual pathway, the signals from both retinae are converged again. From here, they are further distributed to higher visual cortices responsible for extracting certain features from the perceived images. The primary visual cortex is, therefore, sometimes described as a gateway to more specialised visual processing areas (Goebel et al., 2012, p. 1309).

V1 is commonly divided into six layers whose cell types and densities differ so visibly that it has also been called **striate**, or striped, **cortex** (Bruce et al., 2003, p. 47). Most connections from the lateral geniculate nucleus arrive in the fourth layer of V1 and from there are distributed further to the other layers (Goebel et al., 2012, p. 1310). The by far most common type of neurons in V1 are pyramidal cells, whose axons project to other cortical and subcortical brain regions and also to the "**extrastriate**" **visual cortex** containing processing areas V2 to V5 (Goebel et al., 2012, pp. 1309, 1313). These areas are specialised in extracting features from the visual information, such as colour composition, shape and orientation, or motion and direction (Goebel et al., 2012, pp. 1313-1316).

These features are then combined again to form conclusions about perceived objects and their locations. Object recognition and localisation are two distinct aspects of visual information processing that are performed by two separate **processing streams** shown in Figure 3-12: a ventral stream passing through the occipitotemporal regions of the brain determines the identity of an object, while a dorsal stream passing through the occipitoparietal regions of the brain determines the object's location (Goebel et al., 2012, p. 1317). In particular, the ventral stream is concerned with identifying an object based on its characteristics ("What" stream), whereas the dorsal stream is involved in resolving spatial relationships between the viewer and the objects ("Where" stream), which influences sensorimotor responses such as trying to grab the object (Goodale & Milner, 1992). Interestingly, a similar distinction of processing streams has been found in the auditory system as described in Section 3.2.2. The implications of this similarity are discussed in more detail when talking about audio-visual stimulus integration in Section 3.4.

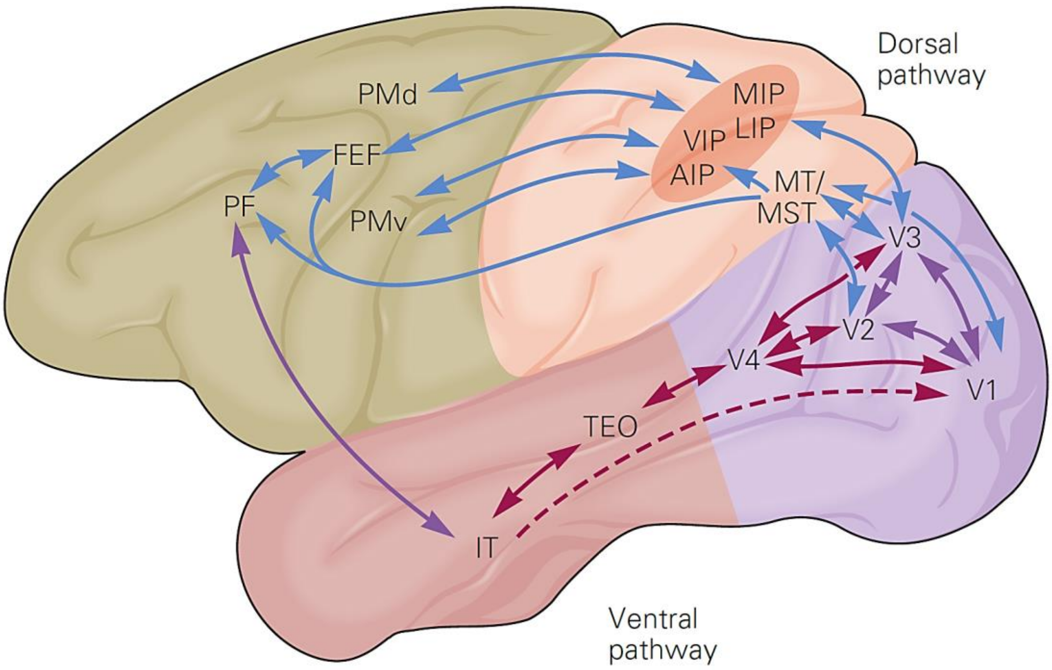


FIGURE 3-12: THE "WHAT" (VENTRAL, RED ARROWS) AND "WHERE" (DORSAL, BLUE ARROWS) PROCESSING STREAMS IN THE VISUAL SYSTEM.[16]

Despite the neural reorganisation along the primary visual pathway, stimuli located next to each other in the original visual scene will still be projected to the primary visual cortex in their original surroundings to preserve important contextual information (Goebel et al., 2012, p. 1316). These arrangements are called **retinotopic maps** (Bruce et al., 2003, p. 43). Studies using functional brain imaging have shown that the primary visual cortex is organised in a way that accurately represents the spatial layout of the original image, albeit with slightly

[16] From *Chapter 25 – The Constructive Nature of Visual Processing* by Gilbert (2013, p. 563). Reproduced with permission. **PF**, prefrontal cortex; **PMd/PMv**, dorsal/ventral premotor cortex; **FEF**, frontal eye field; **MIP/LIP/VIP/AIP**; medial/lateral/ventral/anterior intraparietal area; **MT/MST**, middle temporal area; **IT**, inferior temporal cortex; **TEO**, area of inferior temporal cortex.

distorted features (Dougherty et al., 2003; Polimeni, Fischl, Greve, & Wald, 2010). This means that the primary visual pathway, like its auditory counterpart, tries to retain as much of the original input as possible so that the perceived information can be analysed in full. Apart from the inversion caused by the corneal refraction, the image is also subject to cortical magnification – signals from the foveal regions of the retina are processed in a proportionally larger region of the visual cortex than those of the more peripheral regions (Dougherty et al., 2003; Polimeni et al., 2010). The particularities of retinotopy play a considerable role in the design of the visual processing system presented in this thesis and are described in more detail in Section 5.4.4.

Not all visual signals from the retinae arrive in the lateral geniculate nuclei or the primary visual cortex. Some connections branch off after the optic chiasm to other brain regions, for example, to regulate the circadian rhythm (i.e., sleep-wake-cycle) or to be combined with other sensory information to control the muscle movements of the eyeball (Goebel et al., 2012, p. 1307). In the context of this thesis, an especially interesting aspect is the integration of signals from the auditory system in the visual pathway. Functional brain imaging revealed that auditory information influences activity in the primary visual cortex even if participants were blindfolded (Vetter, Smith, & Muckli, 2014). This suggests that auditory information supports visual perception on an abstract level. The integration of auditory and visual modalities during perception in the brain is discussed in more detail in Section 3.4.

The functioning of the visual system that was briefly described in this section informed the design of the video processing model introduced in Section 5.4. In particular, the functioning of the photoreceptors and the ganglion cells was considered highly relevant for the first step in the model, a retina-inspired video signal transformation. For the rod photoreceptors and peripheral greyscale vision, this is described in Section 5.4.1, while for the cone photoreceptors and foveal colour vision, this is described in Section 5.4.2. The transformed signals were then intended to be mapped into a neural network in a biologically plausible way. For this process, the size and location of the visual cortex were determined as described in Sections 5.4.3 and 5.4.4 so that insights from retinotopy studies could be applied. The visual pathway, which was found to function mainly as a relay station, was not included in the video processing model at this stage for computational simplicity. However, any signal distortions that occurred were captured and modelled. Finally, a novel process of signal compression was developed based on the theory of receptive fields. This is described in Section 5.4.5.

3.4 AUDIO-VISUAL PROCESSING AND INTEGRATION

Multimodal integration is a vital part of different processing stages in the brain in order to fully assess one's surroundings and plan appropriate responses. This section looks at both subcortical and cortical combination of different sensory inputs with a focus on how auditory and visual signals influence each other. While some more general principles discussed below have been well established in the neuroscientific research community, the majority of how these processes function is still under active investigation. The content presented in this section has, therefore, been chosen to illustrate certain aspects in more detail than others, as was deemed necessary to provide a comprehensive background for the work presented in this thesis. The computational framework that was developed based on these findings is presented in detail in Section 5.5.

It is generally assumed that both the auditory and the visual pathway enhance their processing capabilities by incorporating information from other sensory modalities *before* they arrive at their respective cortices (A. K. C. Lee & Wallace, 2019). For the auditory pathway, it has repeatedly been shown that nonauditory signals like visual or somatosensory inputs influence auditory responses in all processing stations from the cochlea to the primary auditory cortex (Atilgan et al., 2018; King, Hammond-Kenny, & Nodal, 2019). Especially the inferior colliculus in the midbrain has been found to be involved in the integration of other sensory modalities like sensorimotor or visual cues (Gruters & Groh, 2012; von Kriegstein, Patterson, et al., 2008). On the other hand, recent evidence suggests that visual cues are in fact *not* integrated at subcortical levels of the auditory pathway, and that measurement results indicating otherwise may reflect the neural activity of descending pathways (Caron-Desrochers, Schönwiesner, Focke, & Lehmann, 2018). The exact mechanisms of subcortical audio-visual integration are hence still actively debated.

From a more applied research perspective, studies in the area of **speech processing** have shown that spoken word recognition is enhanced by visual signals related to facial cues such as lip movement (Campbell, 2008; I. R. Olson, Gatenby, & Gore, 2002). This enhancement even stays in effect when the visual signals are not perceived anymore, indicating that the visual information belonging to an auditory stimulus is learned at some point (Bruns & Röder, 2019; von Kriegstein, Dogan, et al., 2008). Visual cues can also help a listener to better focus on an auditory object of interest (A. K. C. Lee, Maddox, & Bizley, 2019). In sign language users, this cross-modal crossover is even more pronounced. Brain imaging studies with deaf participants showed neural activity in their auditory cortices when presented with visual stimuli in general (Finney, Fine, & Dobkins, 2001) and sign language in particular

(Lambertz, Gizewski, de Greiff, & Forsting, 2005). These findings suggest a certain level of cross-modal cortical reorganisation that differs from person to person, to the extent that it impedes some deaf people's ability to use cochlear implants (Doucet, Bergeron, Lassonde, Ferron, & Lepore, 2006).

For **visual processing**, several studies have shown that the primary visual cortex (V1) receives abstract auditory information that enhances the visual perception of stimuli (Vetter et al., 2014; Vroomen & Gelder, 2000). Activation of V1 could even be shown in participants who were blindfolded and listened to sound that they learned to associate with a visual stimulus beforehand (Petro, Paton, & Muckli, 2017). This indicates that the learning process of visual information described for the auditory pathway applies equivalently to the presentation of auditory stimuli to the visual pathway. It has further been found that visual size perception can be influenced by varying auditory stimuli (Tonelli, Cuturi, & Gori, 2017).

But how does this cross-modal signal integration work, given that the ways in which auditory and visual stimuli are transformed into neural activity by the cochlea and the retina are so different? While auditory spatial cues are encoded in a head-centred **reference frame** or coordinate system, visual spatial information is based on an eye-centred frame. As a result, the brain has to combine these reference frames into a new common "hybrid" frame by taking into account the relative position of the eyes and the head (Willett, Groh, & Maddox, 2019). Over 30 years ago, experiments showed that both auditory and visual signals must be transformed into a common coordinate system in or before the superior colliculus which in the visual pathway is responsible for creating saccadic eye movements (Jay & Sparks, 1987). More recently, it has been suggested that a potential location in the auditory pathway for this reference frame resolution could be the inferior colliculus, based on evidence in monkeys that showed that auditory responses in this region were influenced by eye position (Groh, Trause, Underhill, Clark, & Inati, 2001). However, it has also been found that the posterior parietal cortex, which is organised with respect to action planning, can resolve different reference frames of visual eye-centred and auditory head-centred signals into a common coordinate frame (Andersen, Snyder, Batista, Buneo, & Cohen, 1998). According to the latest review, it is most likely that several processing stations along both pathways are involved in reference frame resolution (Willett et al., 2019).

The underlying reason for the reference framework mismatches lies in the transformation of the **representational format** of signals from different modalities. While the visual and somatosensory senses can quite accurately represent locations and position information in their environment using maps, the auditory pathway has to compute this information in the superior olivary complex as described in Section 3.2.2 (Porter & Groh, 2006). This means that the visual, somatosensory, and motor cortex are based on a representation using spatial and receptive fields, while the auditory cortex uses a neural rate coding in which the firing rates of the neurons are proportional to the azimuth of a stimulus (J. Lee & Groh, 2014). Moreover, the sensory inputs from both auditory and visual sensory inputs have been shown to be "weighted" by the brain based on their expected value for identifying certain aspects of a stimulus, since the auditory system has a better temporal resolution and the visual system has a better spatial resolution (Alais & Burr, 2019). For the work presented in this thesis, this means that a similar way of overcoming the differences in the representational format of the signals has to be found. This is formulated and discussed in detail in Section 5.5.

Another aspect that has to be considered when integrating signals from different modalities is the **timing** of the stimuli. Since stimuli that are in close spatial and temporal proximity can generally be considered connected or even originating from the same source, the brain needs to process these stimuli in correspondence. It has been found that perceiving such stimuli that are close to each other results in enhanced neural responses (A. K. C. Lee & Wallace, 2019). In processing these stimuli, the brain is thought to perform a temporal alignment of signals that compensates for sound waves travelling more slowly than light rays (Burr & Alais, 2006). In contrast to the stimulus presentation times, auditory signals are in fact processed faster than visual information. Based on a study with three epilepsy patients who had electrodes implanted into their superior parietal lobule, a response to auditory stimuli was measured at around 30 milliseconds after stimulus onset, while a response to visual stimuli was measured at around 75 milliseconds after stimulus onset (Molholm et al., 2006). In the same study, the integration of audio-visual signals was measured at between 120 and 160 milliseconds after stimulus onset. This indicates that the brain matches the timing of different sensory stimuli to gain the best possible knowledge from the signals. A discussion of how these timing issues influenced the development of the audio-visual model can be found in Section 5.5.2. In particular, these findings informed stimulus presentation times and temporal alignment of different signal sampling rates for sound and frame rates for videos.

Going beyond the subcortical processing of sensory stimuli, the signals are sent from their respective cortices to higher, more conceptual processing areas. As mentioned in both Section 3.2.2 and Section 3.3.2 when describing the workings of the auditory and visual cortices, respectively, the brain has developed two distinct **cortical processing streams** to extract different types of information. In particular, a ventral stream through the temporal lobes is responsible for object recognition, while a dorsal stream through the parietal lobe has been identified to determine the location of perceived objects. This phenomenon was first described for the visual modality by Goodale and Milner (1992), who at the time called these the streams for perception and action, respectively. The authors later revisited their hypothesis and found that a large amount of evidence had amassed supporting their theory (Milner & Goodale, 2008), for example from brain imaging studies using positron emission tomography to localise these streams (Ungerleider & Haxby, 1994). Similar findings have been made for the auditory system, for example by examining the brains of monkeys (Rauschecker & Tian, 2000), but also more recently in humans (Rauschecker, 2015).

Figure 3-13 shows the two processing streams in a diagram. The neural signals originate in the sensory processing areas, which in this diagram are the auditory, visual, and somatosensory cortices, from where they split up to travel to either the temporal or the parietal association cortices. They are then sent to the prefrontal association cortex, where they are for example combined with emotional responses and create an active thought process (C. R. Olson & Colby, 2013).

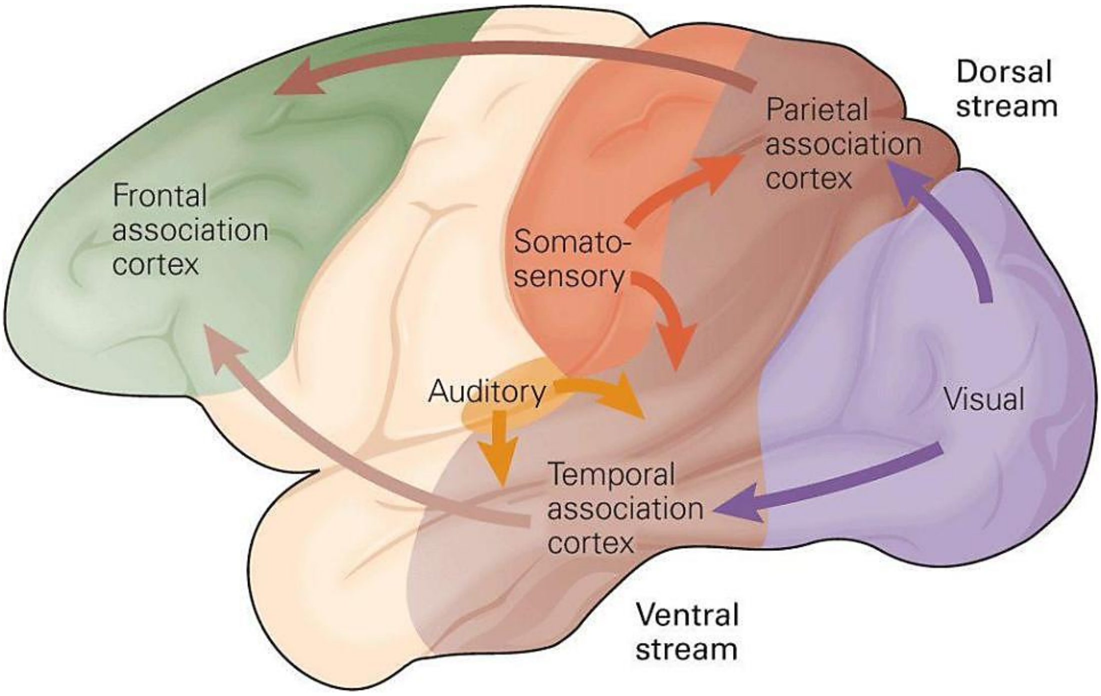


FIGURE 3-13: THE "WHAT" (THROUGH TEMPORAL CORTEX) AND "WHERE" (THROUGH PARIETAL CORTEX) PROCESSING STREAMS IN THE BRAIN.[17]

It has been found that some neurons in the **prefrontal cortex**, especially in its ventrolateral segment, respond to multisensory stimuli to improve communication (Romanski, 2007) and that the frontal lobes converge audio-visual inputs to combine face and vocalisation stimuli for the same reason (Plakke & Romanski, 2019). In general, evidence from a large number of brain-imaging studies shows that signal processing is more concrete and feature-based

[17] From *Chapter 18 – The Organization of Cognition* by C. R. Olson and Colby (2013, p. 397). Reproduced with permission.

near those brain regions that perform stimulus input and transformation and more abstract and concept-based in higher cortical levels (Taylor, Hobbs, Burroni, & Siegelmann, 2015). This applies to all modalities – auditory, visual somatosensory, gustatory, olfactory, and interoceptive.

The following paragraphs detail the two processing streams, with a focus on where and how audio-visual integration might happen along them.

The **ventral processing stream** is concerned with the recognition of objects from all sensory modalities and hence contains information about the "objecthood" of a perceived item (Kubovy & van Valkenburg, 2001). Several studies, usually employing brain-imaging methods like fMRI, have attempted to pinpoint the exact locations of sensory integration in the ventral processing stream. For example, one study by Spitsyna, Warren, Scott, Turkheimer, and Wise (2006) looked at auditory and visual signals during a language comprehension task of spoken and written narratives and found a common neural activation in the left anterior temporal cortex as well as at the junction of the temporal, occipital, and parietal cortices. Another fMRI study attempted to identify the conceptual representation of two categories of objects (animals and tools) and showed an integration site in the posterior superior temporal sulcus and the middle temporal gyrus for both within and across auditory and visual modalities (Beauchamp, Lee, Argall, & Martin, 2004). A third study that measured cross-modal integration identified the left superior temporal sulcus as a major processing area (Calvert, Campbell, & Brammer, 2000). Thus, strong evidence exists about the location and purpose of this processing stream, being mindful of the fact that language seems to have an inherent tendency to be processed mainly in the left temporal lobe (Wilson, Bautista, & McCarron, 2018).

The **dorsal processing stream** is concerned with the localisation of objects in space and in relation to the observer to potentially generate movement, which is facilitated by it passing through the sensorimotor cortex in the parietal lobe (Rauschecker, 2015, 2018). As with the ventral processing stream, several brain imaging and behavioural studies have revealed how the dorsal stream processes information. For example, fMRI data recorded during a motion discrimination task with auditory and visual stimuli showed neural responses in the intraparietal sulcus of the parietal cortex and that both modalities can enhance and suppress each other's activity (Lewis, Beauchamp, & DeYoe, 2000). A recent review by Chaplin, Rosa, and Lui (2018) compiled a wealth of evidence that the integration of auditory and visual signals related to movement and location is processed in the posterior parietal cortex and – interestingly – in the superior temporal sulcus, which is usually ascribed to the realm of the ventral processing stream. However, as mentioned at the beginning of this chapter, this topic is an active area of research, so some inconsistencies can be expected. Independent of the processing region but related to spatial processing, it has also been found that combined audio-visual stimuli improve the relative localisation of objects over unisensory auditory or visual perception (Freeman, Wood, & Bizley, 2018). The evidence further suggests that vision is seen as more reliable by the brain than other senses when it comes to object localisation, as is shown by its dominance over auditory spatial information (Witten & Knudsen, 2005) and during motor reaching tasks (Glazebrook, Welsh, & Tremblay, 2016).

The integration of multisensory signals with a focus on auditory and visual information informs the design of the combined audio-visual processing system described in Section 5.5. In particular, Section 5.5.1 looks at how both sensory modalities were both entered into a single brain-shaped neural network using the learnings from the unisensory auditory and visual models. Section 5.5.2 focuses on signal timing to ensure a biologically plausible combination of presented stimuli. While the two cortical processing streams were not directly implemented in the model, it was an interesting area of exploration to investigate if the general connectivity in the brain-inspired network had the ability to form specialised processing streams.

3.5 CHAPTER SUMMARY

This chapter explained how the human eyes and ears collect and translate information about their surroundings, and how the brain channels and interprets these signals to extract useful information. These findings form the basis of the biologically inspired computational model presented in this thesis and support answering some of the research questions raised in Section 1.3. While the highly specialised biological processes include a myriad of intricate steps and components, it was impossible for the work presented in this thesis to cover this amount of detail. Prior theoretical knowledge that could support the choice of which biological processes to model in the proposed computational system was scarce as described in Chapter 4. The purpose of Chapter 3 was, therefore, not only to provide a background on the relevant biology, but also to assess which aspects might be a useful starting point for the system design.

Considering that, computationally, spiking neural networks were a focus of the work presented in this thesis, the two aspects that were deemed a good starting point for this work were the transformation of real-world data into spikes and the mapping of these spikes into the network architecture. Chapter 5 explains in detail the employed methods of cochlear and retinal encoding that created signals similar to those present in the brain. In nature, these signals are then propagated through the brain into highly specialised processing areas. Information perceived by the ears is primarily processed in the auditory cortices, which are located in both temporal lobes of the brain, and signals from the eye are mainly processed by the visual cortex, which is located in the occipital lobe of the brain. In the computational model developed here, this process is simulated using tonotopic and retinotopic maps to enter the signals in a neurologically plausible way into their equivalent locations in a three-dimensional brain-shaped artificial neural network. In the biological brain, the signals then travel through two designated cortical processing streams that combine multimodal information to extract the identity and the location of a perceived object, respectively. From there, they are further propagated to the frontal cortex or other regions of the brain depending on the nature of the environmental stimuli. The result of this sophisticated procedure defines our thoughts, emotions, conclusions, ideas, and creativity. Since the computational model is in the early stages of system design and very much exploring boundaries, it can not promise such things as of yet. However, it is hoped that by applying it to several different real-world problems (as in Chapters 6, 7, and 8) a potential for further development can be shown.

4 EXISTING COMPUTATIONAL METHODS

"Crime is common. Logic is rare. Therefore, it is upon logic rather than upon the crime that you should dwell."

– Sherlock Holmes in The Adventure of the Copper Beeches

4.1 CHAPTER OVERVIEW

This chapter presents an overview of existing brain-inspired computational methods that can be used for auditory and visual processing. While seminal works of more "traditional" approaches are also mentioned briefly, the focus of the selected literature was on methods that simulated neural mechanisms or claimed biological plausibility. Therefore, this review does not cover in depth currently popular machine learning and deep learning methods such as convolutional neural networks, in favour of emphasising the more biologically plausible group of spiking neural networks. Due to their proximity to the brain as a blueprint for a "supercomputer", brain-inspired methods can be expected to enhance the processing capabilities of computational architectures, as has, for example, been argued by Hassabis, Kumaran, Summerfield, and Botvinick (2017). A variety of biologically inspired "cognitive architectures" has been developed in the past few decades, many of which consist of several sub-systems with distinct tasks that are working together like the processing units of the brain (Goertzel, Lian, Arel, de Garis, & Chen, 2010).

In 2021, researchers can choose from a range of software packages for brain simulations that can be used for neural modelling and computational analysis (Tikidji-Hamburyan, Narayana, Bozkus, & El-Ghazawi, 2017). Artificial neural networks as novel machine learning methods and their arguably most brain-like variant, the SNN (Maass, 1997), provide a versatile (Hong et al., 2020) and promising (Fong, Scheirer, & Cox, 2018) approach to guide computational data analysis. These developments together with the thesis author's intrinsic motivation to unlock the secrets of the brain for novel computational processing approaches spurred the direction of this chapter towards a focus on nature-inspired and biologically plausible methods.

The overview of the computational background given in this chapter informed the design of the system presented in this thesis. Thematically, Section 4.2 looks at methods for sound processing, Section 4.3 looks at video processing approaches, and Section 4.4 explores architectures that combine auditory and visual data. Within this context, the aim of this chapter was to provide an overview of methods irrespective of specific datasets and applications. Literature related to the tasks of speech and video gesture processing is discussed in detail in Chapters 6 and 7, respectively.

4.2 SOUND PROCESSING

The field of sound processing, as it is understood in the context of this thesis, encompasses problems involving auditory data, such as speech recognition, environmental sound classification, music analysis, and sound source separation. Since sound stimuli consist of a mixture of sound waves, processing methods typically look at and try to interpret spectral information. Conventionally, using Mel-frequency cepstral coefficients (MFCC), sound signals can be decomposed into separate frequencies based on **perceptual features** (Zheng, Zhang, & Song, 2001), which can then be analysed individually, for example for speech (Vergin, O'Shaughnessy, & Farhat, 1999) and speaker recognition (Sahidullah & Saha, 2012) or musical genre classification (Tzanetakis & Cook, 2002). With the advent of Deep Learning and Convolutional Neural Networks (CNN) (LeCun et al., 2015; Schmidhuber, 2015), another means of interpreting sound data through **spectrograms** became feasible (Costa, Oliveira, & Silla, 2017; Dörfler, Bammer, & Grill, 2017; Ren et al., 2018; Tjandra et al., 2015; H. Zhang, McLoughlin, & Song, 2015). In this approach, a sound is first converted to an image representation of its spectral and temporal components and then classified using established image recognition algorithms. Other methods to prepare sound files for classification include linear transformation (Gales, 1998), adaptive segmentation (J. X. Zhang, Brooks, & Whalley, 2009) or stochastic mapping (Afify, Cui, & Gao, 2009).

A variety of algorithms (Hinton et al., 2012) can then be used to analyse the pre-processed sound files for their respective application task. Historically, **Hidden Markov Models** (HMM) (Gales & Young, 2008; Rabiner, Wilpon, & Soong, 1989; Wilpon, Rabiner, Lee, & Goldman, 1990) and **Support Vector Machines** (SVM) (Ganapathiraju, Hamaker, & Picone, 2004; Lu, Zhang, & Li, 2003; Wan & Campbell, 2000) were commonly used for speech and speaker recognition. More recently, methods involving **Deep Learning and CNN** have shown promising results for speech and speaker recognition (Abdel-Hamid et al., 2014; L. Deng, Hinton, & Kingsbury, 2013; Sainath et al., 2015; Sainath, Mohamed, Kingsbury, & Ramabhadran, 2013; H. Zhang et al., 2015), and also for environmental sounds (Boddapati, Petef, Rasmusson, & Lundberg, 2017; Çakır, Adavanne, Parascandolo, Drossos, & Virtanen, 2017; Piczak, 2015; Salamon & Bello, 2017; Tokozume & Harada, 2017) and music classification (Choi, Fazekas, Sandler, & Cho, 2017; Costa et al., 2017).

Most noteworthy within the context of this thesis are algorithms and models that claim **biological inspiration and plausibility**. Therefore, the remainder of this section will look at (1) algorithms for sound transformation that are based on the functioning of the cochlea and (2) computational sound processing systems that make use of an SNN or its derivates.

The first computational model that was based on the functioning of the cochlea and received wide acclaim was developed about 40 years ago (Lyon, 1982). The **Lyon model** tries to simulate the physiological behaviour of the cochlea by employing a half-wave detection function in combination with a compression algorithm that is based on automatic gain control used in electronic systems. It focuses on mimicking biological behaviour instead of modelling biological processes (Lyon, Rehn, Bengio, Walters, & Chechik, 2010). In later work, Lyon and Dyer (1986) thoroughly evaluated this model and investigated its biological plausibility, its properties and optimum parameters. The model is still used in speech recognition systems today (Schrauwen, D' Haene, Verstraeten, & Van Campenhout, 2007; Verstraeten, Schrauwen, Stroobandt, & Van Campenhout, 2005; Yong Zhang, Li, Jin, & Choe, 2015). More recently, Reimann (2011) also presented a way to formalise cochlear equations; this work put more emphasis on simulating the functioning of the cochlea's components. Although Reimann's work remained largely theoretical, it is directly based on biological processes and confirms Greenwood's (1990) experimental observations of stimulus-response measurements in mammalian cochleae.

However, these models can not be used directly for biologically inspired machine learning applications since they do not output their results in a spike-based format. Modelled after how neurons communicate in the brain, spikes are the most brain-like form of information transmission in computational systems and it has been shown that a combination of spike count and mean response times capture auditory stimulus information very well (Nelken, Chechik, Mrsic-Flogel, King, & Schnupp, 2005). Three such models that can **transform sound into spikes** to be entered into an SNN are compared by Rudnicki, Schoppe, Isik, Völk, and Hemmert (2015): the Holmberg, Gelbart, and Hemmert (2007) model, the Meddis et al. (2013) model, and the Zilany, Bruce, and Carney (2014) model. While the three models differ in the amount of emphasis they put on certain details of their cochlear models, they all modelled the auditory periphery reasonably closely to what is observed in humans. Furthermore, through the wrapper module provided by Rudnicki et al. (2015), these models are easily accessible through a unified interface that converts given sound files to spike files, which can be processed further by an SNN. Another auditory model that interfaces with an existing neural computation paradigm is the Brian Hears module (Fontaine, Goodman, Benichoux, & Brette, 2011) which interfaces with the well-known Brian SNN simulator (Goodman & Brette, 2008). This module is optimised for speed and parallelisation while trying to remain biologically plausible. It makes use of vectorisation and filter banks to facilitate these optimisations and mainly aims at providing algorithms to simulate cochlear and auditory models.

The last set of models for sound transformation that are related to the work presented in this thesis are biologically inspired hardware systems. These **silicon cochlea chips** capture sounds through microphones and produce spike-based output that can, for example, be entered into neuromorphic hardware. Examples for this category are the AEREAR binaural silicon cochlea (S.-C. Liu, van Schaik, Minch, & Delbrück, 2014; M. Yang, Chien, Delbrück, & Liu, 2016) or the binaural neuromorphic auditory sensor for Field-Programmable Gate Arrays (Jiménez-Fernández et al., 2017). Both approaches employ pulse-frequency modulation to capture the sounds and then create spikes as output using the **Address Event Representation** (AER) format. AER is a communication protocol commonly used in neuromorphic systems where each processing unit is identified by a unique address from which it notifies all other units when it observes an event. The most notable difference between these hardware architectures and the algorithms introduced earlier in this section is that spikes are emitted asynchronously as events happen and that hardware systems are optimised for speed and low energy consumption. Neuromorphic auditory chips have been used widely in systems where real-time processing is required, typically providing input to other neuromorphic processing boards (Dominguez-Morales et al., 2016) and SNN architectures derived from such boards (Dominguez-Morales, Liu, et al., 2018) but also to SVM (Abdollahi & Liu, 2011) and CNN (Dominguez-Morales, Jiménez-Fernández, Domínguez-Morales, & Jiménez-Moreno, 2018) systems.

Apart from inspiring how to encode auditory stimuli in a biologically meaningful way, the mechanisms with which the brain and the auditory cortices handle sound can also be incorporated into other parts of **computational sound processing architectures**. Similar to more general brain-inspired methods, it has been suggested that the design of such architectures could benefit from mimicking observations from biology (Klein, König, & Körding, 2003). One example where SNN have been successfully applied due to their special architecture is the field of sound localisation (Glackin, Wall, McGinnity, Maguire, & McDaid, 2010; Goodman & Brette, 2010; J. Liu, Perez-Gonzalez, Rees, Erwin, & Wermter, 2010; Voutsas & Adamy, 2007; Wall, McDaid, Maguire, & McGinnity, 2012). The asynchronous occurrence of the spikes facilitates precise time-based calculations, which are also used by the brain to determine the location of an auditory stimulus (Grothe et al., 2010). More generally, the asynchronous nature of SNN makes them particularly well-suited to process inherently dynamic sound data since they can capture temporal relationships in the signals (Tavanaei & Maida, 2017b; Voutsas, Langner, Adamy, & Ochse, 2005; Jibin Wu, Chua, & Li, 2018; Jibin Wu, Chua, Zhang, Li, & Tan, 2018; Jibin Wu, Yılmaz, Zhang, Li, & Tan, 2020). However, not much is mentioned in contemporary literature about how exactly the features extracted from sound stimuli are entered into the respective network. It seems that, usually, the data is simply vectorised and then fed into a one-dimensional input layer of the network. Relevant papers do not seem to devote much critique (or words) to this process.

The seeming absence of research activity in this area sparked the question as to whether there was a more **biologically plausible way to map the data into the network**. One preliminary piece of work in this area (Saraceno, 2017, p. 39) used a mapping approach that rudimentarily followed the principle of tonotopic mapping in the auditory cortices. Developed under the supervision of the thesis author, this approach sees input neurons arranged in the same region of the network where Brodmann areas 41, 42, and 22 (primary auditory cortex) would be in the brain. The locations for the input neurons were manually selected based on the frequency to which they were tuned, with lower frequencies in the more frontal parts and higher frequencies in the more occipital regions of the "auditory cortex". There were 20 input neurons for each side of the network, enabling bilateral processing. The first experiments on classifying parts of three musical pieces showed promising results. This work served as an inspiration for the development of the sound processing system presented in Section 5.3.

4.3 VIDEO PROCESSING

The field of video processing, as it is understood in the context of this thesis, encompasses problems involving *dynamic* visual data, such as video labelling, event detection, or gesture and action recognition. Computational visual object recognition has been a prolific research field for over half a century (Andreopoulos & Tsotsos, 2013). **CNN** have seemingly become the method of choice for visual processing in the last decade (Ji, Xu, Yang, & Yu, 2013; Kriegeskorte, 2015; Rawat & Wang, 2017), spurred by the surprising success of such a network in a popular image recognition challenge (Krizhevsky et al., 2012). With the growing popularity of **SNN**, a myriad of papers have also looked at processing images, i.e., static visual data, with various SNN architectures and report impressive results (Beyeler, Dutt, & Krichmar, 2013; Iakymchuk, Rosado-Muñoz, Guerrero-Martínez, Bataller-Mompeán, & Francés-Víllora, 2015; Kerr, McGinnity, Coleman, & Clogenson, 2015; Kheradpisheh, Ganjtabesh, Thorpe, & Masquelier, 2018; C. Lee, Srinivasan, Panda, & Roy, 2019; Q. Yu, Tang, Tan, & Li, 2013; Y. Zeng, Zhang, & Xu, 2017). However, both the nature of the spiking behaviour and the feature detection mechanisms found in biological retinae are inherently dynamic. Therefore, this overview of related literature focused on systems that used dynamic data in a more biologically plausible way.

Dynamic visual data are comprised of a collection of changing light rays, which can be captured and transformed into spikes for computational analysis and feature detection. Several approaches attempted to perform this transformation by mimicking and modelling the retina's functionality and behaviour. For example, Wohrer and Kornprobst (2009) propose a **retinal simulation software** that aims to be biologically accurate. Their system transforms video sequences into spike trains by modelling the functional behaviour of retinal ganglion cells. An extension of this work (Cessac et al., 2017) enhances this algorithm by adding lateral connectivity between these ganglion cells, which simulates the behaviour of amacrine cells, and also provides a graphical user interface for the earlier command-line tool. A similar, but less detailed approach to modelling retinal ganglion cells was also suggested by Vance et al. (2018). By comparing their resulting spike trains to measurement data from experiments with human and mammal retinae, the authors of these studies found that their algorithms provided a biologically plausible output for the given stimuli.

Other biologically inspired **video encoding algorithms** tended to focus on one select aspect of retinal processing. For example, Gütig, Gollisch, Sompolinsky, and Meister (2013) experimented with varying spike times and found that most of the information about the stimulus was encoded in the precise temporal patterns of the spikes. This finding further supported the assumption that computational visual processing systems must retain the **dynamic aspects of the input data** to become more biologically plausible. Another interesting finding in this regard was the work of Akbas and Eckstein (2017), who showed that a system with a moving foveal area with high spatial resolution and surrounding areas with lower resolution can match the performance of a system in which the whole visual field is sampled with high spatial accuracy. However, the former system, like the eye, is more energy-efficient. Interestingly, most works found in the literature do not seem to devote specific attention to the **colours** of the observed light rays, but instead mainly look at changes in light intensity within the visual field (to detect edges) and over time (to detect movement). One exception to this observation was the work by Rafegas and Vanrell (2018), who described a multi-layered CNN that was trained to detect features in coloured images of objects. Layer by layer, the network's neurons developed a selectivity to colour axes, characteristically coloured objects or backgrounds, and contrasted colours of objects and backgrounds. These findings in combination with its importance in human vision (Gerl & Morris, 2008) suggest that colour can be assumed to also play an important role in computational object recognition systems.

The third category of biologically plausible systems for transforming visual data into spikes shifted their attention to **later stages of the visual pathway** by modelling the functionality of the primary visual cortex, V1. For example, Lian, Grayden, Kameneva, Meffin, and Burkitt (2019) created a biologically inspired model of the pathway itself, modelling the interactions between V1 and its preceding station (the lateral geniculate nucleus). The authors evaluated their system on image data and observed the emergence of receptive fields and orientation tuning within their neural representation. In another study, H. Liu, Shu, Tang, and Zhang (2018) used a layered SNN to mimic the behaviour of V1 in conjunction with the middle temporal cortex for motion detection to create feature vectors representing human actions. These feature vectors were then classified using an SVM. As part of their work, the researchers also presented a biologically inspired approach for modelling surround suppression, which is an important feature of retinal ganglion cells and V1 that has been observed in nature. Other works in this category have made use of the cognitive processes in the visual apparatus as a blueprint for semantic feature extraction and concept formation (Yin et al., 2018) or mimicked the hierarchical organisation of V1 to increase the efficiency of their system when learning new visual concepts (Rule & Riesenhuber, 2021).

Finally, visual data can be transformed into spikes using specialised **neuromorphic hardware**. One of the systems that received widespread attention in this area was the Dynamic Vision Sensor (DVS) developed by Lichtsteiner, Posch, and Delbrück (2008). This event-based camera does not capture a visual scene frame by frame like a conventional camera would, but instead detects changes in brightness in the visual field over time and transforms them into spikes using the AER format. This means that the origin of the "brightness change" event can be traced back to a unique location on the sensor and, hence, in the visual field. Further development of these sensors means that they are nowadays also capable of colour (Berner & Delbrück, 2011) and stereo vision (Domínguez-Morales et al., 2019).

The described algorithms and systems can transform visual data into spikes using biologically inspired approaches. Therefore, the next logical step would be to **process these spikes further** in a biologically plausible way. However, the majority of the works mentioned so far in this section do not employ such methods but rather input their spikes into more static architectures like SVMs or layer-based CNN. Notable exceptions to this practice are the architectures by Hadjiivanov (2016) and Hopkins, Pineda-García, Bogdan, and Furber (2018). The system presented by Hadjiivanov (2016) includes a new encoding algorithm that mimics retinal ganglion cells and Gaussian receptive fields, and the resulting spikes are then also processed in a simple SNN to evaluate the algorithm's performance. As a second example, Hopkins et al. (2018) proposed a **fully integrated hardware setup** where spikes from an event-based camera are processed with a neuromorphic SNN architecture, the SpiNNaker chip (Furber, Galluppi, Temple, & Plana, 2014). Several papers have also described software-based **simulations of DVS** cameras where the spikes are then entered into an SNN, with varying application data but generally good results (George, Banerjee, Dey, Mukherjee, & Balamurali, 2020; D. Liu, Bellotto, & Yue, 2020; Q. Liu et al., 2020; J. Shen, Zhao, Liu, & Wang, 2020). Of these models, Q. Liu et al. (2020) further propose a new biologically inspired feature extraction algorithm that is based on the functioning of V1. These features are then entered into the SNN for classification.

However, like for the sound processing systems described in Section 4.2, exactly how the signals are **mapped into the network** is rarely presented as a matter of thorough investigation. In the majority of the works, the extracted features and spikes are simply entered into the first layer of the network as a linear feature vector. One contrary example is the architecture presented by Ge, Liang, Yuan, and Thalmann (2019). In their experiments, the researchers analysed hand poses by spatially mapping the captured **depth data** of those hand poses into a three-dimensional CNN. They then extracted features from the modified network states to estimate the original hand pose, with good results. Another example is the work by Paulun et al. (2018), which employs a **retinotopic mapping** approach into a three-dimensional SNN architecture that is shaped like V1. Developed under the supervision of the thesis author, this mapping approach arranges the input neurons that receive visual stimuli from a DVS simulator in the same way in which visual signals are mapped in the brain. The mapping also follows the mirroring and distortion of these signals along the visual pathway based on the origin point of a signal in the visual field. As part of the research presented in this thesis, the Paulun approach was studied and improved further, as presented in Section 5.4.

4.4 COMBINING AUDITORY AND VISUAL DATA

This section explores biologically inspired computational architectures in which auditory and visual data were combined and analysed together. As explained in Section 3.4, observations from neuroscience suggest that the brain integrates these modalities (and those perceived by other senses) along two main processing streams that look at object identification and localisation. Furthermore, an increased recognition rate was observed in the brain when multiple modalities were combined. This raises the question as to whether the same can be expected for computational systems that follow a similar data combination approach.

Several approaches to signal fusion have emerged in the space of audio-visual signal processing architectures, with varying success (Abdelaziz, 2018). In the **feature fusion** approach, extracted features are simply concatenated and then fed into a processing unit together. One early example of this approach is the work by Roy and Pentland (2002), who estimated the probability of a sound representing a certain phoneme using recurrent neural networks and then combined these probabilities with histograms representing visual features. In more recent literature, Arandjelovic and Zisserman (2017) introduce a large-scale unsupervised learning method that receives input from concatenated vectors in fusion networks. Features from corresponding sound and video data were first extracted separately each using a CNN, where the sound was converted to spectrograms and videos were divided into frames. Another example of feature fusion is the work on audio-visual speech detection by Thermos and Potamianos (2016). Sound signals are first processed with the MFCC algorithm and the resulting features are then combined with depth data from videos in an SVM.

A second approach of merging modalities is **decision fusion**, in which the pre-processing methods typically contain multiple steps that are assessed and compared for their informational value before a decision is made about which predicted label should be retained. For example, Cruz, Parisi, Twiefel, and Wermter (2016) combine speech and gesture data in this way to train robots with a neural network-based associative architecture and reinforcement learning. Their network grows dynamically with each input that it receives from the pre-processing streams. In another example, Poria, Cambria, Howard, Huang, and Hussain (2016) apply a mixture of feature and decision fusion to the task of extracting sentiments from multimodal web videos, with large improvements in classification accuracy when compared to existing methods.

While these two fusion methods performed well in the past and remain popular, a recent benchmarking study found that another approach, the **multi-stream HMM**, achieve even better recognition rates (Abdelaziz, 2018). First described in detail by Dupont and Luettin (2000), multi-stream HMM contain several input streams that can work independently and in parallel up to a pre-defined point where the signals are synchronised and combined. The input streams can then automatically align their internal temporal characteristics to match this pre-defined anchor point. This method has, for example, been successfully applied by Noda, Yamaguchi, Nakadai, Okuno, and Ogata (2015) to the problem of spoken word recognition where the speakers were also filmed. In their architecture, sound signals are processed using a deep denoising autoencoder, while visual features are extracted by a CNN. Their two encoding systems were trained and optimised separately and then integrated in a multi-stream HMM for analysis and classification.

More recently developed methods for combing auditory and visual signals typically involve architectures that are inspired by **neural processing**. For example, Huang and Kingsbury (2013) introduced a novel feature fusion method that is based on deep belief networks, which are composed of several simple learning modules containing layers of neurons (Hinton, 2009). The researchers show that their system outperforms existing fusion methods such as multi-stream HMM but requires more training data. This was also observed in a second approach (S. Zhang, Zhang, Huang, Gao, & Tian, 2018) that used deep belief networks to combine features extracted from CNN for emotion recognition. In this work, sound features were extracted using a standard CNN, while video data were processed in a three-dimensional CNN. After being combined by the deep belief network, the stimuli were classified with an SVM with promising results.

Further to these examples of neural-network-based methods, several approaches for audio-visual processing have been developed from **biological inspiration**. Early work in this area (Kasabov, Postma, & van den Herik, 2000) attempted to simulate the hierarchical nature of the biological auditory and visual pathways in separate subsystems. In this architecture, each pathway draws its own conclusions about the correct label for a given sample of a person authentication task. An overarching conceptual subsystem assesses and combines the results of the pathways in a decision fusion process and arrives at a final conclusion. An extension of this work (Wysoski, Benuskova, & Kasabov, 2010) employs similar hierarchical subsystems for separate pre-processing of sound and video stimuli using MFCC and rank order coding, respectively. The conceptual subsystem that combines both streams is now based on an SNN that enables interaction between the unimodal processing layers so that they can inform each other's decisions. Like its predecessor, this method can again be categorised as a decision fusion approach.

A final piece of work that shall be mentioned in this section is based on **neuromorphic hardware** (Neil & Liu, 2016). As mentioned in Sections 4.2 and 4.3, specialised sensors have been developed that can transform sound and light into spikes in a fast and biologically plausible way, mimicking the behaviour of the ears and eyes. Those signals can then be combined using "sensor fusion" and serve as input for a neural network architecture. Neil and Liu (2016) used the AEREAR and DVS hardware sensors to create spikes from auditory and visual stimuli, respectively. The spikes were then combined using a software-based deep fusion network consisting of deep CNN and recurrent neural networks, which analysed and classified the data. However, the researchers reported no experiments on SNN hardware, even though this would seemingly be the next logical step. More recently, a novel approach based on neuromorphic hardware was introduced that draws heavily on findings from neurobiology for their system design. Oess, Löhr, Schmid, Ernst, and Neumann (2020) present a neurophysiologically plausible model of multisensory integration that is then deployed on the neuromorphic TrueNorth chip (Cassidy et al., 2014). The model simulates several concepts of neural interaction and behaviour. Another very recent publication (Rathi & Roy, 2021) employed an SNN to combine speech and video data. While the sound files are encoded using the Lyon model (Lyon, 1982), the video files are transformed into Poisson spike trains that are extracted from pixel intensity. Each modality contains several SNN layers with cross-modal connections formed between them that are modified using the brain-inspired spike-timing-dependent plasticity algorithm. It is encouraging to see the direction and findings of these works and hopefully more research interest will arise in the future

While the presented methods differed greatly in their chosen techniques, all papers that reported results for comparing **unimodal versus multimodal** systems found that the latter outperformed the former, sometimes considerably. Similar observations can also be made in the brain. Both biological and computational systems can benefit from diversified sources of information about their environment as it allows them to supplement missing data from one modality with information extracted by the other. However, this requires the system to be able to connect the knowledge gained from these separate sources in a meaningful and logical way. For example, temporal misalignment between semantically connected signals has been found to adversely influence the system's recognition rate, to an even greater extent than fully omitting the signals from one modality (Rao, 2016).

While this problem of temporal data stream synchronisation between different modalities has been addressed particularly by those systems employing multi-stream HMM, another aspect of integration is the **spatial arrangement of merged input signals**. To the best of the thesis author's knowledge, this topic has not been commented on in published works to date. However, spatial signal integration is arguably a fundamental concept of audio-visual information processing in the brain, as described in Section 3.4. Therefore, this topic is investigated in detail in the work presented in this thesis.

4.5 CHAPTER SUMMARY

This chapter summarised and introduced seminal works in the area of sound, video, and audio-visual processing, followed by a more detailed description of approaches that aimed to be more biologically plausible. While conventional machine learning methods achieved impressive results in several domain areas, they still lack the ability to understand context and more fluid concepts like mood and social behaviour. The brain, on the other hand, struggles with fast, precise calculations (of which there are many in more traditional approaches) but shines at scene understanding and knowledge integration. The work presented in this thesis explored some of these bio-inspired approaches further by taking inspiration from how the human hearing and vision systems process sound and video data. It aims to fill the gaps that were identified particularly in the area of signal mapping.

5 SYSTEM DESIGN AND ARCHITECTURE

"One's ideas must be as broad as nature, if they are to interpret nature."

– Sherlock Holmes in A Study in Scarlet

5.1 CHAPTER OVERVIEW

The main objectives of the research presented in this thesis were to build a biologically inspired model for audio-visual data processing and to evaluate its capabilities in order to assess the potential of such an approach. This chapter describes how the model was built, how decisions were made on which biological mechanisms to include, and how the model's parameters were informed by their natural counterparts. Chapters 6, 7, and 8 then describe the experimental setups to evaluate the model and its parts. The entirety of this chapter and the presented bio-inspired approach form a reply to Research Question 1a that was raised in Section 1.3 about how the biological background can inform the design of the computational model.

Irrespective of the modality of the input data, the modelling process always followed the same four steps shown in Figure 5-1. However, the modality influences *how* these steps were performed so that a satisfactory biological plausibility could be achieved, as indicated by the blocks and icons in the figure. The numbers in the figure denote the section of this thesis in which the steps are explained in detail. Those sections highlighted in bold red font denote novel contributions of the work presented in this thesis, while italic blue font marks sections that were largely based on existing work that was reused and adapted. A summary of the original and adapted contributions of this work is given in Section 5.6.

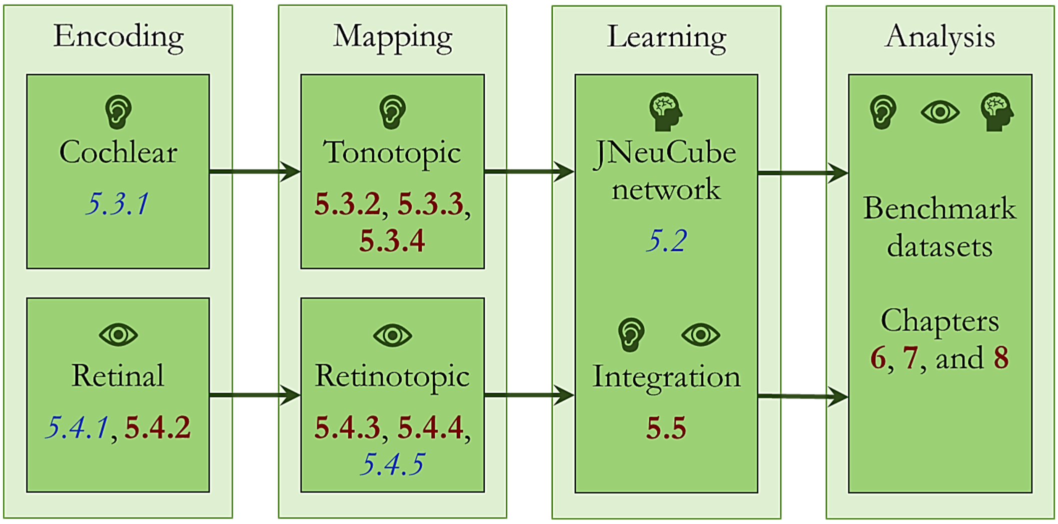


FIGURE 5-1: THE FOUR STEPS OF THE MODELLING PROCESS. SECTIONS IN BOLD RED FONT MARK NOVEL CONTRIBUTIONS, WHILE SECTIONS IN ITALIC BLUE FONT WERE BASED ON EXISTING WORK.

The following paragraphs give a general overview of the purpose of the steps and outline what they entail with respect to the two stimulus types.

Step 1. **Encoding**. The raw signals – WAV sound files for the auditory model and MP4 video files for the visual model – were transformed into spikes in a process that aimed to resemble the transformation of sound and light into brain signals by the cochlea and retina, respectively. This process answers Research Question 2a that was asked in Section 1.3. For the cochlear encoding, described in Section 5.3.1, an existing model was adapted and used. In contrast, the retinal encoding algorithm was a novel development that was based on how event-based cameras process video data. It contains two parts, peripheral greyscale vision described in Section 5.4.1 and foveal colour vision described in Section 5.4.2.

Step 2. **Mapping**. This step answers Research Question 2b that was asked in Section 1.3. The encoded signals are entered into the neural network in cortically plausible locations based on their data type. Since the network was created to be three-dimensional and brain-shaped, these input locations could be chosen based on the position of the auditory and visual cortices in the brain. The sets of input coordinates required for this step were developed from scratch and based on relevant brain imaging data that had been acquired and made available by neuroscientists. The process of deriving the tonotopic maps for the auditory input signals is described in Section 5.3.3 and the process of creating the retinotopic mapping for the visual input signals can be found in Section 5.4.4. Furthermore, since several different network sizes were developed as part of this research (see Section 5.2.3), not just the location but also the number of input neurons had to be determined so that the voluminal ratio between the size of the functional cortical area and the whole brain could be preserved. How these figures were calculated for the sound processing system is described in Section 5.3.2 and for the video processing system in Section 5.4.3. Moreover, the data created in the encoding step had to be compressed to a certain degree to fit into these calculated available numbers of input locations. A novel, flexible method for compressing auditory data was developed for the sound processing system and is described in Section 5.3.4, while for the video processing system, an existing method of splitting frames into meaningful sections was revised and enhanced as described in Section 5.4.5.

Step 3. **Learning**. In this step, a brain-inspired SNN architecture called NeuCube, which is equipped with unsupervised and supervised learning algorithms, was used to recognise patterns in the encoded data that are connected to specific events in the original stimuli. The neural network had a three-dimensional, brain-shaped layout that facilitated the mapping of the input signals in the Mapping step. For the models developed here, a NeuCube implementation in the Java programming language was used, called JNeuCube to distinguish it from the earlier MATLAB implementation that had been used for the pilot study in Chapter 2. The characteristics and features of the learning process are described in detail in Section 5.2. Another aspect of the learning step was the possibility to integrate different input modalities, as was asked in Research Question 2c in Section 1.3. In this research, auditory and visual signals were combined based on the location-dependent and temporally aligned connection of different modalities in the brain. The process for this integration is described in Section 5.5.

Step 4. **Analysis**. Making use of both the unsupervised and the supervised learning modes in the JNeuCube, the analysis step included traditional classification of benchmark data (shown in Chapters 6 and 7) as well as an unusual, yet potentially insightful means of visualisation (shown in Chapter 8). In this novel visualisation, the connections between the neurons that were modified during the unsupervised learning process could be visualised in the brain-shaped network and provide clues as to which areas of the network were most active when being presented with particular stimuli.

Contrary to what might intuitively be expected by the reader, the structure of this chapter does in fact not follow the order of the steps from 1 to 4, but rather the order in which the software and data that were necessary to perform these steps were developed and acquired. While the steps as shown in Figure 5-1 are more easily explainable from a top-down, systemic point of view, the amount of detail and explanation in the remaining sections of this chapter warrants a more topical bottom-up approach that follows the functional aspects of the model rather than its chronological parts. With this in mind, the **outline of the chapter** is as follows:

Section 5.2 introduces the NeuCube architecture that was used in Steps 3 and 4 of the modelling process. The section explains the design of the system and how the neural network was adapted and utilised for this research. This section also includes an explanation of the parameters that could be specified for the JNeuCube implementation as well as an overview of the different network sizes that were developed for this research. Section 5.3 describes all parts of the model that were specifically related to sound processing, including the cochlear encoding for Step 1 and the tonotopic mapping for Step 2. Analogously, Section 5.4 describes all those parts of the model that were used for processing visual stimuli, including the retinal encoding for Step 1 and the retinotopic mapping for Step 2. Section 5.5 then goes on to discuss the two key aspects of data integration, namely the spatial arrangement of the input data in the network and the temporal synchronisation of events, which became relevant in the Learning step of the model after both the auditory and the visual stimuli had been encoded and mapped using their respective processing pipelines. Finally, Section 5.6 concludes the chapter and summarises its contributions.

5.2 A BRAIN-INSPIRED SPIKING NEURAL NETWORK

This section describes the theoretical background of the neural network architecture that was used to process the encoded stimulus data. It then focuses on an implementation of this architecture in the Java programming language, called JNeuCube, its parameters, and what is known about how these influence the model's performance. The section closes with a description of the different network sizes that were created to be used for this research.

5.2.1 SYSTEM ARCHITECTURE

The human brain contains about 86 billion neurons (Azevedo et al., 2009). Each of these neurons is connected to numerous other neurons through its axon and dendrites, creating trillions of synaptic communication links that facilitate the brain's abilities to evaluate, reason, and create. In line with this research's objective to create a brain-inspired model, the characteristics of this "biological neural network" were taken into consideration when designing the features of the audio-visual processing system. Therefore, an artificial neural network (ANN) was chosen to be the central processing and learning unit for the system, specifically, a spiking neural network (SNN).

SNNs are considered the third generation of ANNs (Maass, 1997). Their main improvement over first- and second-generation ANNs is that SNNs do not just consider the strength of the signals as determined by the weight of the neural connections, but also their temporal characteristics, which represent events in the original stimuli. While in traditional ANNs, signals are systematically propagated through the network layer by layer, effectively forcing the neurons to send signals at fixed time steps, the firing time of a spiking neuron is dependent on its post-synaptic potential being built up by incoming signals until it reaches a defined threshold, a process which is not bound by synchronised time intervals (Pfeiffer & Pfeil, 2018). Naturally, this feature is expected to improve their ability to recognise patterns in data that are dynamic, where information is captured in temporal relationships in addition to possibly existing spatial or spectral components. Since this research focused on analysing spectro-temporal auditory data and spatio-temporal visual data, which are inherently dynamic, using a specialised SNN for this task was considered a promising approach.

The specific SNN architecture used for this research was the so-called **NeuCube** (Kasabov, 2014). This architecture consists of a "reservoir" of spiking neurons that are arranged in three-dimensional space and connected to each other, whereby the patterns of the incoming signals are captured in the weights of these connections by specialised learning algorithms. NeuCube employs **four stages** with multiple adjustable parameters in its data processing pipeline.[18] In the first stage, the network is **initialised**. Neurons are positioned in the reservoir using coordinates specified by a location file and then connected to some of their neighbours by a small-world connectivity algorithm with randomly chosen connection weights. Some of these neurons are designated as input nodes that receive incoming signals; these are also specified by a location file with coordinates.

The second stage is the **unsupervised learning** of the incoming data using *spike-timing-dependent plasticity* (STDP; Sjöström & Gerstner, 2010). In this process, the pre- and post-synaptic weights of the neural connections are modified as the input signals are propagated through the network, leading to long-term potentiation and long-term depression effects. Temporal associations between neurons are learned by adjusting the weights of connections based on the firing order of pre- and post-synaptic neurons. If the pre-synaptic neuron fires just before its post-synaptic neuron fires as well, the weight of their connection is increased. Likewise, the weights of connections where this firing order is reversed are decreased. The degree to which these adjustments are performed is determined by a learning function (Sjöström & Gerstner, 2010).

The third stage of NeuCube's data processing pipeline is the **supervised learning** of the trained network using the so-called *dynamic evolving SNN* algorithm (deSNN; Kasabov, Dhoble, Nuntalid, & Indiveri, 2013). This algorithm classifies the now modified connections in the network by connecting every active neuron in the reservoir to a single, newly created output neuron that represents the sample on which the network has just been trained. Only neurons that spiked at least once during the unsupervised learning stage are connected to an output neuron since inactive neurons would not produce any meaningful signals. The weight of the connections to the single output neuron is always initialised with the same value but is then modified based on the spikes that are generated by the reservoir neurons when processing the input sample. The presence of spikes increases the weight of the output connection, while their absence decreases it.

In the fourth and final stage, the weights of these output connections are associated with the **labels** of the original input samples using the k-Nearest Neighbour algorithm (kNN). The weights of the connections to the output neuron for a sample are compared to the output connection weights of other samples and the best match is chosen. This approach is based on the assumption that samples representing the same original stimulus will create similar patterns in the network and hence similar output connection weights (Kasabov et al., 2013).

[18] For better readability, this section only gives a general overview of the processing pipeline. A detailed list of all available parameters is provided in Section 5.2.2

The originally proposed NeuCube architecture also included several algorithms for data encoding that could transform the input data into spikes with varying level of detail (Kasabov, 2014; Petró, Kasabov, & Kiss, 2020). However, this feature was not used for this research. Instead, specialised encoding algorithms that were based on biological cochlear and retinal functionality were developed and are an original contribution of this thesis. Likewise, the mapping algorithms included in NeuCube (Tu, Kasabov, & Yang, 2017) were replaced by those developed as part of this research, These new mapping algorithms were based on tonotopy for the auditory signals and retinotopy for the visual signals.

The **final architecture** that was used for this research is depicted in Figure 5-2. Stimulus data in the form of sound and video files were transformed into spikes by the cochlear and retinal encoding algorithms, respectively. These spikes were then mapped into the network at the locations of the primary auditory and visual cortices. The STDP algorithm adjusted the connections in the brain-shaped reservoir based on the incoming signals as they were propagated through the network. After that, the deSNN algorithm connected the whole network to single output neurons, which were then associated with a class label by the kNN algorithm based on their connection weights. A similar architecture to the one described here was used successfully by Vanarse, Espinosa-Ramos, Osseiran, Rassau, and Kasabov (2020) for odour data classification.

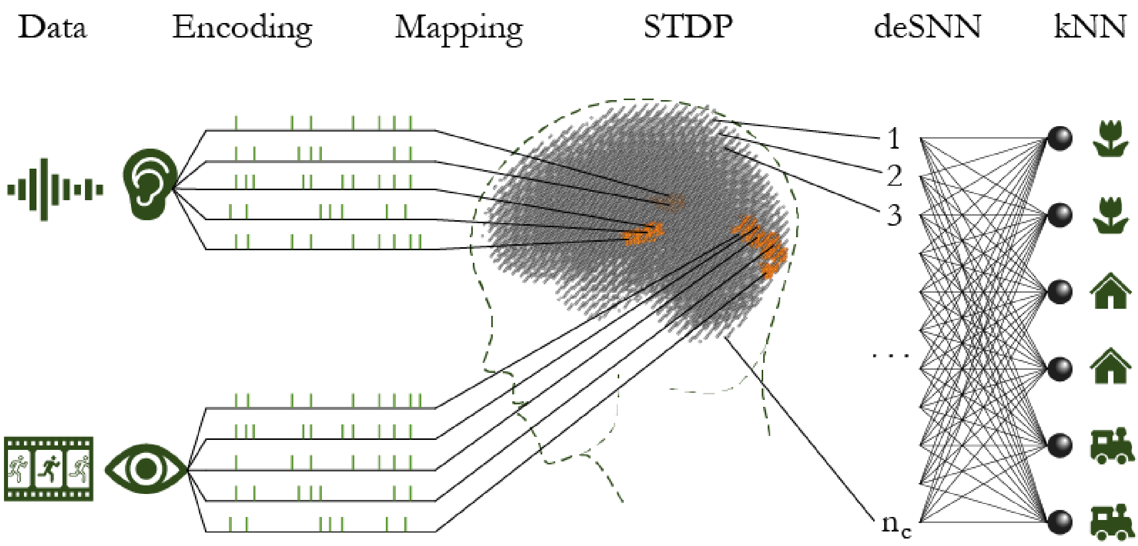


FIGURE 5-2: THE ADAPTED NEUCUBE ARCHITECTURE THAT WAS USED IN THIS RESEARCH.

Note that while the two stimulus types could be presented to the model at the same time, there was no obligation to do so. Both the sound and the video processing system could operate independently from one another. However, due to the biologically inspired mapping algorithm, audio-visual data could be combined easily, and it was one objective of this research to investigate if their combination could improve the performance of the model.

The NeuCube architecture was already briefly introduced in Chapter 2, where it was used to analyse and classify brain data collected through an electroencephalograph while a study subject was presented with written and spoken words and images of ten objects and living beings. One conclusion of this study was that in future projects, a more reliable, flexible, and accurate implementation of the NeuCube should be used to achieve better and more robust results. The implementation that was used for the models presented here is the JNeuCube. This Java-based version of the NeuCube supports the automated execution of experimental setups for large datasets as well as several algorithms for parameter optimisation. The following section explains the parameters of the JNeuCube in detail.

5.2.2 SYSTEM PARAMETERS

The JNeuCube implementation includes several parameters that could be configured to set up and run experiments on the raw data. The values for these parameters were passed into the system in a text file with the "properties" extension. They can be broadly classified into two categories, administrative, i.e., folder and file names, and algorithmic, i.e., any values that were used in the functions and formulas. This section focuses on the algorithmic parameters as these could directly influence the learning and classification performance of the model. The rationale behind choices for or against certain configurations for the models developed for this research is explained where relevant and the listing of the parameters below follows the order of the modelling stages described in Section 5.2.1.

The following four parameters are high-level parameters for the JNeuCube that were relevant to the general configuration of the system and needed for subsequently performing the processing steps described in Section 5.2.1.

Problem type – Describes the kind of problem that was being investigated. This could be set to *classification* or *regression*, depending on the desired result. For the research described in this thesis, it was always set to *classification*.

Standard directory – The folder containing the data. Each data sample in this folder should be a comma-separated text file and be prefixed with *samX* where X is the incremental sample count. Maintaining the order of the samples was important as their labels were recorded in a separate "target file".

Standard target class label file – If the file with the data labels was not included in the standard directory with the prefix *tar_class*, its location and name could be specified here.

Is Encoded Data – A flag indicating if the data in the standard directory had already been encoded, i.e., transformed from raw data into spikes. For the experiments conducted as part of this research, this parameter was always set to *true* since the cochlear and retinal encoding algorithms were not part of the JNeuCube framework. All data were encoded beforehand and then made available to the system. Therefore, all parameters related to the encoding step are omitted from the explanation provided here and instead included in the relevant sections on cochlear (Section 5.3.1) and retinal encoding (Sections 5.4.1 and 5.4.2).

The following 19 parameters were relevant in the initialisation stage of the network including layout specification and description of the neuron properties. Parameters that were not relevant for the models developed in this research are omitted from this list.

Mapping mode – Set to *auto* or *file*, this parameter determined how the neurons in the reservoir would be laid out. Note that despite its name, this parameter was not related to the Mapping step but instead described the positions of the non-input neurons in the reservoir. In *auto* mode, they would be arranged in a cuboid grid for which the user could specify the number of neurons in each of the three dimensions, whereas for the *file* option the user could specify a file location with the exact coordinates. For this research, the *file* option was used so that the customised network sizes described in Section 5.2.3 could be evaluated.

Mapping coordinates file name – The location and name of the file that contained the coordinates of the neurons in the reservoir. Note that despite its name, this parameter was not related to the Mapping step but rather to the arrangement of the non-input neurons in the network.

Reservoir builder – This parameter defines the general behaviour of the neurons. For the research described in this thesis, it was set to *neucube_reservoir*, which meant that all neurons were excitatory neurons. Positive and negative values could then be assigned to the synapses of these neurons to simulate excitatory and inhibitory behaviour.

Input mapping – With the options of *algorithm*, *file*, or *image*, this parameter defines how the input neurons were selected. For this research, only the *file* input mapping was relevant as the locations of the input coordinates were developed in a separate process using tonotopic (see Section 5.3.3) and retinotopic data (see Section 5.4.4).

Input coordinates file name – The location and name of the file that contained the coordinates of the input neurons.

Allow inhibitory input neurons – Specifies whether incoming spikes could be negative (inhibitory) or not. Due to the nature of the encoding mechanisms used for this research, the spikes entered into the network were always positive (excitatory), so this parameter was set to *false*.

Spiking neuron type – This parameter could be set to *excitatory* or *inhibitory* and describes the behaviour of the neurons in the reservoir. Like the behaviour chosen for the input neurons, this parameter was set to *excitatory* because the network would only process positive spikes.

Core name – The neuron model that was used for all neurons in the network. Unlike the neuronal structure of the brain, which is shaped by numerous as of yet poorly understood cell types (H. Zeng & Sanes, 2017), all neurons in the JNeuCube exhibited the same behaviour to decrease the computational complexity of the system. The available options in the JNeuCube were leaky integrate-and-fire (LIF) neurons, which were first described more than one hundred years ago (Brunel & van Rossum, 2007); a simplified LIF model that was used in the MATLAB implementation of the NeuCube; the neuron models described by Izhikevich (2004); and probabilistic neurons (Kasabov, 2010). Only the original LIF and the Izhikevich neurons were considered for building the models described in this thesis. According to a comparative study by Izhikevich (2004), LIF neurons were not as biologically plausible as Izhikevich neurons, but about two-and-a-half times as computationally efficient. Some results from the body of literature suggest that using Izhikevich neurons can decrease the performance of the system (Tavanaei, Masquelier, & Maida, 2016). It is not yet understood if this also applies to the JNeuCube, so these two neuron models were deemed a good selection for the experiments presented in this thesis. The simplified LIF neurons and the probabilistic neurons did not have a sufficient literature base to warrant further investigation as part of this research. Therefore, these two neuron types were not considered here. The neurons' behaviour could be configured further with the parameters below.

LIF threshold voltage – When using the LIF neuron model, the energy of all incoming spikes during the unsupervised learning stage was added to the neuron's post-synaptic potential (PSP). If the neuron's PSP reached the voltage threshold specified here, it emitted a spike as well. This parameter was expected to contribute significantly to the classification accuracy of the system (J. H. Lee, Delbrück, & Pfeiffer, 2016), which is why it was chosen as a parameter for optimisation and not fixed at this stage.

LIF reset voltage – The PSP of the LIF neuron was reset to this value after the neuron had emitted a spike.

LIF threshold refractory time – After emitting a spike, the LIF neuron would not spike again for a fixed number of time steps. This so-called refractory time is also present in biological neurons (Purves et al., 2017, pp. 55-56).

LIF resistance – In each time step, a neuron loses a small amount of its voltage potential. This "leakage" is described by the resistance parameter.

LIF capacitance – When integrating the incoming spikes with the existing electric potential of the neuron, the capacitance determined the degree of this integration.

Izhikevich behaviour – This parameter labelled the behaviour exhibited by the neurons based on the features described by Izhikevich (2004). Since this was expected to significantly influence the performance of the model, this parameter was not fixed at this stage but rather chosen as a parameter for optimisation in the experiments described in Chapters 6 and 7.

Connection algorithm name – With the locations and the behaviour of the neurons being defined, the JNeuCube offered three algorithms with which initial connections between them could be formed: the small-world connectivity (SWC) algorithm, an SWC variant for images of topographic maps, and an SWC variant for the "EPUSSS" algorithm. In the standard SWC algorithm, neurons were connected to a random subset of their spatially close neighbours whereby the weights of these connections were randomly chosen within a defined range. The specialised SWC for map data was seen as not suitable for this research due to its different application area, and the EPUSSS algorithm is still under development and was hence not studied as part of this research. Consequently, for the models presented here, the standard SWC algorithm was used.

Minimum weight value – The lower boundary of the range with which the random SWC connections were initialised. This could be a negative number, indicating an inhibitory connection.

Maximum weight value – The upper boundary of the range with which the random SWC connections were initialised.

Small world radius – This number defined the maximum distance that two neurons were allowed to have to be considered "neighbours" by the SWC algorithm. Only neighbouring neurons were randomly connected by the algorithm, disregarding the proven functional importance of long-distance connections (Knösche & Tittgemeyer, 2011) for model simplification and computational efficiency. The value of this parameter was dependent on the coordinate system that was used and hence scaled according to the size of the network from a base value of 2.5 distance units.

Small world positivity rate – The fraction of all connections created by the SWC in this step that were assigned positive weights and hence became excitatory connections. Research by Hendry, Schwark, Jones, and Yan (1987) showed that about 70% of neurons in monkeys' brains are excitatory,[19] which is why this value was left at 0.7 for all models presented here.

The next processing stage in the JNeuCube was the learning stage where the connection weights of the neurons were first adjusted by an unsupervised learning algorithm to capture the patterns in the data. Following the unsupervised learning, the JNeuCube passed the network's connection weights through a supervised learning algorithm. The following 13 parameters were relevant in this stage.

Unsupervised learning algorithm name – In addition to the "basic" STDP algorithm (Sjöström & Gerstner, 2010) that was briefly explained in Section 5.2.1, the JNeuCube offered three variations of STDP as well as two other learning algorithms that could adjust the connection weights in the network. Since the plasticity mechanisms of the brain vary widely across regions and are also influenced by neuromodulators and other factors (Sjöström & Gerstner, 2010), it is a difficult and as of yet unsolved problem to find an artificial algorithm that can fully mimic this level of detail and still be computationally efficient. While the research in this thesis aimed to be biologically plausible, the experiments that were run to evaluate the model had to be run on conventional von-Neumann hardware. Therefore, the basic STDP algorithm was chosen for its biological inspiration and its computational efficiency. The basic STDP function that was used by the JNeuCube to adjust the weights was:

$$W(x) = \begin{cases} A_{pos} * e^{\frac{-x}{\tau_{pos}}}, & x > 0 \\ A_{neg} * e^{\frac{x}{\tau_{neg}}}, & x < 0 \end{cases}$$

The four parameters for this function (***A* positive, *A* negative, τ positive, τ negative**) could be specified in the properties file. While the A values determined the rate at which the weights were changed, the τ values effectively defined the time window over which this change was calculated. In addition, an **upper bound** and a **lower bound** for the connection weights could be set. These two parameters restrict the connection weights to a specified range, mimicking the limited voltage range of biological neurons (Sjöström & Gerstner, 2010). All six parameters were left at fixed values for all the models presented in this thesis based on suggestions by the lead developer of the JNeuCube software. Both A positive and

[19] Unfortunately, these kinds of measurements are rather invasive, so no data for humans were available at this stage.

A negative were set to 0.001 and both τ positive and τ negative were set to 10. The boundaries for the weight range were chosen to be -2 and +2. Within the limited time and scope of this research, it was considered more valuable to first investigate the capabilities of the newly developed methods for encoding and mapping before trying to improve the performance of the JNeuCube reservoir, which is why these parameters were left at their default values.

Training rounds – Determines how many times the input data were passed through the network. With very few or short stimulus data, this could increase the pattern recognition ability of the network, since significant events in the data were repeated and the patterns that could be captured by the weights would be reinforced. For the data that were used with the models presented here, one training round was considered a sufficient starting point.

Saving weights mode – If this parameter was set to *true*, all connection weights of the network would be saved at every step of the learning process. While this feature could be used to analyse the evolution of the connections that the stimulus patterns imprinted into the reservoir, it was not considered essential for this research and hence left at *false*. The insights from this additional analysis were not expected to be significant enough to warrant the required increased training time and storage space.

Supervised learning algorithm name – There were two supervised learning algorithms implemented in the JNeuCube, which were both based on the deSNN algorithm. For all models in this research, the standard deSNN algorithm described by Kasabov et al. (2013) was used. With this algorithm, the spiking activity of the reservoir was captured by connecting all neurons that had spiked at least once during the unsupervised learning to a single output neuron per sample.

Modulation factor – This parameter was used to calculate the connection weights for the connections to the output neurons. The weights were determined by calculating the modulation factor to the power of the order of the spike. Since this factor has to be smaller than 1, the order of the spikes thus determines their significance for the connection.

Positive drift – If a spike was sent through one of the output connections, this parameter determined by how much its weight was increased.

Negative drift – If no spike was sent through an output connection in a given time step, its connection weight was decreased by this factor.

The final stage of the learning in the NeuCube was the classification of the data. More precisely, a class label was assigned to the single output neuron's connection weights based on their similarity to the connections of other samples' output neurons. In combination with those parameters that were needed for the general setup of the system to form an experimental setup that could produce reliably correct results, this section contains the following seven parameters:

Classifier name – In the JNeuCube implementation, there were two classifiers available, the kNN algorithm and a weighted kNN. In the default kNN algorithm, samples were classified by identifying a specified number *k* of their spatially closest neighbours and then being assigned the most common label among them. This required knowledge of the correct labels of all samples in the dataset, which is why the kNN algorithm is categorised as a supervised learning method. In the weighted kNN, the labels are not simply counted to determine the majority but also weighted based on the distance between the to-be-classified sample and its neighbours – the smaller their distance, the more weight is given to a neighbour's label and vice versa. For the research presented here, the simple kNN algorithm was used because this step does not have a sufficient biological base from which to draw conclusions and thus the computationally less complex method was favoured.

kNN *k* – The number of neighbours that were considered for the kNN algorithm. This is typically an odd number to avoid a stalemate situation when voting on the majority label.

Classifier distance – The method with which the distances between a sample and its neighbours were calculated. In the JNeuCube implementation, Euclidian and Gaussian distance calculations were available. For the models developed here, Euclidian distance was used since it was the most straightforward option. The Euclidian distance is calculated by computing the square root of the sum of the differences of all the samples' features in the feature space.

Cross-validation method name – The JNeuCube offered two methods for cross-validating the models, k-fold and Monte Carlo. For this research, the k-fold cross-validation method was chosen due to its lower complexity. In k-fold cross-validation, the dataset was split into a defined number of folds, of which all but one were used to train the model while the remaining fold was used to test the model. This was repeated until all folds had been used as test fold once. The overall classification accuracy of the model was then determined by calculating the arithmetic mean of all fold-models' accuracies.

Num folds – The number of folds that were used for the k-fold cross-validation method.

Training rate – Described the fraction of the total number of samples that were used for training the model, for example, 0.7 for 70% of the samples (the value used in this research). The remaining samples were used for validating the model. This split occurred *before* the data were further split into folds for the cross-validation, as shown in Figure 5-3. The network models that were created during all six runs (five cross-validation runs and one final validation run) were always initialised with the same connection weights and then processed the different sample populations (varying combinations of four folds for the cross-validation and all five folds for the final run). While the cross-validation runs provided an insight into how well the model could classify the existing data, labelling the validation samples tested its generalisability. This procedure ensured that the model's reliability could be verified with completely unseen test samples, which increased the reliability of the results.

Training samples (70%)					Validation samples (30%)
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	

FIGURE 5-3: DATA SPLITTING APPROACH FOR TEST AND VALIDATION.

Number of experiments – While not an explicit parameter in the properties file, this number still had to be determined before executing the JNeuCube software to build the models. Due to the randomised initialisation of the neural connections, both favourable and unfavourable configurations could occur. By running each experimental setup 30 times and then reporting the algorithmic mean of the achieved classification accuracies as the final result for the experiment, it was expected that extreme variants would be smoothed out and that the validity of the results would be further improved.

The majority of the parameters described in this section were left at their default values that had been chosen by the lead developer of the JNeuCube implementation. Since only limited theoretical work had been done on optimum parameter settings for SNN in general and the NeuCube architecture in particular, and it was expected that the nature of the data would influence to what extent the parameters would affect the model's performance, finding the best configuration was a difficult endeavour. The values chosen by the lead developer reflect anecdotal evidence from previous work on the same system, where they had led to satisfactory results. Furthermore, by using the same parameters as for other research projects, it was hoped that a comparative data basis could be created in the future to better inform these choices.

5.2.3 NETWORK TEMPLATE SIZES

A secondary objective of this research was to investigate if and how the size of the neural network influenced the model's performance, where performance is measured by classification accuracy. As described in Section 5.2.2, the layout and hence the size of the neural reservoir of the JNeuCube could be specified through a list of three-dimensional coordinates. If these coordinates resembled the brain and were provided at different scales, varying network sizes could be modelled easily. However, since no suitable dataset with such scaled coordinates was publicly available, it had to be developed first as part of this research.

The most suitable source for a "base template" that could then be scaled in size was thought to be research involving brain imaging data. However, trying to create a biologically plausible set of points that resembled *the brain* turned out to be not as straightforward as initially thought. There is a huge range of diversity in how brains are shaped and where functional regions are located (Brett, Johnsrude, & Owen, 2002). While some landmarks, like Heschl's gyrus as the location of the primary auditory cortex and the calcarine fissure as the location of the primary visual cortex, can be found reliably well in every brain, other areas, for example along the temporal lobe, can show large discrepancies between individuals (Brett et al., 2002). This meant that any network templates based on brain imaging technology were effectively limited in their resolution, and hence their biological plausibility since a common set of coordinates could not be as precise as individual brain data. However, the standardisation of brain volumes does create advantages for the neuroscientific research community in terms of comparability and reproducibility of their work, so attempts have been made by these researchers to create such a template despite its anticipated shortcomings.

The first such template was developed by Talairach and Tournoux (1988). It was based on a single brain exemplar and has been widely used in brain imaging studies. The Talairach template was however limited in its generalisability due to underlying neurological conditions of the brain donor. This limitation prompted the development of a more generalised template by the Montreal Neurological Institute (MNI). The MNI template was based on several hundred brain scans that were averaged into a common template (Evans et al., 1993). Since then, the MNI template has increasingly replaced the Talairach template in brain research, however, both brain atlases are still used widely today (Laird et al., 2010). Therefore, these two well-recognised and widely adopted brain templates were chosen as the basis to develop the different network coordinates for the models in this thesis. The anatomical inaccuracies existing in these templates were not expected to limit the performance of the models because the locations of the two regions most relevant for this research – primary auditory and visual cortex – had already been determined as being reliably identifiable across individuals (Brett et al., 2002).

The original Talairach template was downloaded from www.talairach.org where it was available as an annotated NIfTI file (Lancaster et al., 1997; Lancaster et al., 2000). Files in NIfTI format contain a three-dimensional matrix with meaningful values, which in this case were anatomic labels that described the hemisphere, lobe, gyrus, tissue type, and cell type of each element. With dimensions of 141*172*110 elements, the NIfTI matrix contained a total of 2,667,720 data points, of which 1,527,747 were annotated with a label indicating their position in the brain. Therefore, these **1,527,747 elements** were chosen as the basis for a **Talairach coordinate template** that was adapted to the needs of the work presented in this thesis. The NIfTI file also contained metadata specifying a matrix transform function with which the elements of the index-based matrix could be converted to coordinates. While the indices in the NIfTI matrix must be positive for computational reasons, the anatomical coordinates of both the Talairach and the MNI template typically contain negative values because their coordinate origin is located in the centre of the brain. The transform function could thus be used to align the NIfTI matrix back to the original coordinates, moving the anterior, inferior, and left parts of the brain into negative coordinate space.

The MNI template was downloaded from www.alivelearn.net/xjview/download-link/ where it was available as part of the xjView toolbox in the form of a database that contained a list of coordinates for each brain label. According to the toolbox's documentation (X. Cui, n.d.), the database was obtained from a third party who had taken the Talairach coordinates and brain labels described above and applied a non-linear transformation (Laird et al., 2010; Lancaster et al., 2007) to convert them into MNI space (Pakhomov, 2014). The xjView database contained 268 lists of coordinates, one for each anatomical feature of the brain such as hemisphere, cell type, lobe, and Brodmann area. Since one neuron coordinate can naturally be part of several of these lists,[20] duplicates had to be removed and the final **MNI template** created for this research contained **241,606 coordinates**.

Besides the obvious discrepancy between the numbers of coordinates, the templates also looked noticeably different, as shown in Figure 5-4. In this figure, the blue dots represent the coordinates from the Talairach template, and the orange dots represent those from the MNI template. The first difference can be observed in the size of the networks. As expected from the origin of the templates, the Talairach template was smaller than the MNI template. The size discrepancy seemed to be especially pronounced in the extreme superior and anterior regions of the brain and is visible in the view from the left (left tile of the figure) and the back (middle tile). Another noticeable variation was the difference in density. This was also expected since there were a lot more coordinates available for the Talairach template than for the MNI template. However, it was also evident that both templates were aligned to the same coordinate origin point and that they were using the same scale. Based on the information provided with the Talairach NIfTI metadata, each coordinate point represented a brain volume of one cubic millimetre.

[20] A neuron in Brodmann area 17 will also be located in the occipital lobe, for example, and additionally be part of a hemisphere and of grey matter.

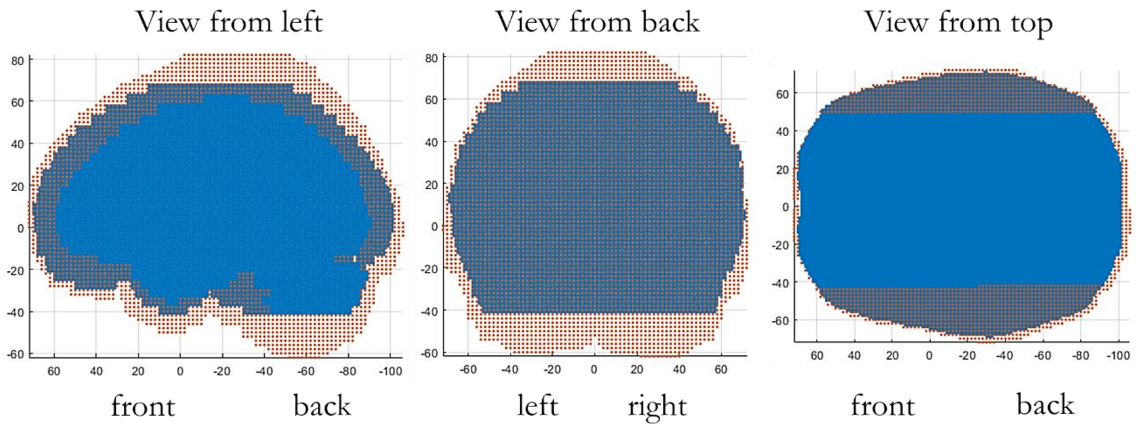


FIGURE 5-4: VISUAL COMPARISON OF THE COORDINATES FOR THE TALAIRACH (BLUE) AND MNI (ORANGE) BRAIN TEMPLATES.

Coming back to the objective mentioned at the beginning of this section, these two "original" brain templates had to be scaled to create several different network sizes that could be used to build the models for this research. This was done by performing a stepwise selection of neurons from the templates based on their three-dimensional structure: along each of the three dimensions, only every *n*th neuron was selected to remain in the template while all others were discarded, as shown in Figure 5-5. The example in the figure shows a scaling factor of two, where only every second neuron remained in the network. This stepwise selection was applied to both the Talairach and MNI templates until a low four-digit number of remaining neurons was attained for each. Smaller templates would not have been viable when performing the biologically plausible mapping since the number of input neurons was going to be dependent on the voluminal ratio between functional cortex and brain. If the brain template would become too small, the number of input neurons would sink to fractions of neurons and hence not be representable anymore.

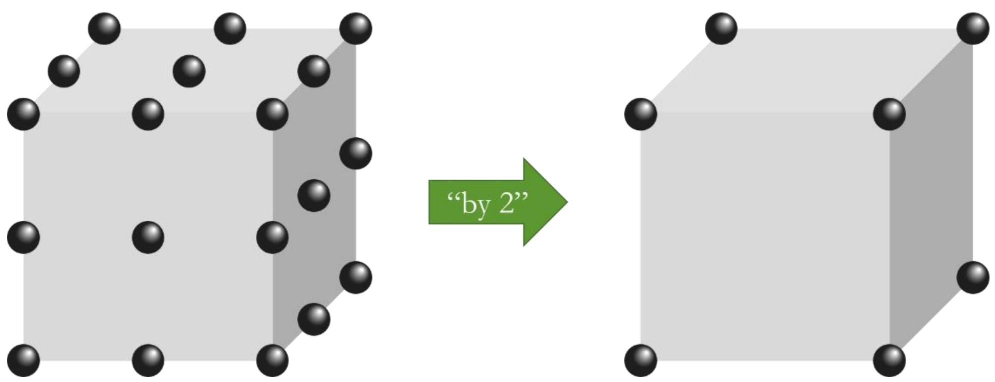


FIGURE 5-5: VISUALISATION OF THE STEPWISE SELECTION OF NEURONS IN EACH DIMENSION TO CREATE THE SCALED BRAIN TEMPLATES.

The final number of neurons for each scaled template is shown for the Talairach atlas in Table 5-1 and the MNI atlas in Table 5-2. A scaling factor of *n* indicates that every *n*th neuron in each dimension was kept in the network while all remaining neurons were removed. The template names were used throughout this research to refer to these specific sets of coordinates. Furthermore, since the original MNI template contained only even numbers, one additional enlarged MNI template was created by filling in the absent odd coordinates.

TABLE 5-1: NEURON COORDINATE NUMBERS OF THE BRAIN TEMPLATES THAT WERE CREATED FOR THIS RESEARCH BASED ON THE TALAIRACH ATLAS.

Template name	Number of neurons	Scaling factor
TAL_orig	1,527,747	1
TAL_by_2	192,600	2
TAL_by_3	56,770	3
TAL_by_4	23,550	4
TAL_by_5	12,150	5
TAL_by_6	7,199	6
TAL_by_7	4,452	7
TAL_by_8	2,960	8
TAL_by_9	2,086	9
TAL_by_10	1,525	10

TABLE 5-2: NEURON COORDINATE NUMBERS OF THE BRAIN TEMPLATES THAT WERE CREATED FOR THIS RESEARCH BASED ON THE MNI ATLAS.

Template name	Number of neurons	Scaling factor
MNI_times_2	1,932,848	½
MNI_orig	241,606	1
MNI_by_2	30,182	2
MNI_by_3	8,907	3
MNI_by_4	3,747	4
MNI_by_5	1,939	5

While this downscaling approach decreased the number of neurons in the network, it did not alter the network's shape or dimensions. Smaller templates still resembled their original sources visually, albeit with lower neuron density. Figure 5-6 shows this for both the Talairach (top row) and the MNI (bottom row) templates. The left column (blue) shows the original templates, the middle column (orange) shows a medium-sized template (TAL_by_5 and MNI_by_3), and the right column (green) shows the smallest available template for each atlas (TAL_by_10 and MNI_by_5).

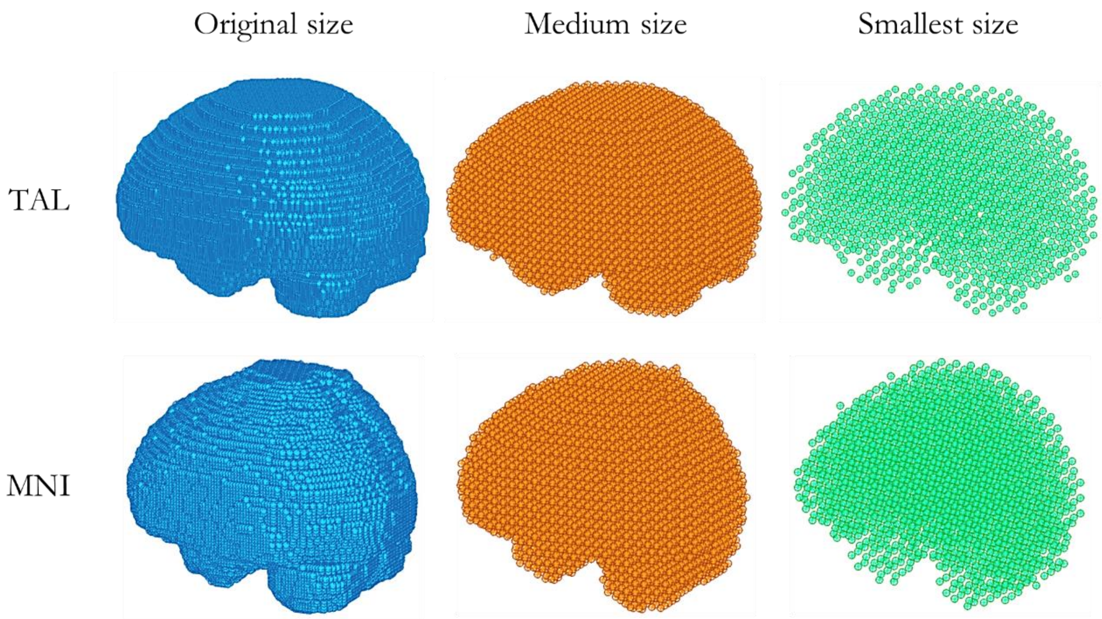


FIGURE 5-6: VISUALISATION OF DIFFERENCES BETWEEN TEMPLATE SIZES. NOTE THAT WHILE THE NEURON DENSITY DECREASES, THE OVERALL SHAPE IS RETAINED.

The two templates that were obtained by converting brain imaging data found in literature and the further 14 templates that were developed from them as part of this research form an important foundation for the models presented in this thesis. The final lists of coordinates for the different network sizes constitute an original contribution of this research that serves as an intermediate step to answering Research Question 3c, which in Section 1.3 asked if the size of the neural network influences its performance.

5.3 SOUND PROCESSING SYSTEM

This section describes the design of a biologically inspired sound processing model that is based on the human auditory system. It focuses on four areas:

- the transformation of sound waves into electrical signals based on the functionality of the cochlea,
- the position and size of the auditory cortices within the brain,
- the frequency-based (tonotopic) mapping of the electrical signals into the primary auditory cortices, and
- the merging of these signals along the primary auditory pathway.

These four focus areas were identified as a suitable starting point for the design of a computational system after studying the relevant literature, which showed that they were comparatively well researched and understood by hearing scientists.[21] Since they were found to play key roles in the hearing process, the computational system developed as part of this research largely follows their functionality, albeit with some simplifications. Related literature from biology and neuroscience informed the design of the model and is discussed where applicable. The performance of the developed model was then evaluated using benchmark data, the results of which are reported and compared to existing systems in the same domain in Chapter 6.

5.3.1 COCHLEAR ENCODING

This section describes how the sound data were converted into spikes by the sound processing model developed in this research. As described in detail in Section 3.2.1 and illustrated in Figure 3-3 and Figure 3-4, the cochlea has the remarkable ability to transform mechanical sound waves into electrical brain signals. Due to its unique structural mechanics, the basilar membrane running inside the cochlea starts to vibrate at a point that is directly related to the frequency of the sound stimulus. So-called hair cells located on top of the basilar membrane are depolarised as their “hairs” (stereocilia) are pushed into the tectorial membrane located above them by these vibrations. This leads to rapid depolarisation of the hair cells, creating electrical signals that are transmitted via auditory nerve fibres to the next processing stages along the auditory pathway. This process of cochlear encoding is simulated in the sound processing model described in this chapter.

---

[21] An introductory explanation of the human auditory system can be found in Section 3.2.

In any mammalian cochlea, the locations where the sound vibrations cross the basilar membrane and excite the hair cells can be calculated using the **Greenwood function** (Greenwood, 1961). This function expresses the relationship between specific frequencies and positions in the cochlea:

$$f = \int_0^x \Delta f_{cb} = A(10^{ax} - k)$$

where f is the frequency of the sound, A, a, and k are species-specific constants, and x is the distance between the apex and the point of interest on the basilar membrane. The three species-specific constants can be determined through experimentation by observing the reaction of hair cells on the basilar membrane when playing tones of known frequencies. For the human cochlea, which has to be studied on deceased subjects typically no longer than a few hours after death (Greenwood, 1990), the following values are suggested:

$$f = 165.4(10^{2.1x} - 0.88)$$

assuming that x is given as a fraction of the length of the cochlea (Greenwood, 1990). Since a typical human cochlea is about 34 millimetres in length (Wright et al., 1987), the value for a changes to 0.06 if x is given in millimetres. With this information, the Greenwood function can be used to calculate the sound frequency to which a given location on the basilar membrane responds and vice versa.

The discovery of the Greenwood function was an important prerequisite for the development of applications that rely on a biologically plausible cochleotopic mapping of frequencies in the inner ear. For example, it is commonly applied in artificial cochlear devices for the hearing-impaired (Stakhovskaya, Sridhar, Bonham, & Leake, 2007). The Greenwood function is a fundamental part of the sound processing system presented in this chapter as it provides an algorithm for signal transformation in the very first step of the model.

The second neuroscientific finding that has to be considered in the development of the signal transformation and encoding step of the model is related to the **auditory nerve fibres** that are connected to the inner hair cells of the cochlea, the so-called *Type I* fibres. M. C. Liberman (1978) discovered that there are three types of Type I fibres, based on their **spontaneous firing rates**. The spontaneous firing rate indicates how many signals are created by auditory nerve fibres that are not a direct result of an auditory stimulus. In terms of electromechanical properties, spontaneous firing rates facilitate a distinction between different levels of loudness. Fibres with high spontaneous firing rates emit more than 18 spikes per second and are very sensitive to quiet sounds because they are easily excited. However, due to the recovery time needed between spikes, these fibres will saturate more quickly than those with low or medium spontaneous firing rates, and hence, temporarily cease to function effectively in louder environments. In contrast, fibres with medium (between 0.5 and 18 spikes per second) and low (less than 0.5 spikes per second) spontaneous firing rates mainly react to louder sounds and saturate less quickly, allowing a broader spectrum of loudness perception.

While there is limited[22] research on auditory nerve fibres of humans, Liberman's extensive experiments on cats have shown a **distribution of fibres** with high, medium, and low spontaneous firing rates of about 61%, 23%, and 16%, respectively (M. C. Liberman, 1978). Since the auditory nerves of cats seem to share several characteristics with the human auditory nerve, it is generally assumed that this distribution would be similar in humans (Nadol, 1988). Therefore, these figures were applied to the development of the sound processing system presented in this chapter.

The next step was then to decide how many hair cells and auditory nerve fibres to include in the model. Fortunately, several quantification studies of the inner ear have been performed, and it was found that there are approximately **3,500 inner hair cells** along the human cochlea (Wright et al., 1987) that are connected to about 30,000 Type I auditory nerve fibres (Spoendlin & Schrott, 1989). This means that, on average, each inner hair cell is connected to approximately 8.57 auditory nerve fibres. Furthermore, considering Liberman's distribution of spontaneous firing rates, there should be 18,300 high-spontaneous, 6,900 medium-spontaneous, and 4,800 low-spontaneous Type I fibres in total, or approximately **five, two, and one per hair cell**, respectively (M. C. Liberman, 1978).

This is, of course, a very simplified assumption to build the computational model described here. The different kinds of nerve fibres are, in fact, not evenly distributed between the hair cells (M. C. Liberman, 1978), and the number of hair cells along the cochlea as well as the number of auditory nerve fibres per hair cell varies widely between speech-relevant and non-speech-relevant frequencies (Spoendlin & Schrott, 1989). However, some level of **simplification** is necessary at this step for the model to be computationally feasible and hence potentially usable in real-world applications. Therefore, the sound processing system developed in this research modelled 3,500 hair cells per ear that are each connected to five auditory nerve fibres with high, two fibres with medium, and one fibre with low simulated spontaneous firing rates.

Based on the aforementioned studies and several other discoveries in the field of hearing science, numerous research groups have developed detailed **computational models of the**

[22] After all, while the ethics committee of the 1970's might have considered it acceptable to dissect and study "A litter of four cats, born and raised in a soundproofed chamber" (M. C. Liberman, 1978, p. 442), attempting this experimental setup on humans would have, then and now in 2021, raised ethical concerns.

inner ear, each focusing on different aspects that mimic human hearing (Clark, Brown, Jürgens, & Meddis, 2012; Holmberg et al., 2007; Meddis et al., 2013; Sumner, Lopez-Poveda, O'Mard, & Meddis, 2002; Verhulst, Dau, & Shera, 2012; Zeddies & Siegel, 2004; Zilany et al., 2014). In a recent review paper (Rudnicki et al., 2015), the authors systematically compared some of these models and found that in terms of biological accuracy and signal coding, the model developed by Zilany et al. (2014) performed best in their experimental setup. This 2014 model is an improved version of a previous model developed by the same group (Zilany, Bruce, Nelson, & Carney, 2009) that was based on observations of the auditory periphery. The authors also published the source code of their experiments on the online code repository GitHub as a module for the Python programming language (Rudnicki & Hemmert, 2016), making it easily available and reusable for other researchers.

Since it was not the intent of this thesis to develop and assess novel models of the inner ear, the recommendations by Rudnicki et al. (2015) were followed and the published implementation of the 2014 Zilany model was used as a basis for the auditory processing system presented in this thesis. The encoding software was downloaded from GitHub and customised to the needs of the model (see Appendix A, Listing II). The following **parameters** were used:

- *Sound*. A vector containing one channel of sound data.
- *Fs*. The sampling rate of the sound. For computational reasons, the model limited the sound frequency to between 100,000 and 500,000 Hertz. Since most available benchmark sound datasets used a lower sampling rate and it was unclear how resampling would affect the computational model, it was decided to fix this parameter at the lowest available value of 100,000 Hertz.
- *Anf_num*. This parameter contained the number of auditory nerve fibres per characteristic frequency grouped by their spontaneous firing rate (high, medium, low). Taking into consideration the quantification studies by M. C. Liberman (1978) in combination with the findings by Spoendlin and Schrott (1989), eight auditory nerve fibres were used, of which five were assumed to have a high spontaneous firing rate, two to have a medium spontaneous firing rate, and one to have a low spontaneous firing rate.
- *Cf*. This parameter contained the number and range of characteristic frequencies to which the auditory nerve fibres responded. Based on the work by Wright et al. (1987), 3,500 characteristic frequencies per cochlea were assumed to be biologically plausible. The frequencies were chosen to be in the range of 125 and 8,000 Hertz so they would match the frequency range covered by Langers, Sanchez-Panchuelo,

Francis, Krumbholz, and Hall (2014). Their work became relevant in the later steps of mapping and data compression described in Sections 5.3.3 and 5.3.4.

- *Species*. This parameter provides different options to simulate the species-specific size of the cochlea. This parameter was set to 'human'.
- *Seed.* Since there was no documentation for this parameter, it was left at its default value of 0. Presumably, this is used to initialise the randomisation of the spontaneous firing of the auditory nerve fibres.

The two model parameters that needed the most attention from a research perspective were the *number of auditory nerve fibres* and their *characteristic frequencies*. While the former was covered reasonably well by the literature in this section, the latter required a more detailed consideration of the overall network configuration for the final models. The following sections discuss in detail the relationship between the network size and the number of input neurons as well as their spatial arrangement and tonotopic distribution.

5.3.2 NUMBER OF AUDITORY INPUT NEURONS

The auditory cortices in the human brain take up a certain amount of space within the overall brain matter. Since the number of neurons in the neural network model is freely definable and is, in fact, a parameter for model optimisation, the size of the signal input region has to be kept proportional to the overall network size in order to retain biological plausibility. Therefore, this section looks at neurological evidence for the size of the auditory cortices.

A short excursion into the beginnings of neuroscientific research shows that the early 20th century saw a surging research interest in the structure and functionality of the human brain. Probably most well-known is the work of German neurologist Korbinian **Brodmann** who developed a brain atlas by identifying areas in the cortex of the brain based on their cytoarchitectonic (cell structure) characteristics. Brodmann's atlas is still widely used today. Other schools of thought developed at that time, such as myelinogenesis (studying the development of myelin in different fetal stages), which was pioneered by German neurologist Paul **Flechsig** and investigated the structure and composition of the brain from other perspectives, with somewhat different conclusions and considerable debate.[23] With regards to the auditory cortices, for example, Brodmann acknowledges that the area he identified as area 41 in his atlas was the previously discovered Heschl gyrus (Brodmann, 1909, p. 145). However, Brodmann was convinced that auditory processing is such an important feature of human consciousness that it can not be restricted to such a small area and must instead be

[23] While researching this section, I had the pleasure of reading the original papers by Brodmann and Flechsig in German, which contained an entertaining hint of rivalry.

distributed across the temporal lobe (Brodmann, 1909, pp. 313-315). This area 41 has now been verified as the location of the main auditory processing centre (Morosan et al., 2001). Interestingly, Flechsig through his myolegenetic studies had already discovered that auditory functionality is contained in the Sylvian fissure he called "Hörrinde" (German for "hearing cortex"). He first calculated that the auditory processing centres made up around two per cent of the overall surface area of the cortex (Flechsig, 1905). Later, he wrote that the regions he described as auditory cortices, which are also known as *temporal anterior transverse gyri*, had a total surface area of approximately two square centimetres (Flechsig, 1908). He also discovered that these gyri are a projection area for the cochlear nerves by looking at the pathways the nerve fibres take in the brain, and he noted that they seem much more complicated and tangled than the comparatively orderly arranged nerve fibres extending from the retina to the visual cortex (Flechsig, 1908).

Since the last century has seen immense advancements in available research methods, the feud between Brodmann and Flechsig could finally be scientifically settled using modern medical imaging technology. Morosan et al. (2001) in combination with Rademacher et al. (2001) used magnetic resonance imaging to study the brains of ten people and identify the size and location of their auditory cortices, specifically an area they called Te1. Area Te1 is nowadays considered the location of the primary auditory cortex as identified by cytoarchitectonic characteristics in combination with functional imaging technologies (Morosan et al., 2001). By averaging the sizes of the identified auditory cortices in their study subjects, Rademacher et al. (2001) determined that **area Te1 has a volume of 1,683 mm^3 in the left hemisphere and a volume of 1,581 mm^3 in the right hemisphere of the brain**.[24] Anecdotally, this difference between left and right had already been mentioned by Flechsig (1908). However, his observation that it was reversed in a left-handed person and that both cortices were of the same size in three musicians' brains he analysed could not be confirmed by more recent studies.

The computational neural network model described in Section 5.2 assumed that the overall network reservoir covers the whole brain. Just like the auditory cortices are part of the whole brain, their size had to be put into relation to the **whole brain volume**. Like the studies described above that determined the volume of area Te1, modern medical imaging technology has equally facilitated accurate whole-brain measurements. A notable paper by Allen, Damasio, and Grabowski (2002) describes how the authors studied the size and shape of the brains of 46 subjects using magnetic resonance imaging to investigate volume differences between the sexes. They found that male brains had an average volume of

[24] For better mental visualisation, these numbers are roughly between the cubes of 11 and 12 mm.

1,273.6 cm^3, while female brains had an average volume of 1,131.1 cm^3. Since a distinction between male and female is not relevant for the development of the computational sound processing system presented here, the arithmetic mean of both figures – **1,202.35 cm^3** – was used when designing the system.

In order to develop a biologically plausible model, the number of input neurons was determined using the following formula that preserves their proportions:

$$\frac{vol\ Te1\ (Rademacher)}{vol\ Brain\ (Allen)} = \frac{number\ of\ input\ neurons}{number\ of\ reservoir\ neurons}$$

Area Te1 makes up about 0.27% of the total brain volume, of which roughly 0.14% belong to the left hemisphere and roughly 0.13% belong to the right hemisphere. Combining these figures with the different network sizes that were generated for this research as described in Section 5.2.3 results in concrete **numbers for input neurons** shown in Table 5-3 and Table 5-4. These were the figures that were used later to determine the locations of the input neurons, which could then be supplied to the JNeuCube system as input coordinates that would match with the varying network sizes. This setup facilitated investigating Research Question 3c that asked if the size of the network influenced the model's performance.

TABLE 5-3: NUMBER OF AUDITORY INPUT NEURONS DEPENDING ON THE SIZE OF THE BRAIN TEMPLATE FOR THE TALAIRACH ATLAS.

Brain template	Number of reservoir neurons	Number of input neurons Left side	Right side
TAL_orig	1,527,747	2,138	2,009
TAL_by_2	192,600	270	253
TAL_by_3	56,770	79	75
TAL_by_4	23,550	33	31
TAL_by_5	12,150	17	16
TAL_by_6	7,199	10	9
TAL_by_7	4,452	6	6
TAL_by_8	2,960	4	4
TAL_by_9	2,086	3	3
TAL_by_10	1,525	2	2

TABLE 5-4: NUMBER OF AUDITORY INPUT NEURONS DEPENDING ON THE SIZE OF THE BRAIN TEMPLATE FOR THE MNI ATLAS.

Brain template	Number of reservoir neurons	Number of input neurons	
		Left side	Right side
MNI_times_2	1,932,848	2,706	2,542
MNI_orig	241,606	338	318
MNI_by_2	30,182	42	40
MNI_by_3	8,907	12	12
MNI_by_4	3,747	5	5
MNI_by_5	1,939	3	3

Naturally, the volume differences between the left and right sides only came into play for the larger network sizes. For smaller networks, fractional numbers had to be rounded because only entire neurons could be represented in the network. This caused inconsistencies like the discrepancy between the "TAL_by_6" and the "MNI_by_3" models. Even though "MNI_by_3" was larger, it was assigned the same number of input neurons for both the left and right sides because the calculated values were 12.47 and 11.58, respectively, both of which were rounded to 12.

After the number of input neurons in the network had been determined, their exact locations also had to be found. The next section describes this tonotopic mapping process.

5.3.3 TONOTOPIC MAPPING – LOCATION OF AUDITORY INPUT NEURONS

The previous section discussed how the number of input neurons was identified depending on the size of the neural network. Based on those numbers, this section reports on the investigation of the spatial distribution and location of these input neurons. The same literature that talked about the volume of the primary auditory cortices also discussed their location in the brain and found that it differed from person to person within a certain range (Rademacher et al., 2001). The difficulty when applying these data to a computational model was in converting the often inaccurate and inconsistent biological locations into exact computer-usable coordinates. The efforts described in this section present an attempt to solve this problem in a scientifically sound manner.

Tonotopy describes the relationship between sound frequencies and their corresponding processing regions in the primary auditory cortex. Sound stimulus frequencies that are close to each other are processed in neighbouring areas, and this overall spectro-spatial arrangement of the auditory cortex was found to be similar across humans (Saenz & Langers, 2014). This also means that signals created by the cochlea as a reaction to a specific frequency band will likely take the same path through the brain and terminate in the same neural cluster.

Because the basilar membrane located in the cochlea will only vibrate in an area that correlates to the perceived sound frequency, its hair cells will only emit electrochemical signals in that particular area.[25] It has been shown that the signals transmitted to the auditory cortices by the auditory nerves are directly spatially and temporally related to the frequencies perceived by the Corti (Delgutte, 1997). This means that detailed information about the frequency composition of the sound stimulus is captured and preserved along the auditory pathway until it arrives in the auditory cortices. While the brain templates that were used to build the different network sizes described in Section 5.2.3 did provide labels for their coordinates that indicated their anatomical region, this information was too broad to define the frequency-dependent locations of the input coordinates. Therefore, this research again looked at employing data directly derived from neuroscience when determining these locations.

As shown in Figure 5-7, most parts of the auditory pathway are **tonotopically organised**, meaning that the anatomical locations of the neurons are directly related to their characteristic frequencies (Saenz & Langers, 2014). Thus, the spatial (frequency-dependent) and temporal (time-dependent) characteristics of the original sound waves are preserved, which means they can be mapped to their corresponding neurons on the cortical frequency map of the auditory cortex (Dahmen & King, 2007).

The abbreviations used in the diagram identify neural processing clusters along the primary auditory pathway. These neural processing clusters and their functionality are described in detail in Section 3.2.2. For the purpose of designing the computational sound processing model described in this chapter, the most important finding is that the signals arrive at their destination, the primary auditory cortices, in a tonotopically organised fashion that can be observed in most humans. This finding has to be preserved in the model.

[25] A detailed description of this process can be found in Section 3.2.1.

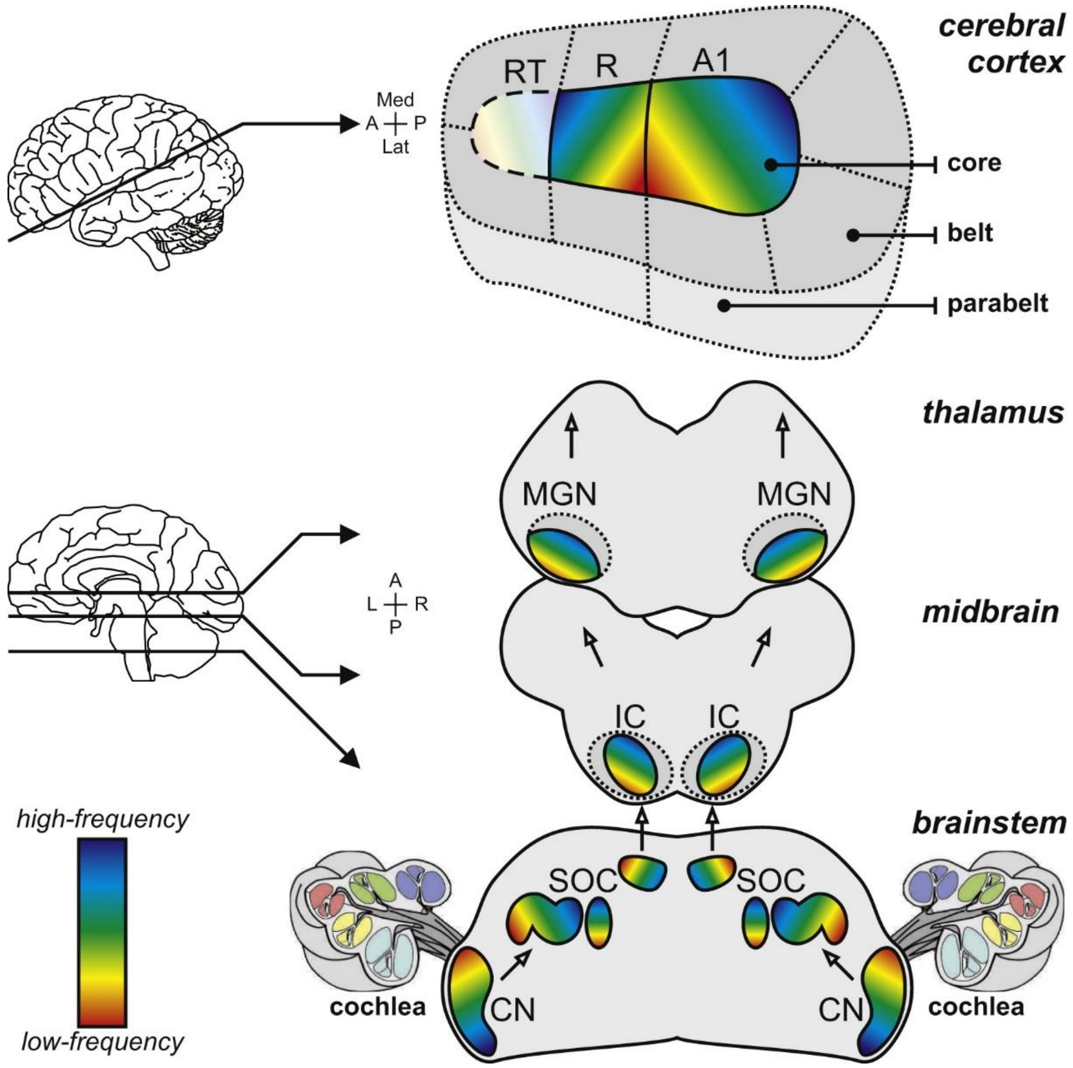


FIGURE 5-7: TONOTOPIC MAPPING OF FREQUENCIES ALONG THE PRIMARY AUDITORY PATHWAY.[26]

Tonotopic mapping has been studied using several different paradigms for measuring responses to auditory stimuli with functional imaging technologies. Langers et al. (2014) investigated how three different research protocols – continuous, clustered, and sparse signal collection – influenced the quality of the tonotopic maps. During their experimentation, the authors created several tonotopic maps, one of which they generously agreed to share with the author of this thesis so she could use it as the base for her tonotopic mapping approach. The data created by Langers et al. consisted of a list of 26,988 three-dimensional coordinates in the MNI space (14,527 coordinates in the left hemisphere and 12,461 coordinates in the right hemisphere) and their corresponding "response phase values". These phase values roughly indicate which frequencies were presented to the study subject when the corresponding brain area exhibited neural activation. The study paradigm used by Langers et

[26] From *Tonotopic mapping of human auditory cortex* by Saenz and Langers (2014, p. 43). Reproduced with permission. **RT**, rostro-temporal area; **R**, rostral area; **A1**, auditory area 1 (primary auditory cortex); **MGN**, medial geniculate nucleus; **IC**, inferior colliculus; **SOC**, superior olivary complex; **CN**, cochlear nucleus.

al. to create this map is explained below Figure 5-8, which shows a visualisation of the tonotopic map. Red colours represent smaller phase values, while blue colours represent larger phase values. Areas coloured yellow are prominent because they cover the frequency range of normal human speech and thus activate a disproportionally large number of receptive neurons.

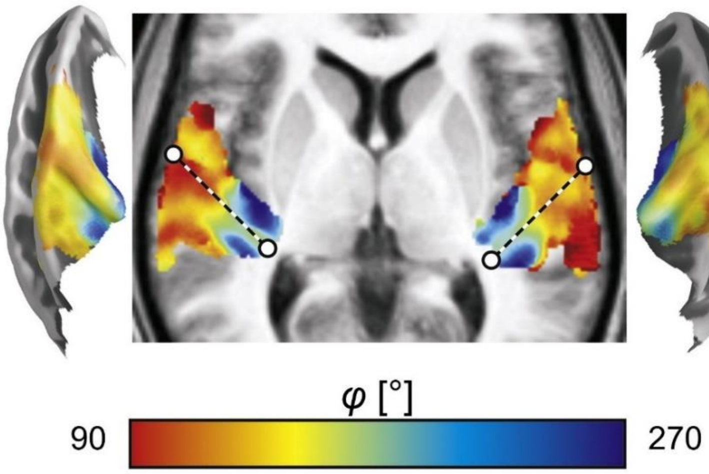


FIGURE 5-8: A VISUALISATION OF THE TONOTOPIC DATA THAT WERE USED TO MAP FREQUENCY RESPONSES INTO THE NEURAL NETWORK MODEL.[27]

The dataset was created by Langers et al. using fMRI with a sparse paradigm of collecting data points while seven participants were listening to a sweep-stimulus ranging from 125 to 8,000 Hertz. The authors found that, while all three collection protocols produced qualitatively similar tonotopic maps, the sparse data collection protocol was preferable over the other two studied methods because "it best avoids the obvious frequency-specific interference from [Acoustic Scanner Noise]" (Langers et al., 2014, p. 672). Since acoustic scanner noise has a substantial impact when measuring exact responses to sound frequencies, its exclusion during data collection is desirable.

Due to the difficulties in mapping concrete frequency stimuli to specific responses, Langers et al. employed a data acquisition technique in which the neural activation level caused by amplitude and frequency tuning is reflected in **response phases**. These response phases indicated at what time of the stimulus presentation a neural area was activated. Since the frequency composition of the stimulus was known, the phase values could be roughly traced back to some frequency values, even though no direct association was possible. The phase-frequency value pairs that were mentioned throughout the original paper are summarised in Table 5-5.

[27] Adapted from *Neuroimaging paradigms for tonotopic mapping (II): The influence of acquisition protocol* by Langers et al. (2014, p. 668). Reproduced with permission.

TABLE 5-5: APPROXIMATE MAPPING OF AUDITORY RESPONSE PHASES TO SOUND FREQUENCIES BASED ON THE DATA PROVIDED BY LANGERS ET AL. (2014).

Phase degree	Approximate frequency
0°	125 Hertz
90°	300 Hertz
120°	500 Hertz
150°	700 Hertz
220°	1,600 Hertz
240°	2,100 Hertz
270°	3,400 Hertz
360°	8,000 Hertz

For the computational sound processing system described in this chapter, concrete frequency values were needed to cover the whole sound range that was created during the cochlear encoding step. The given value pairs were, therefore, fitted to an exponential equation using the MATLAB Curve Fitting Toolbox (The MathWorks Inc., 2019a) so that the missing values could be approximated. An exponential equation was chosen because it became clear after some experimentation that it provided the best fitting characteristics when compared to, for example, linear or polynomial equations. The resulting formula is as follows:

$$frequency = 158.2 * e^{0.01092 * phase}$$

Further to transforming the data points to frequency values, they were visually analysed using Virtual Reality technology and compared to the original data shown in Figure 5-8.[28] Since the neural coordinates were located in a three-dimensional space, they were difficult to visualise on a flat computer screen. Therefore, a specialised visualisation tool was used that included Virtual Reality goggles for a fully immersive experience (Marks, 2017) to gain a better understanding of the spatial distribution of points and their values. As a result of this analysis, it was found that the given coordinates matched very well with the areas identified as the primary auditory cortices and with Heschl's gyri described earlier (Morosan et al., 2001). The phase/frequency values were also distributed as expected from the literature (Saenz & Langers, 2014).

[28] The resulting video can be found at https://youtu.be/nXuqib3B838.

The output of the tonotopic mapping step described in this section was a dataset with **26,988 coordinates** in the three-dimensional space of a brain-shaped neural network model **combined with sound frequency values** to which the neurons in these locations are likely to respond. The next section describes how this mapping was combined with the findings on the volume of the auditory cortex so that it could be adapted to different network sizes.

5.3.4 COMPRESSION OF SIGNALS

After the sound data were encoded into electrical signals as described in Section 5.3.1, they had to be compressed to account for the inflated time scale introduced by the computational requirements of the cochlear encoding module and to match them with the desired number of input channels identified in Section 5.3.2. In its biological original, the auditory system does not need such an explicit signal compression – there are many more signal-receiving neurons in the auditory cortices than there are signal-producing hair cells. However, due to the setup of the computational model presented here, a mechanism was needed to flexibly reduce the amount of data being processed. The compression was performed in two dimensions, namely frequency range and time. Since the compression by frequency range required a more complex approach, this section begins by explaining time compression first.

A requirement for using the cochlear encoding module described in Section 5.3.1 was to resample all sound files to a sampling rate of 100,000 Hertz. This meant that for every second of data, there were 100,000 data points that had to be considered by the cochlear encoding module. Initial experiments with these data showed that training the neural network with that many data points led to extremely long training times (almost a whole week for the smallest network template, which was expected to be the fastest) and frequent over-saturation of the network, which decreased the meaningfulness of the model. It was, therefore, decided to **downsample the encoded data** to facilitate the computation of larger network sizes and to speed up the training time of the model. This was done by merging every 100 time points into one, effectively downsampling the encoded data to a sampling rate of 1,000 Hertz. The scaling factor of 100 was expected to facilitate viable computation times while not losing too much information from the data. Since no literature could be found that had investigated the trade-off between sampling rate, computation time, and model performance, this factor may be revised if the experimental results were not satisfactory.

The compression by frequency range was not as straightforward as the compression by time. The outcomes of the previous three sections provided varying figures for the **number of channels** that were involved in the signal transformation:

- The cochlear module described in Section 5.3.1 defined 3,500 data streams per ear, one for each characteristic frequency, as the most biologically plausible output of the cochlear encoding, while
- the volume model described in Section 5.3.2 defined different amounts of desired input neurons ranging from four to 5,248 depending on the network size, and
- the tonotopic model described in Section 5.3.3 could map up to 26,988 data points into specific locations in the network. Using all 26,988 points as an input for the neural network would imply that the corresponding neural reservoir would need to contain almost ten million neurons to retain a biologically plausible volume ratio. Simulating a network of these dimensions was not just computationally challenging but also conflicted with the other two given figures.

Since the design of the computational sound processing system did not include a detailed model of the auditory pathway, which in nature would take care of such numerical discrepancies, a computational way had to be found that could reduce the data size to the desired amounts without losing too much information. As the hair cells typically project into a larger number of auditory nerve fibres and an even larger number of neurons in the auditory cortex, this challenge does not occur in nature. Therefore, a **novel computational mechanism** was developed as part of this research that could group and merge the signals to reduce their number for correct mapping into the defined cortical input neurons. This mechanism drew heavily on the distribution of characteristic frequencies of the 26,988 data points provided with the tonotopic dataset that was created by Langers et al. (2014).

The first step in this approach was to sort the data points of the tonotopic dataset by their phase values, which corresponded to sound frequencies so that values that were close to each other could be grouped and later merged. Visualising the sorted data points as in Figure 5-9 showed their uneven distribution across frequency bands. When trying to **select uniformly distributed points** (marked in orange in Figure 5-9) so that all frequency bands were reasonably evenly covered, two choices emerged: on one hand, neurons could be uniformly distributed based on their phase value; on the other hand, they could be uniformly distributed based on their position in the ordered dataset.

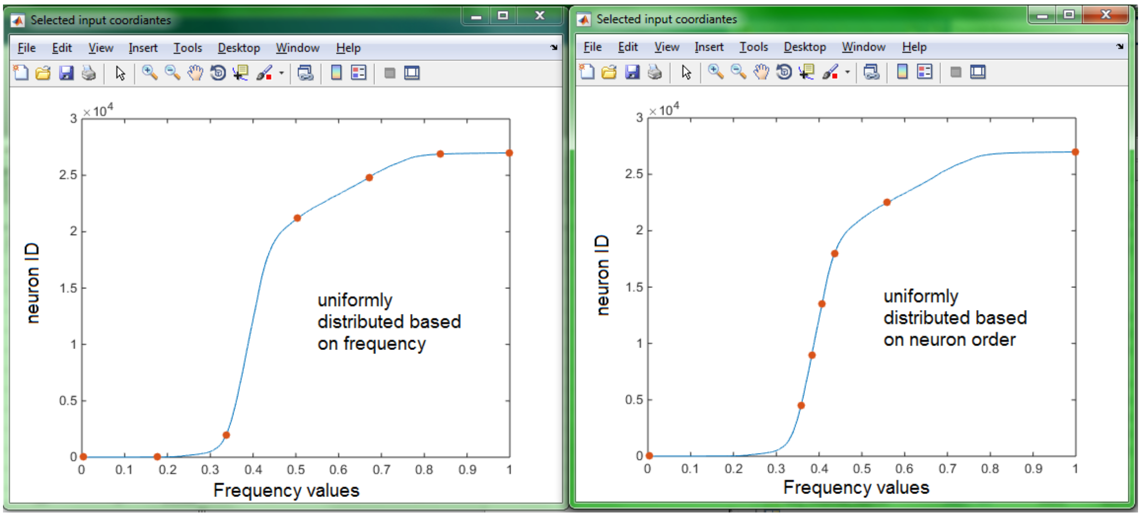


FIGURE 5-9: VISUALISATION OF THE TONOTOPIC DATASET WHERE VOXELS WERE SORTED BY PHASE VALUE. ORANGE DOTS REPRESENT VOXELS THAT WERE SELECTED WITH UNIFORM DISTRIBUTION EITHER BASED ON THEIR PHASE VALUE (LEFT PANEL) OR THEIR POSITION IN THE SORTED DATASET (RIGHT PANEL).

The second step was to decide which of these two choices was the better one. One advantage of selecting by phase value was that the distances between the values were similar, which meant they were covered evenly (see Figure 5-10 bottom panel). However, it was debatable whether this approach was biologically plausible. After all, the tonotopic dataset suggested that the number of neurons reacting to the frequency bands did not seem to be distributed evenly since the majority of the voxels had phase values between 0.35 and 0.5, which corresponded roughly to a frequency range between 626 and 1,129 Hertz. This coincided with other findings from the literature that suggested a larger representation of neurons were tuned to the frequency bands in the range of normal human speech (Delgutte, 1997). An even distribution by neuron order, and not by phase value, was, therefore, expected to be more biologically accurate. Figure 5-10 visualises how the dataset can be distributed evenly into nine sections. It can be seen that the two approaches covered the centre and the edges of the dataset to different extents. While the top panel (containing the distribution by neuron order) shows more groups in the centre and only a few larger groups in the periphery, the bottom panel (with the distribution by phase value) had one prominent group in the centre and the remaining sections were split roughly between the other eight groups.

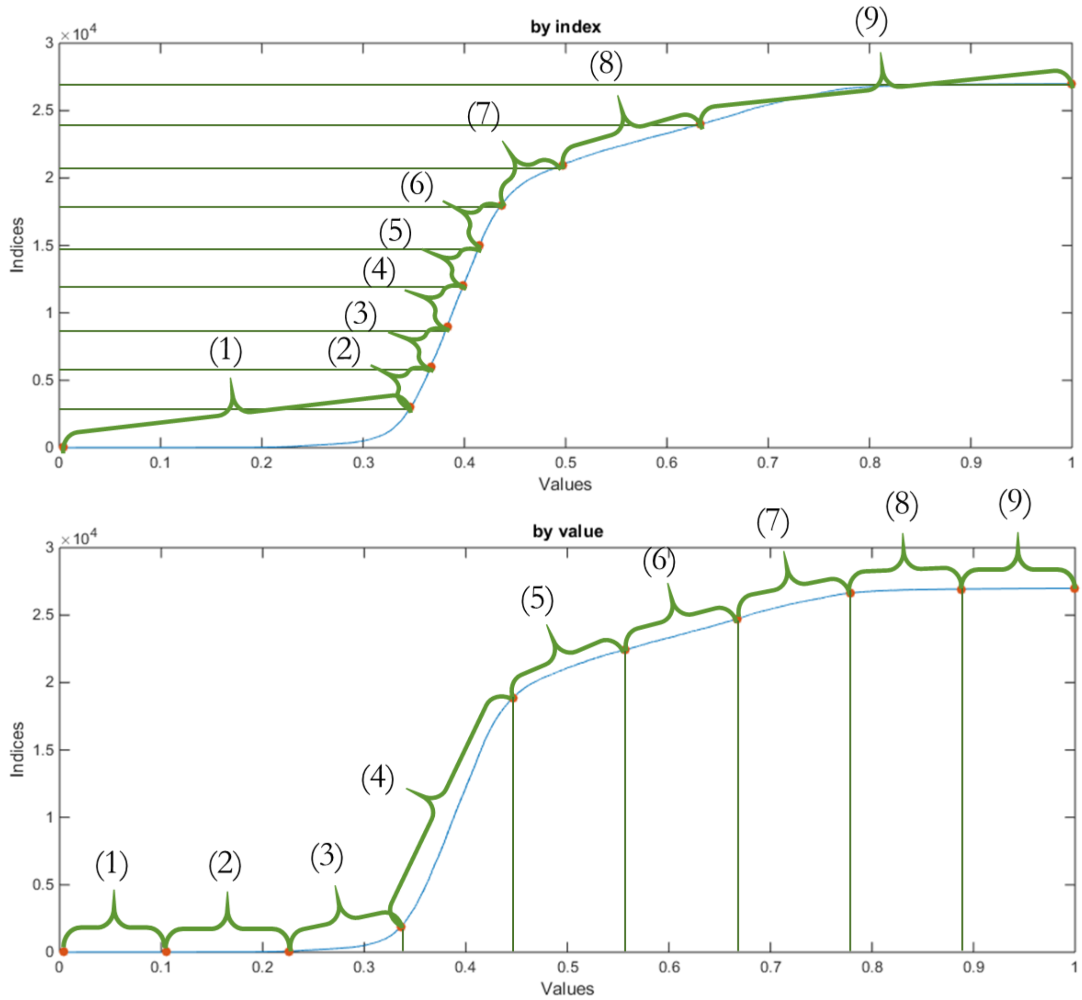


FIGURE 5-10: TWO WAYS OF DIVIDING THE TONOTOPIC DATASET INTO NINE SIMILAR SECTIONS. GREEN LINES SHOW UNIFORM DISTRIBUTION BY ORDER OF NEURONS (TOP PANEL) OR BY PHASE VALUE (BOTTOM PANEL).

Taking into consideration that most of the neurons fell into the central categories, it became apparent that the first option provided what appears to be a more biologically realistic distribution. By first ordering the voxels by their corresponding phase value and then grouping them evenly based on their order, this approach ensured that there was always the same number of processing neurons in one group. All signals emitted by the neurons in a group were then summarised so they could be used in future steps of the sound processing system that required fewer input channels than were provided by the encoded sample created by the cochlear encoding module.

But how was this approach applied to the data? Figure 5-11 shows an example of a stereo sound file of the spoken digit “zero”. For better visibility, only the input of the channels from the left ear is shown (however, the input from the right ear looked similar), and the data had already been compressed along the time dimension with a scaling factor of 100. Each black dot represents a spike or electrical impulse created by a hair cell in the computationally simulated cochlea. Point columns corresponded to time, where the first time point of the sound sample was the column on the left of the diagram and the last time point was the column on the right. Rows corresponded to the frequencies, with the lowest frequencies at the top and the highest frequencies at the bottom. The red lines show the borders between the nine groups that were identified earlier in this section and shown in the top panel of Figure 5-10 to be compressed based on neuron order.

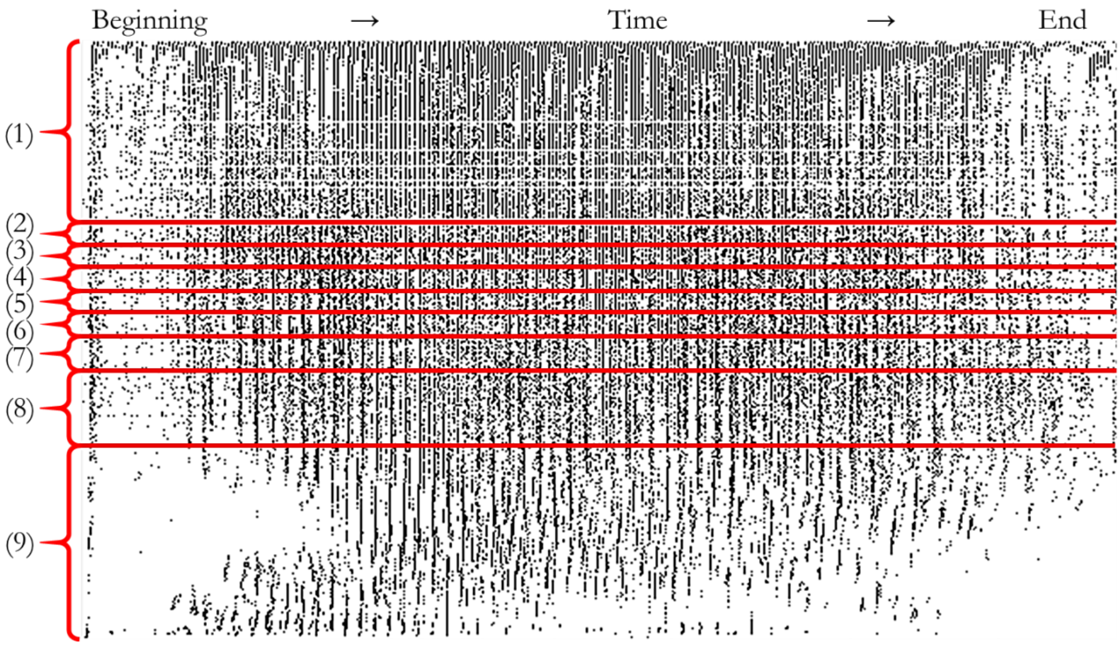


FIGURE 5-11: VISUALISATION OF DATA COMPRESSION IN A SOUND SAMPLE THAT WAS DIVIDED INTO NINE GROUPS BASED ON NEURON ORDER. RED LINES SHOW BORDERS BETWEEN THE GROUPS.

The **result** of this example compression would be a grid of nine rows and as many columns as there were in the original time-compressed sample, again containing black dots that would show the presence of an electrical impulse at a particular time point within the defined frequency range. When then entering these data into the neural network model, the nine input locations for these groups were chosen to represent the median of the frequencies covered within one group's range. The corresponding coordinates could be found easily by referring back to the original dataset by Langers et al. (2014).

For the example described above that illustrated the compression method, the number of nine groups was chosen mainly to facilitate visualisation. As described in Section 5.3.2 when discussing the most suitable number of input neurons depending on the number of neurons in the whole network, there were different sizes for brain templates available, and, hence, **different numbers of input neurons** were required. While the data shown in Figure 5-11 were based on the original MNI template with a total of 241,606 neurons, which used

338 input channels for the left ear and 318 input channels for the right ear, Figure 5-12 shows the same data compressed down to a network with 30,182 neurons that used 42 input channels for the left ear (top panel) and 40 input channels for the right ear (bottom panel). So, if instead of identifying nine groups as in the example above, the speech sample had been divided into 42 sections, the top panel of Figure 5-12 would be the result of this compression.[29]

FIGURE 5-12: SOUND SAMPLE AFTER ENCODING AND COMPRESSION DOWN TO FIT THE "MNI_BY_2" TEMPLATE WITH 42 INPUTS INTO THE LEFT EAR (TOP PANEL) AND 40 INPUTS INTO THE RIGHT EAR (BOTTOM PANEL).

A first visual comparison of both figures showed that the general shape of the signals was retained, albeit with less detail. This phenomenon was also observed when scaling down further as shown in Figure 5-13. The target template for this figure was "MNI_by_3", which contained 8,907 neurons in the reservoir and hence needed 12 input neurons for each hemisphere.

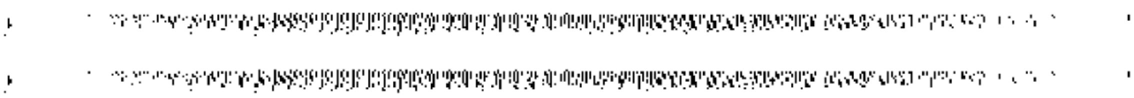


FIGURE 5-13: SOUND SAMPLE AFTER ENCODING AND COMPRESSION DOWN TO FIT THE "MNI_BY_3" TEMPLATE WITH 12 INPUTS INTO THE LEFT EAR (TOP PANEL) AND 12 INPUTS INTO THE RIGHT EAR (BOTTOM PANEL).

In practice, the time aspect of the merging process was combined with the frequency aspect so that the merging could be performed in one pass. Figure 5-14 shows an example of how the two dimensions were merged in an encoded sample of the spoken digit "zero". While the actual number of groups in this figure was arbitrarily chosen for this illustrative purpose, the process remained the same for all combinations of time/frequency pairs, with all spikes contained in a rectangle being merged into one data point of the final training sample.

[29] Since the uncompressed example presented above showed only the input into the left ear, it would be compressed using the number corresponding to the left side of the smaller template.

A rectangle of data would create a spike if its average spike rate was at least twice the average spike rate of the whole original sample, with a distinction being made between spike rates in the right and left channels of the data. This introduction of a threshold meant that regions with higher than average spike activity would always be preserved, while regions with lower than average spike activity would be filtered out, effectively performing noise cancelling on the data. The factor of twice the average spike rate was chosen because it quickly became apparent through visual inspection of the data that the threshold for creating a spike should be dependent on the spike characteristics of the original encoded sample, and it was found by visual comparison of the uncompressed and compressed sound samples that the factor two helped to preserve the characteristics of the original data. Unfortunately, no previous research results were available on this topic, so this value was chosen as a starting point for exploration and should be investigated further in future work.

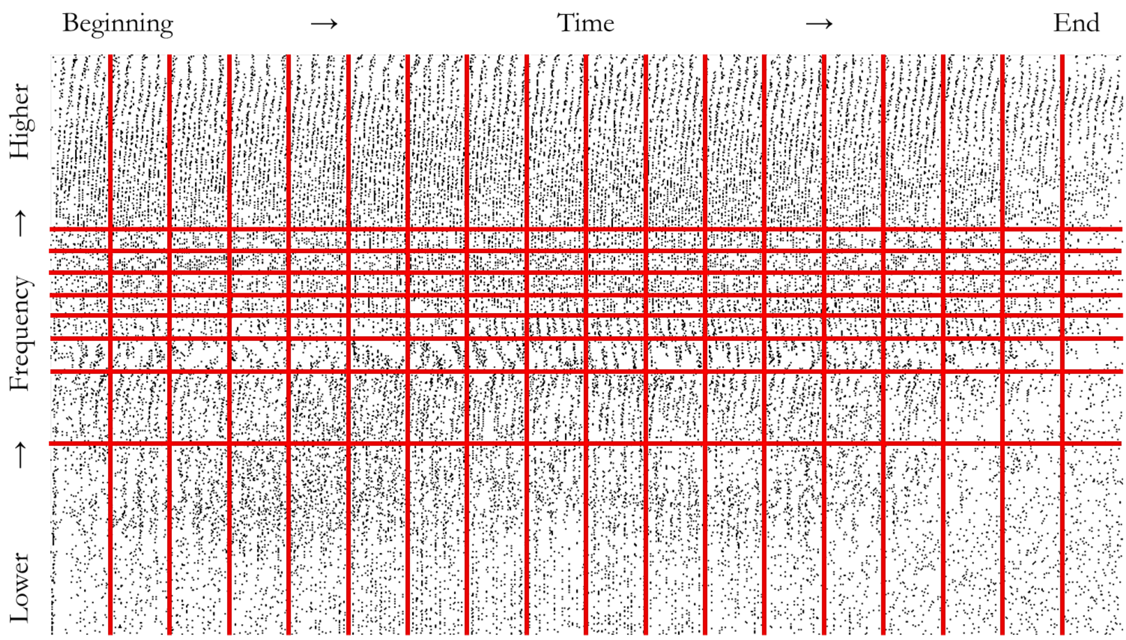


FIGURE 5-14: VISUALISATION OF THE MERGING PROCESS COMBINING TIME COMPRESSION AND FREQUENCY COMPRESSION. RED LINES SHOW BORDERS BETWEEN GROUPS.

As a final remark, an overlap of groups at their borders was considered but postponed at this stage, because the compression approach was first intended to be tested in its general functionality. Fine-tuning the algorithm by including more inspirations from nature will be the topic of future research.

5.3.5 SECTION SUMMARY AND CONTRIBUTIONS

This section presented the design of a biologically inspired sound processing system that replicated four aspects of the primary auditory pathway:

1. **Cochlear encoding**. Using the implementation of an established cochlear model, designing this part involved careful parameter selection. Based on a survey of related literature, it was found that 3,500 characteristic frequencies per cochlea ranging from 125 Hertz to 8,000 Hertz could be assumed. It was also found that using eight auditory nerve fibres per characteristic frequency with five having a high, two having a medium and one having a low spontaneous firing rate would be a computationally feasible, yet biologically plausible simplification.
2. **Number of input neurons**. It was explained that a biologically defined ratio of input to normal processing neurons had to be preserved to retain biological plausibility. This ratio was found by putting measured volumes of the auditory cortices into relation to the overall brain size. Concrete figures for the left and right hemispheres were calculated for the different network templates that had been developed for the computational model in Section 5.2.3.
3. **Location of input neurons**. A dataset containing information about neural responses to tones in a tonotopic map with 26,988 data points was obtained and transformed into a format that could be used to provide biologically plausible locations for the auditory signals. Virtual reality tools were employed to analyse the dataset and create a three-dimensional visualisation of regions in the auditory cortices that were mapped to specific frequencies.
4. **Compression of signals**. In order to overcome the differences in signal numbers created by the first three steps, this aspect focused on simulating a way of grouping and merging signals along the auditory pathway based on their response to certain frequency bands. A novel way of signal summarisation was developed that could flexibly compress input spikes with limited loss of information.

As a result of these efforts, auditory signals sent into the cochlear encoding module could then be mapped into defined locations of the computational neural network model in a biologically plausible way. The algorithms and methods presented here partly answer Research Questions 2a and 2b that were asked in Section 1.3 about signal transformation and mapping of auditory and visual data. Detailed information and results of experiments conducted with the sound processing system can be found in Chapter 6.

5.4 VIDEO PROCESSING SYSTEM

This section describes the design of a biologically inspired video processing model that was based on the human visual system. It focuses on five areas:

- the transformation of changes in light intensity into electrical signals based on the functionality of the rod photoreceptors in the retina, modelling peripheral greyscale vision;
- the transformation of changes in colour appearance into electrical signals based on the functionality of the cone photoreceptors in the retina, modelling foveal colour vision;
- the position and size of the primary visual cortex in the brain;
- the location-based (retinotopic) mapping of the electrical signals into the primary visual cortex; and
- the merging of these signals based on the organisation of photoreceptors and retinal nerve cells into visual receptive fields.

These five focus areas were identified as a suitable starting point for the design of a computational system after studying the relevant literature, which showed that they were comparatively well researched and understood by vision scientists. [30] Since they were found to play key roles in the vision process, the computational system developed as part of this research largely follows their functionality, albeit with some simplifications. Related literature from biology and neuroscience informed the design of the model and is discussed where applicable. The performance of the developed model was evaluated using benchmark data, the results of which are reported in Chapter 7.

Sections 5.4.1 and 5.4.2 explain the functioning of the retinal encoding module that was developed for the video processing system as part of this research. This module was based on the behaviour of the photoreceptors (rods and cones) in the human retina, where the rods perceive the presence of light photons and the cones perceive the presence of specific wavelengths of light depending on the type of cone. Unlike the cochlear encoding in the sound processing system, which made use of an existing computational model of the cochlea, the retinal encoding module for the video processing system was developed as an original part of this research to include colour vision capabilities (explained in detail in Section 5.4.2) and to explore a new approach of simulating receptive fields (explained in detail in Section 5.4.5). The peripheral encoding algorithm and the receptive field approach originally

[30] An introductory explanation of the human visual system can be found in Section 3.3.

introduced by Paulun et al. (2018)[31] were enhanced and refined for the research presented in this thesis. These modifications enabled greater flexibility to study different network configurations and incorporated more design decisions for the revised system that were based on biological observations.

5.4.1 RETINAL ENCODING

This section explains how the retinal encoding module transformed differences between subsequent frames of a video file into electrical impulses. It begins with an overview of the biological mechanisms that were considered during the design and implementation of the module and then outlines the process from pixels to spikes. After presenting the general encoding algorithm, this section focuses more specifically on peripheral greyscale vision, while Section 5.4.2 describes foveal colour vision.

As detailed in Section 3.3.1, in the eye, light is transformed into electrical impulses by the **photoreceptors** on the retina in a process called phototransduction. Named after their appearance, the photoreceptors reacting to the presence and absence of light are called rods, while those reacting to different colours are called cones. The biological inspiration for the peripheral greyscale encoding developed in this thesis lay in the functioning of the brightness-sensitive rods in combination with the layers of retinal neural cells to which they are connected. When light photons enter the eye, they activate a molecule in the outer segments of the rods called *retinal*. Through a cascade of chemical reactions brought on by the modified *retinal* molecules, the rod cell rapidly hyperpolarises, creating an electrical impulse. In darkness, the activated *retinal* molecules revert back to their inactive state and the rod cell depolarises again. The electrical impulses created by the rod cells are passed on through several layers of retinal neurons, including bipolar, horizontal, amacrine, and retinal ganglion cells, among others. These specialised neurons perform the task of meaningfully modifying the raw photoreceptor signals by summarising the output of neighbouring rods arranged into receptive fields and intensifying certain signal combinations while weakening others. This constitutes an early preprocessing step of visual signals before they are passed on through the axons of the ganglion cells. These form the optic nerve that relays the signals into the visual cortex, where they are processed further.

Due to the complexity of the interaction between photoreceptors and retinal neural cells, and as-of-yet unknown aspects of this interaction, the biological process could only serve as an inspiration for the visual encoding algorithm developed in this research. Furthermore, one

[31] The work published by Paulun et al. (2018) was based on a collaboration under the supervision of the thesis author.

original intention of the research presented in this thesis was to provide the opportunity to later port the developed systems to biologically inspired hardware. Since at the time of system conception, a biologically inspired cochlea chip was not yet available, the auditory processing system presented in this thesis made use of a biologically inspired software package that was unrelated to any hardware. For the visual system, however, such a hardware system was available in so-called **event-based cameras**. Specifically, the Dynamic Vision Sensor (DVS) developed by Lichtsteiner et al. (2008) was used as an inspiration when designing the retinal encoding mechanism for this research.

Event-based cameras like the DVS focus on high processing speed and hence take a less computationally expensive approach in their capture of events in a scene compared to standard frame-based cameras: Only those parts of a frame that *differ* from the previous frame by more than a defined threshold are recorded. This enables the DVS to achieve a very high temporal resolution and also operate reliably under varying light conditions while arguably capturing the most important parts of a scene (Lichtsteiner et al., 2008). This adaptability combined with the fast processing speed and focus on movement can also be found in the behaviour of the eye (Bergua, 2017). Furthermore, event-based cameras were designed to interface easily with pixel-based video formats while outputting spike data. Therefore, the functioning of the DVS system formed the base for the retinal encoding module presented here. This meant that instead of processing the absolute brightness and colour of the incoming light, spikes were created based on *changes* of brightness and colour in subsequent pixels in a video frame. For the computational system presented here, this process was further enhanced by the introduction of a block-based summarisation algorithm. The **common steps** of both the peripheral and the foveal encoding process were as follows:

Step 1. Processing all frames of the video in sequence, the differences of pixels in two successive frames were quantified. For peripheral vision, this was calculated based on brightness, while for foveal vision, a specialised colour difference formula was used. Each frame change reflected a "time step" in the output spike file.

Step 2. If the difference exceeded a predefined "pixel threshold", a "pixel spike" was created for this pixel.

Step 3. Pixel spikes that were located close to each other were then summarised into "blocks" modelled after visual fields. If the number of spikes in a block exceeded a predefined "block threshold", a spike was created for the block and recorded in the output file for the current time step.

Step 4. The block with the highest activity was set as the new focal centre of the frame as described in Section 5.4.5.

The expected input for the encoding module were videos consisting of two or more frames in any standard pixel-based format.[32] The final output file for the video was a matrix with $m \times n$ entries, where m was the number of time steps, i.e., frame changes, and n was the number of blocks. The pixel spikes were only relevant as an intermediate step and not recorded in the output file.

The encoding module was implemented in the Python programming language using the OpenCV package for video and image processing (Bradski, 2000) and the NumPy package for general data handling (Harris et al., 2020). This code can be found in Appendix A, Listing V. The code includes both greyscale and colour processing as well as the blocking mechanism since these were all performed at the same time for each video sample.

After this general description of the encoding process, the following paragraphs detail how they were applied to the aspect of peripheral greyscale vision.

Since rod photoreceptors can only process differences in brightness independent of colour, the whole image frame was first converted to greyscale using OpenCV's inbuilt functions. This process translated the three byte-based values for red, green, and blue, which each ranged from 0 to 255, into one value representing the brightness of the pixel, again in the range of 0 and 255. The following equation was used:

$$g(p) = 0.299 * p_R + 0.587 * p_G + 0.114 * p_B$$

where g is the greyscale value of the pixel and p_R, p_G, and p_B represent the red, green, and blue values of the pixel, respectively. After that, the absolute difference between the brightness of all the pixels of the two subsequent frames was calculated. One alteration between this method and the one previously described by Paulun et al. (2018) was that here, the greyscale values were not converted to a logarithmic scale before the comparison, but rather left at their absolute values. This conversion would have enhanced the processing of pixels with lower brightness and in return condensed those with higher levels of brightness. However, the human eye can adapt well to both low and bright light (Barlow, 1972; Rose, 1948), so this conversion was not applied here. The following equation was used to create pixel spikes:

$$s_p(f) = \begin{cases} 1, & g(f_{t+1}) - g(f_t) > \theta_p \\ 0, & g(f_{t+1}) - g(f_t) \leq \theta_p \end{cases}$$

where s_p is the pixel spike, f is the frame, g is the greyscale conversion equation, t is the time step, and θ_p is the pixel threshold.

[32] Any sounds contained in the video file should be encoded and processed separately by the sound processing system described in Section 5.3 and then merged back with the encoded video data using the approach described in Section 5.5.

With respect to the creation of pixel spikes, another modification to the Paulun approach was that here, no “negative spikes” were generated. While negative spikes were initially considered based on the encoding mechanism in the DVS camera, they are not biologically plausible and were hence not implemented in the final version of the encoding module. Although the retina does react differently to changes from darkness to light than vice versa, this is coded in varying spike rates rather than in the polarity of the spikes (Kuffler, 1953). Negative spikes do not have a biological equivalent and were also not used in the auditory processing system developed for this research, so it seemed a more fruitful and consistent approach to work with only positive spikes at this point.

One open question for the retinal encoding module was which value to use for the pixel threshold θ_p. A change in brightness needs to be significant enough to trigger a reaction. This significance had to be quantified with respect to the 256 grey levels that existed in the video frames. However, the eye is capable of perceiving continuous levels of grey, so fixing this value based on discrete levels is challenging and the figures reported in the literature are considerably wide-ranging. For example, the earliest calculations by König (1895) resulted in 660 different levels of brightness, although he noted that this was dependent on the hue. More recent literature stems largely from the field of medical imaging, where the grey level resolution of the screen can influence how well the images can be analysed by a physician. The numbers of identified perceivable grey levels range from 8 (Dambrosio, Amy, & Colombo, 1995) over 80 (Okkalides, 1996) to 720 (Kimpe & Tuytschaever, 2007). A formula to calculate the number of grey levels (Fetterly, Blume, Flynn, & Samei, 2008) resulted in 557 different perceivable shades of grey using the specifications of the LCD screen available to the thesis author.[33] On the other hand, two commonly reported figures are 30 and 50 (Berg, 1996; Fukui, 2001; Kreit et al., 2013). However, these could not be verified by reliable sources. Eventually, one paper was found that was deemed most relevant to the problem at hand: X. Yu, Dou, and Li (2018) studied the same 256 grey levels that were available per pixel in the encoding module presented here. The researchers changed the values at small intervals and asked 100 volunteers to indicate when they perceived a difference. X. Yu et al. (2018) found that around 85 levels of grey could be perceived. With 256 available levels of grey, these results indicated that a **peripheral pixel threshold** of 3 would be a reasonable value to be used in the encoding module of the visual processing system developed in this thesis.

[33] https://www.philips.co.nz/c-p/241B4LPYCB_75/brilliance-lcd-monitor-led-backlight-with-powersensor

5.4.2 SIMULATING COLOUR VISION

Within the context of the video processing system introduced here, the area in the focal centre of the video frames was processed using a novel computational model that was inspired by the functioning of the human fovea and developed as part of this research. In its biological counterpart, the photoreceptors in the foveal region of the retina are mainly cones, which react to changes in colour or, more precisely, to combinations of specific wavelengths of incoming light. Based on this mechanism, a computational model that simulated human colour vision was developed here that could encode colour changes of the pixels in the focal centre of subsequent video frames into spikes.[34] As explained in detail in Section 5.4.5, the location of the focal centre was updated after each frame comparison so it was centred in the periphery block that exhibited the most spike activity during the previous time step. The foveal colour encoding was only applied to the area around the focal centre, and the resulting spikes were eventually combined with those created by the peripheral greyscale encoding algorithm to form the encoded dataset.

This section explains the aspects that were considered during the development of the foveal encoding module. It first outlines contemporary understandings of how visual inputs are processed by the fovea. Conceptual colour models are then explored and their implications on the computational model developed in this research are described.

The human colour vision system has fascinated scientists for centuries. German physicist Hermann von Helmholtz first proved a previously hypothesised theory that wavelengths of light can be combined to create the impression of different colours (von Helmholtz, 1852). He further showed that there must be three types of nerve fibres in the human eye that are sensitive to red-, green-, and indigo-coloured light, respectively, with their combined perception making up all visible colours (von Helmholtz, 1867, pp. 291-294). In computing terms, the eye, therefore, uses an additive colour system, in which white light is created by the simultaneous presence of all three base colours. This is facilitated by three types of cone photoreceptors in the human retina that react to the three colours identified by von Helmholtz (Brown & Wald, 1964; A. R. Hanson, 2012). The exact wavelengths also influenced the naming of the cones (Merbs & Nathans, 1992):

- S-cones react most to light rays with a **s**hort wavelength of 426 nm (**blue** light)
- M-cones react most to light rays with a **m**edium wavelength of 530 nm (**green** light)
- L-cones react most to light rays with a **l**ong wavelength of 552 and 557 nm (**red** light)

[34] A detailed explanation of how the photoreceptors transform light waves into electrical impulses can be found in Section 3.3.1.

The electrical impulses created by the three types of cones are combined at the retinal level to form a distinguishable impression of the perceived colour (Schmidt, Neitz, & Neitz, 2014). The signals are then further processed in the visual cortex, where they are combined with information about the object's location and movement (Gegenfurtner, 2003; Solomon & Lennie, 2007). While the neural encoding of colour perception at the level of the retina is well understood, the exact mechanisms of colour processing in the cortex are still a mystery. This is mainly due to colour perception being a highly subjective sensation that is shaped by past experiences, but also because the eyes are the only sensory organ that can perceive specific wavelengths, making a more objective assessment difficult (Zaidi & Conway, 2019).

The first implementation of the foveal colour vision model developed for this research was based on the **RGB colour space**. RGB stands for Red, Green, and Blue and this scheme is based on the three colours that are perceptible by the cone photoreceptors in the human eye. Pixels in modern computer screens normally consist of three light-emitting diodes producing these colours, whereby each of the diodes can be dimmed or brightened independently. Their brightness values typically range from 0 to 255, which facilitates the display of more than 16.7 million possible colours per pixel. Due to its prevalence, it first seemed an obvious choice to use the RGB colour space for the computational model developed here. Similar to the peripheral encoding module, spikes would be created by simply comparing the differences of the RGB values of subsequent pixels to a set threshold. However, this approach was found to be not very biologically accurate, since the RGB colour space is not perceptually uniform. This meant that colour pairs with the same numerical difference could be perceived by a human observer as being more or less similar depending on their position in the colour space, a phenomenon that is caused by the different sensitivity levels of the three cone types (Gagin et al., 2014; Rabin, Gooch, & Ivan, 2011). Basing the implementation of the foveal encoding module on a colour space that did not address these non-uniform cone responses was, therefore, considered unfavourable.

While humans have three types of cones in our eyes, we can generally distinguish between *four* main hues – red, green, blue, and yellow (Neitz & Neitz, 2008). This discrepancy can be observed from early childhood and can be explained in part by higher-level processing in the lateral geniculate nucleus (Stoughton & Conway, 2008). The consideration of four main hues provides the neurological basis for several colour spaces that have been developed as an alternative to the RGB colour space with the explicit intention of being perceptually uniform (Ortiz-Jaramillo, Kumcu, Platisa, & Philips, 2019). One well-studied colour space is the **CIE L* a* b* colour space** that was developed in 1976 by the International Commission on Illumination (CIE). Like RGB, it is a three-dimensional colour space; however, it uses two axes (*a** and *b**) for describing the hue while the third axis corresponds to lightness (*L**) as shown in Figure 5-15. The values on the *a** axis are defined by the CIE standard to range from red on the positive end to green on the negative end, while the values on the *b** axis range from yellow on the positive end to blue on the negative end. With increasing distance from zero, the represented colours become more saturated. The *L** axis commonly ranges from 0 to 100, with 0 standing for absolute black (the absence of light) and 100 corresponding to the brightest white (Sharma, 2003).

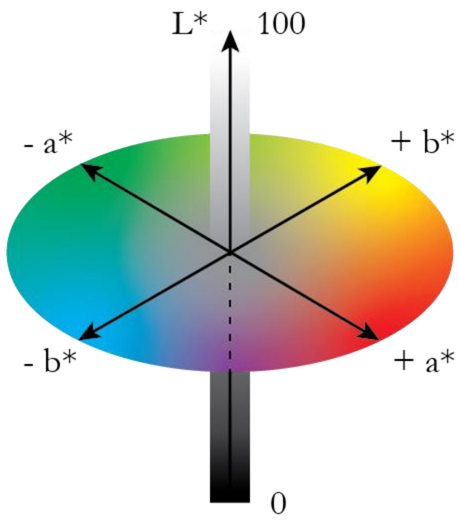


FIGURE 5-15: DIAGRAM OF THE CIE L* A* B* COLOUR SPACE.

The difference between colours, more formally described as **colour distance**, is the magnitude and character by which two colours can be distinguished under specified conditions. It is formalised as ΔE, or Delta E, where the letter E stands for the German word Empfindung, meaning sensation. The CIE L* a* b* colour space defined in 1976 was intended to be perceptually uniform, which meant that the ΔE between two colours in this space could be computed by simply calculating the Euclidian distance between the pair's colour coordinates. However, it was subsequently found that this colour space did not completely agree with human perception, particularly in the blue regions (Sharma, 2003). As a consequence, the CIE revised the colour distance formula (but not the colour space), once in 1994 and, after discovering further minor inaccuracies, again in 2000. This latest formula, called CIEDE 2000 (Luo, Cui, & Rigg, 2001), employs several scaling parameters to address the identified non-uniformity issues.

The challenges in finding an appropriate representation of colour were mainly caused by the subjectiveness of colour perception, making it a hard problem to solve objectively (Ortiz-Jaramillo, Kumcu, & Philips, 2016). Besides the initiatives by the CIE, other organisations with an interest in standardised colour representations such as the Colour Measurement Committee of the Society of Dyers and Colourists (CMC) have attempted to create more suitable colour spaces within the last few decades (Sharma, 2003). Another colour space that is little known internationally, but widely used in Germany is the DIN99[35] colour space (G. Cui, Luo, Rigg, Roesler, & Witt, 2002). However, the by far most studied colour spaces and colour difference formulas are those published by the CIE (Ortiz-Jaramillo et al., 2016). Several research groups have investigated the accuracy of the formulas in comparative studies with human participants that were asked to judge colour differences. The experience of the participants varied widely from laypeople to professionals involved in judging colours. While the CIEDE2000 formula still has unaddressed shortcomings (Kuehni, 2002; Luo, Cui, & Rigg, 2002), most comparative analyses conclude that it performs comparatively well or better than other colour distance metrics in most cases (Habekost, 2013; J.-G. Kim, Yu, & Lee, 2009; Luo, Minchew, Kenyon, & Cui, 2004; Ortiz-Jaramillo et al., 2016; S. Shen & Berns, 2011). Therefore, the implementation of the retinal encoding module of the visual processing system developed in this thesis made use of the CIEDE2000 formula to calculate the colour difference between two subsequent pixels in the foveal region of the video frame.

In the implementation of the retinal encoding module, this comparison consisted of two steps: firstly, transforming the pixels from their native RGB colour space into the CIE L* a* b* colour space and secondly, determining the colour distance using the CIEDE2000 formula. Due to the prevalence of the CIE L* a* b* colour space and its associated colour distance formulas among researchers, implementations of both were available as open-source Python modules. The OpenCV package that was used for handling the video data natively supported conversions from the RGB colour space into the CIE L* a* b* colour space and vice versa (Bradski, 2000). Furthermore, the Colour-Science module (Mansencal et al., 2020) provided implementations of the most commonly used colour distance metrics such as the CIEDE2000, CMC, and DIN99 formulas. The correctness of the CIEDE2000 implementation in the Colour-Science module was verified using the test data provided by Sharma, Wu, and Dalal (2005). The resulting code for these two steps can be found in the function *get_fovea_spikes* in Appendix A, Listing V and the following equation formalises this process:

$$s_p(f) = \begin{cases} 1, & CIEDE2000\big(c(f_{t+1}), c(f_t)\big) > \theta_p \\ 0, & CIEDE2000\big(c(f_{t+1}), c(f_t)\big) \leq \theta_p \end{cases}$$

where s_p is the pixel spike, f is the focal area of the frame, c is the colour space conversion equation[36], t is the time step, and θ_p is the pixel threshold.

[35] DIN is an abbreviation for Deutsche Industrienorm, or German Industrial Standard.

[36] This formula involves multiple steps and was hence not included here. It can be found at https://docs.opencv.org/3.4/de/d25/imgproc_color_conversions.html#color_convert_rgb_lab

Having thus defined the metric by which the subsequent pixels would be compared left open the question of setting an appropriate and meaningful pixel threshold θ_p that would prompt the creation of a spike. The original CIE standard developed in 1976 defined a ΔE value of 1 as the **just noticeable difference** (JND) between two colours (Habekost, 2013). The JND and its implications on vision have been studied for more than a century, with contemporary applications for example in the area of hidden watermarking of images (Lin, 2016; Tan et al., 2019; Thongkor, Amornraksa, & Delp, 2018; J. Wang, Wan, Li, Sun, & Zhang, 2020). König (1895) first defined the JND as the difference between two colours that makes them distinguishable. Analogous to the subjectiveness of colour perception in general, the JND is dependent on the observer and especially on the observer's surroundings. A phenomenon called simultaneous cone contrast describes the influence that colours can have on each other based on their proximity (Chevreul, 1855). This phenomenon implicated that some very similar colours could only be distinguished if they were placed next to each other but not otherwise, making it difficult to define an exact and universally applicable threshold.

The challenge of determining a fixed threshold for the computational video processing system from theoretical knowledge alone sparked an unorthodox diversion in the design of the foveal encoding: A very practical application of assessing the JND could be found in the area of dentistry, where the colour of dental prostheses has to match the surrounding teeth.[37] Studies in this area have tried to quantify a JND based on different ΔE metrics. Dentists and chemists who were professionally involved in producing dental prosthesis were asked to assess pre-calculated colour differences to determine perceptibility and acceptability thresholds (Douglas & Brewer, 1998; Ruyter, Nilner, & Möller, 1987). The perceptibility threshold was defined as the difference between two colours that could be seen by the observer and the acceptability threshold was the value at which the observer would reject a prosthesis because its colour difference to the existing teeth was too large. While the acceptability threshold was found to be a ΔE of up to 3.3 colour difference units, the perceptibility threshold was much lower at around 0.4, which is seemingly in disagreement with the original definition of the JND representing a ΔE of 1. However, it should be considered that the human observers in these studies were experts who were highly trained in the area of colour difference detection and that the lighting during the study was optimised for best observation (Douglas & Brewer, 1998). Furthermore, the colours that were relevant in this application area were naturally inclined to be mainly in the yellow hue spectrum. Therefore, certain shortcomings of the discussed colour spaces in the blue hue spectrum

[37] Judging by the amount of available literature, matching the correct colour of dental prostheses is a very prolific research field.

were not applicable here, so that, for example, all three CIE formulas were considered equivalent in this research field (J.-G. Kim et al., 2009; Y.-K. Lee, 2005). In consequence, the thresholds reported by the dental professionals in these experiments could be assumed to be smaller than those of an untrained observer in a more general context.

Since the foveal encoding system developed in this research should, however, be based on the "average" human observer, the threshold that was used here was based on the dentists' *acceptability* threshold of 3.3 colour difference units and then increased slightly to account for the untrained eye. The acceptability threshold was chosen because it was assumed that the experts would only judge those differences as "acceptable" that they would deem to be *not* perceivable by a member of the public when interacting with a person wearing the dental prosthesis. For these reasons, a ΔE value of 5 colour difference units was chosen as the **foveal pixel threshold**. This threshold determined if the colour difference between the pixels in the foveal region of two subsequent frames was large enough to create a spike.

5.4.3 NUMBER OF VISUAL INPUT NEURONS

The pixel differences recorded by the retinal encoding module described in the previous two sections were entered into the neural network using a newly developed approach of signal mapping. For this, both the number and the locations of dedicated input neurons had to be defined in a biologically plausible way that made use of the characteristics of the visual processing mechanisms in the human brain. Most notably, the primary visual cortex (V1), also known as Brodmann Area 17 (Brodmann, 1909, pp. 140-142), has been described as the "gateway" to the visual system because it serves as the first entry point for the signals arriving from the retinae (Goebel et al., 2012, p. 1309). This area was, therefore, chosen as a reference region for the input of signals into the computational model, while preserving the voluminal relationship between the input region and the overall brain template to achieve higher biological plausibility. Like in the auditory system, the proportion between the volume of V1 and the volume of the whole brain should be preserved to increase the biological plausibility of the model developed here. Since the number of neurons in the full-brain templates[38] used in this research varied widely between a few thousand and a few million, the number of input neurons for the visual computational model had to be scaled accordingly.

This section looks at the size of V1 in relation to total brain volume, while Section 5.4.4 describes how their locations were derived. The number of input neurons in the computational model was **scaled proportionally** to the total number of neurons in the

[38] A detailed explanation of different brain template sizes is provided in Section 5.2.3.

network, based on the characteristics of its biological counterpart. While most studies investigating the volume of V1 also estimated the number of neurons in this area, these estimates were usually based on cell densities, which were not considered in the model developed here. Instead, all neurons in the SNN model were evenly spaced apart for computational simplification. To circumvent complications arising from varying cell densities in different regions of the brain, which could influence neuron numbers, this section looks at volume measurements in relation to the overall brain volume, rather than using neuron numbers and putting these in relation to the estimated number of neurons in the brain.

Several studies have attempted to **measure the volume of V1** in cohorts with varying numbers of human subjects. While some researchers investigated differences between age groups (Bush & Allman, 2004; Klekamp, Riedel, Harper, & Kretschmann, 1991; Leuba & Kraftsik, 1994a; G. M. Murphy, 1985), genders (Amunts et al., 2007), or healthy people versus patients with neurological disorders (Dorph-Petersen, Pierri, Wu, Sampson, & Lewis, 2007; Leuba & Kraftsik, 1994b), others tried to relate the size of the visual cortex to brain structure and cognitive ability (Andrews, Halpern, & Purves, 1997; Bergmann, Genç, Kohler, Singer, & Pearson, 2014; de Sousa et al., 2010). Most of these researchers reported the results of their volume measurement in detail in their papers, and five of these papers were chosen as the base for the volume estimation. Table 5-6 shows an overview of their resulting figures and a short description of the underlying cohort for each study. Where this information was available, only data for neurologically healthy adults were selected here.

Further literature reporting on the size of V1 was not included in this overview because it either reused data from previously published papers that were already considered here (de Sousa et al., 2010; Leuba & Kraftsik, 1994b), did not correct for shrinkage caused by post-mortem dehydration (G. M. Murphy, 1985), or provided inconclusive information on the origin and characteristics of the subjects (Bergmann et al., 2014; Bush & Allman, 2004).

The overall volume for each hemisphere was calculated as the simple arithmetic mean of all figures reported for that hemisphere, regardless of the number of study subjects. This approach was chosen because the preparation and measurement techniques varied between studies, so calculating a weighted mean using the number of study subjects would have given more significance to a particular methodology that happened to have more data.

TABLE 5-6: OVERVIEW OF MEASURED VOLUMES OF THE HUMAN PRIMARY VISUAL CORTEX (BRODMANN AREA 17) IN LITERATURE.

Source	**Volume of V1 in cm^3**	**Data basis**
Klekamp et al. (1991, Table 2)	6.53	Mean volume for the *right* hemisphere of 19 subjects
Leuba and Kraftsik (1994a, Table 1)	5.645	Mean volume for the *right* hemisphere of 18 subjects older than 16 years
Andrews et al. (1997, Table 2)	5.69278	Mean volume for the *right* hemisphere of 15 subjects
Andrews et al. (1997, Table 2)	5.11955	Mean volume for the *left* hemisphere of 14 subjects
Amunts et al. (2007, Table 1)	7.591	Mean volume for the *right* hemisphere of 10 subjects
Amunts et al. (2007, Table 1)	7.653	Mean volume for the *left* hemisphere of 10 subjects
Dorph-Petersen et al. (2007, Table 2)	5.99	Mean volume for the *left* hemisphere of 10 healthy subjects
	6.364695 **6.254183**	**Mean volume of right hemispheres** **Mean volume of left hemispheres**

Results were usually reported separately for the left and right hemispheres due to a small but noticeable size difference. Although the computational model developed in this doctoral research did not distinguish between left and right visual cortex, they were kept separate for the calculation of the number of input neurons. Unlike the auditory system, the biological counterpart for the visual system employed a location-based mapping, which meant that the location of the stimulus in the visual field determined where in V1 it was processed. [39] Hence, a distinction between hemispheres was not necessary at this level. The algorithmic mean of the measured volumes in the five studies was about 6.254 cm^3 for volumes measured in the left hemisphere and about 6.365 cm^3 for the right hemisphere, totalling a volume of about **12.619 cm^3 for V1**. This cross-study mean was based on the brains of 72 neurologically healthy subjects between 17 and 93 years of age.

[39] This is explained in detail in Section 5.4.4.

The layout of the neural network that was used for the computational model presented in this thesis was based on the full human brain. Therefore, the volume of V1 as the input region was considered in relation to the **whole brain volume** to calculate the number of input neurons. Similar to the studies mentioned above that determined the volume of V1, whole-brain measurements have also been undertaken by neuroscientists. For example, Allen et al. (2002) describe studying 46 brains using magnetic resonance imaging to investigate volume differences between the sexes. They found that, on average, their studied brains had a volume of **1,202.35 cm³**. This figure was used in the design of the visual system presented here after it had also been used for the auditory system described in Section 5.3.2.

In order to calculate the number of input neurons for the different brain templates developed for this research, the overall brain volume was scaled in proportion to the mean volume of V1 using the following formula:

$$\frac{vol\ V1\ (total\ mean)}{vol\ Brain\ (Allen)} = \frac{number\ of\ input\ neurons}{number\ of\ reservoir\ neurons}$$

In total, V1 was found to make up about 1.05% of the total brain volume. Combining this figure with the different network sizes generated for this research, as described in Section 5.2.3, resulted in concrete numbers for input neurons which are summarised in Table 5-7. The table contains the numbers of all standard processing "reservoir" neurons for the respective templates and then three figures for the number of input neurons, where the total number had to be split into two categories to distinguish between foveal and peripheral neurons based on the origin of their respective signals – greyscale encoding for the periphery and colour encoding for the fovea.

TABLE 5-7: NUMBER OF VISUAL INPUT NEURONS DEPENDING ON THE SIZE OF THE BRAIN TEMPLATE FOR THE MNI ATLAS.

Brain template	Number of reservoir neurons	Number of input neurons		
		Total	Foveal	Peripheral
MNI_times_2	1,932,848	20,286	10,176	10,110
MNI_orig	241,606	2,536	1,292	1,244
MNI_by_2	30,182	317	156	161
MNI_by_3	8,907	93	42	51
MNI_by_4	3,747	39	20	19
MNI_by_5	1,939	20	12	8

TABLE 5-8: NUMBER OF VISUAL INPUT NEURONS DEPENDING ON THE SIZE OF THE BRAIN TEMPLATE FOR THE TALAIRACH ATLAS.

Brain template	Number of reservoir neurons	Number of input neurons		
		Total	Foveal	Peripheral
TAL_orig	1,527,747	16,034	7,990	8,044
TAL_by_2	192,600	2,021	1,020	1,001
TAL_by_3	56,770	596	288	308
TAL_by_4	23,550	247	132	115
TAL_by_5	12,150	128	64	64
TAL_by_6	7,199	76	36	40
TAL_by_7	4,452	47	25	22
TAL_by_8	2,960	31	16	15
TAL_by_9	2,086	22	12	10
TAL_by_10	1,525	16	9	7

While the total number of neurons was calculated using the ratio formula, the figures for foveal and peripheral neurons were derived when developing the block summarisation described in Section 5.4.5. In general, however, the numbers for foveal and peripheral neurons were based on the principle of **cortical magnification** (Wässle, Grünert, Röhrenbeck, & Boycott, 1989). This phenomenon describes a characteristic of the spatial organisation of the visual cortex that was discovered through brain imaging studies: Signals from the fovea, which is very small but the most sensitive area of the retina, are processed by a relatively large portion of the neurons in V1 in the posterior region of the calcarine sulcus. In contrast, the larger, more peripheral areas of the retina are processed in progressively smaller and more anterior regions of the calcarine sulcus. In a retinotopy study that investigated a large spatial extent of the visual field, the authors concluded that the central ten degrees of the visual field are mapped into around 50% of the surface of the primary visual cortex (Jinglong Wu, Yan, Zhang, Jin, & Guo, 2012). Their findings are visualised in Figure 5-16.

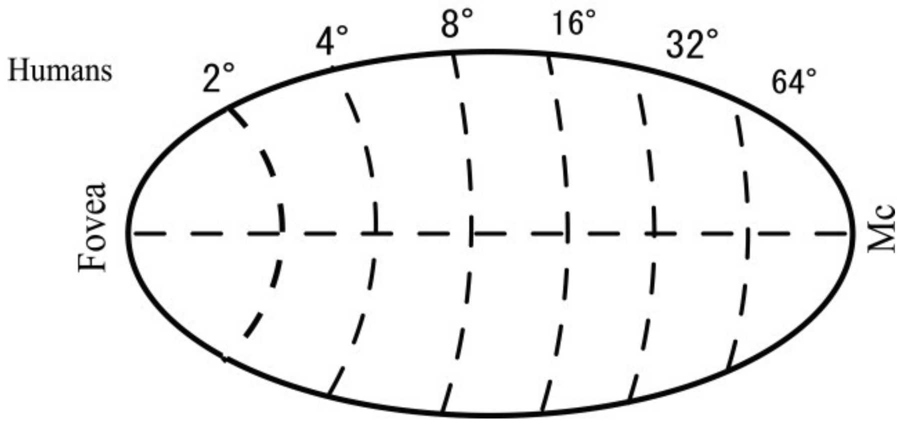


FIGURE 5-16: CORTICAL MAGNIFICATION IN HUMAN PRIMARY VISUAL CORTEX.[40]

Taking the principle of cortical magnification into account for determining the numbers of input neurons per template, about half of the calculated input neurons were designated as foveal input neurons processing the signals from the colour encoding module, while the remaining half were used to process the output of the greyscale encoding module. These figures are shown in the two right columns of Table 5-7. The slight discrepancies between the given numbers and what would be the actual half were caused by the design of the blocks that were used for signal summarisation as described in Section 5.4.5.

After the number of input neurons in the network had thus been determined, their exact locations also had to be found. The next section describes this process.

5.4.4 RETINOTOPIC MAPPING – LOCATION OF VISUAL INPUT NEURONS

While the previous section talked about *how many* neurons should function as input neurons in the neural network model, this section describes *where* they should be located. Like for the auditory system, where tonotopy determined which sound frequencies were processed in which regions of the auditory cortex, the term **retinotopy** describes the mapping of signals from specific regions of the visual field into distinct areas of V1 and further into higher visual processing areas (Goebel et al., 2012, p. 1309). However, the different levels of visual cortex differ in their spatial tuning – while V1 is clearly retinotopically organised, this structure is not as well defined in higher visual cortical areas (Henriksson, Karvonen, Salminen-Vaparanta, Railo, & Vanni, 2012). Therefore, the focus of the retinotopic mapping developed for this research lay on V1.

The first systematic observations between the location of objects in the visual field and their projection into cortical areas were made in the early 20th century by examining soldiers with

[40] From *Retinotopic mapping of the peripheral visual field to human visual cortex by functional magnetic resonance imaging* by Jinglong Wu et al. (2012, p. 1738). Reproduced with permission.

brain injuries. For example, Inouye (1909) studied 30 Japanese soldiers that had been wounded in the Boxer Rebellion and the Russo-Japanese War. Inouye systematically aligned regions in the soldiers' brains that had been affected by bullet wounds with impairments in their visual field. Similarly, Holmes (1918) analysed 16 British soldiers that had suffered head injuries in World War I and lost vision in parts of their visual field. Both researchers concluded that there must be a spatial relationship between a location observed in the visual field and the corresponding processing area in the brain. About 80 years later, their hypothesis was verified by Engel et al. (1994) using fMRI. Since then, brain imaging techniques have been used in a myriad of studies to further refine the resolution of these retinotopic maps and to extend them to areas beyond V1 (Wandell & Winawer, 2011).

Based on the original experiments by Engel et al. (1994), the acquisition paradigm for retinotopy data typically involves two moving visual stimuli with checkerboard patterns that are presented to a study participant while measuring their brain activity (Wandell & Winawer, 2011). One stimulus is an "expanding ring" that increases and decreases its size around the centre of the presented screen, while the second stimulus is a "rotating wedge" that moves either clockwise or anticlockwise around the central point. Because both the position of the presented stimuli and the location of the activated neural clusters are known, they can then be temporally aligned to create the retinotopic map. This process is illustrated in Figure 5-17.

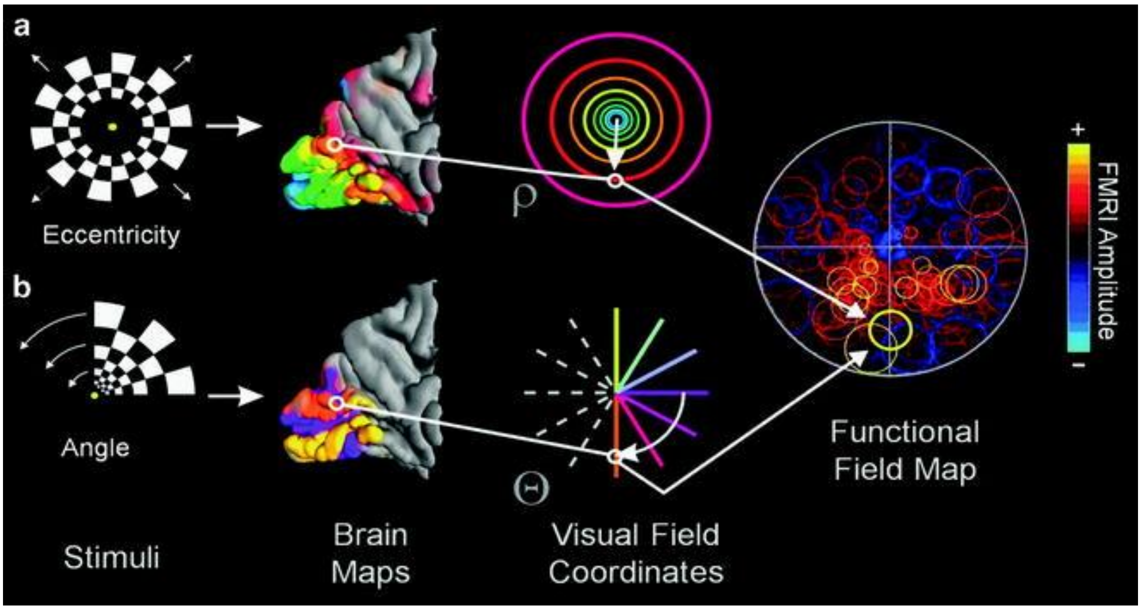


FIGURE 5-17: RETINOTOPIC MAPPING PARADIGM AND ITS RELATION TO THE VISUAL FIELD.[41]

The top part of the figure shows the mapping of the eccentricity of stimuli, i.e. how far away from the centre they are located, while the bottom part of the figure shows their polar angle,

[41] From *fMRI of Human Visual Pathways* by DeYoe et al. (2012, p. 487). Reproduced with permission.

i.e., their circular distance from a fixed axis in the visual field. With respect to the polar angle, the arrangement of objects in the visual field is both flipped from left to right and turned upside down. Thus, objects appearing in the left top corner of the visual field are processed in the right inferior area of V1. Likewise, the eccentricity processing is "inverted", meaning that the smaller areas in the centre of the visual field take up relatively more volume in V1 than the larger peripheral areas, as shown in the Brain Maps column in Figure 5-17.

When designing the video processing system presented in this thesis, the findings from retinotopy studies were taken into consideration for the spatial arrangement of the input neurons. This way, the signals created by the retinal encoding module could be retinotopically mapped into the network model. However, the distortions occurring in biological retinotopic maps due to factors such as cortical magnification were found to be quite significant, so to increase the biological plausibility of the overall model, it was decided that they had to be preserved as accurately as possible. For the research presented in this thesis, both the pixel locations of the input video data and the coordinate locations of the neurons in the network were known at this point because they had already been developed. The next step was thus finding a method to connect these two sets of location data in a meaningful way and pick those neurons that would serve as input neurons. While there were algorithms available to determine the relationship between visual field and V1 (Benson, Butt, Brainard, & Aguirre, 2014), an approach employing brain imaging data was deemed the more fruitful approach. Not only would this help to eliminate any potential geometric inaccuracies, but it also seemed more straightforward to tap into the myriad of available retinotopy data from previous research to build a mapping for the model presented here.

The retinotopy data that were used to create the mapping between video pixel positions and visual input coordinates were collected as part of the Human Connectome Project (HCP)[42]. A sub-group of 181 HCP study participants were scanned using very high resolution fMRI equipment while viewing a retinotopy paradigm (Benson et al., 2018b). Besides the previously described "expanding ring" and "rotating wedge" stimuli, the researchers collecting the data also used a "moving bar" stimulus where a rectangular block moved diagonally through the visual field of the study participant. Benson et al. then analysed the data of the 181 participants using a population receptive field approach to build a retinotopic mapping dataset, which for each participant contained values for the polar angle, eccentricity, and receptive field size, among others. Benson et al. also created three sets of grouped data where two groups were averaged across one half each of the participants and the third group was

[42] https://www.humanconnectome.org/study/hcp-young-adult

averaged across all 181 participants. Furthermore, Benson et al. created three "model fits" that covered either the first half, the second half, or all of the recorded fMRI data.

For the visual processing model developed here, the retinotopic map from Benson et al. (2018b) that was averaged across all participants and covered the whole data collection was used as a base to determine the best locations for the input neurons. This map contained 91,282 so-called **grayordinates** located across the brain and their corresponding values for polar angle and eccentricity to which they showed the highest level of neural activation. Grayordinates describe points on the surface layers of the grey matter in the brain (Glasser et al., 2013). They were introduced as a concept by HCP researchers to decrease the storage space of the data compared to conventional formats such as NIfTI files and to provide a more accurate way of comparing brain data from different people. Since everyone's brain is shaped slightly differently, aligning cortical locations across individuals would otherwise be difficult (Brett et al., 2002). However, due to their definition as surface points, grayordinates do not intrinsically hold information on volumetric coordinates unless they are associated with a projection area provided in a separate file. Therefore, the same grayordinate data can be displayed in different volumetric spaces, such as an averaged brain, an inflated brain, or a sphere. In the research presented here, a volumetric projection area representing the cortical white matter averaged over all participants and aligned to the MNI atlas was used to assign three-dimensional coordinates to the grayordinates. This meant that the resulting coordinates could be easily aligned with the network templates created to study the influence of neural network size.[43] From the resulting list of coordinates and corresponding polar angle and eccentricity data, only those entries were selected where the grayordinate had been labelled as being located in the primary visual cortex. This selection was based on an atlas developed by L. Wang, Mruczek, Arcaro, and Kastner (2015). The code for this alignment and selection is provided in Appendix A, Listing VI. The data and their values for the coordinates are visualised in Figure 5-18.

[43] This process is described in Section 5.2.3.

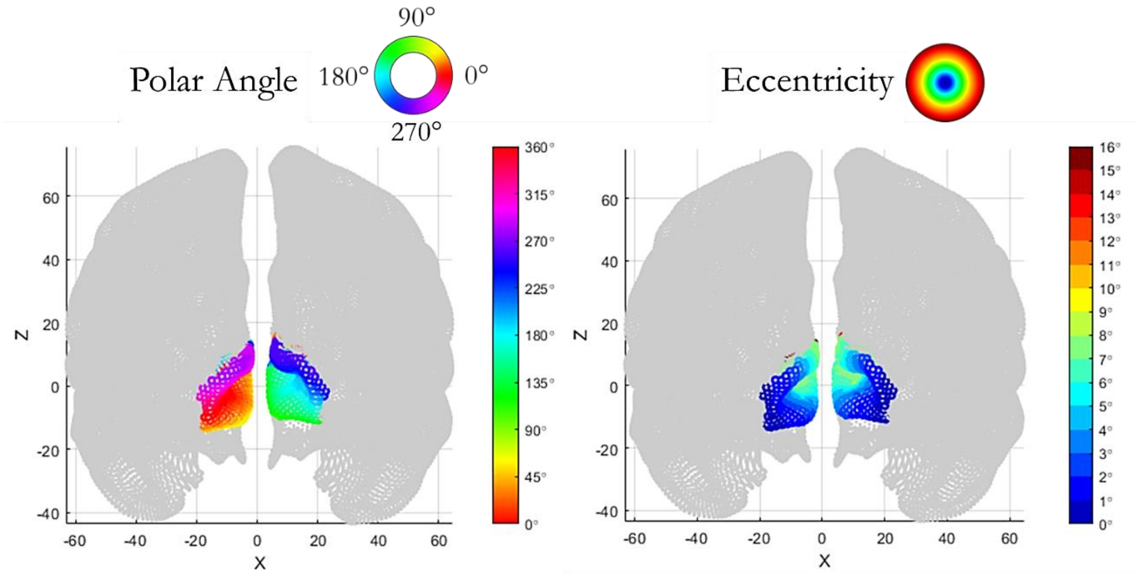


FIGURE 5-18: THE COORDINATES USED FOR THE RETINOTOPIC MAPPING AND THEIR CORRESPONDING VALUES FOR POLAR ANGLE AND ECCENTRICITY. BOTH BRAINS ARE SHOWN FROM THE BACK VIEW.

The values for the polar angle were given in degrees ranging from 0 to 360 and the starting point for this measurement was the positive x-axis with values increasing in an anticlockwise direction. The eccentricity values were given in degrees of the visual field, where 8 degrees corresponded to 100 pixels of the originally presented stimulus (Benson et al., 2018a). Using these two values and the principles of trigonometry, an exact pixel position could be calculated for each pair of polar angle and eccentricity. The code for this calculation can be found in Appendix A, Listing VII.

Analysing the now pixel-based data showed that the origin points of the visual stimuli were not distributed evenly across the visual field, as shown in Figure 5-19. Most notably, there is an absence of points in the vertical centre of the visual field, and the density of the points decreases with increasing distance from the centre. The reason for this is that while the grayordinates were uniformly distributed across the cortical surface, the phenomenon of cortical magnification introduced a distortion of the visual field. Moreover, as can be seen in Figure 5-18, the interhemispheric gap of the grayordinates is quite pronounced, creating an absence of data for the central vertical axis of the visual field.

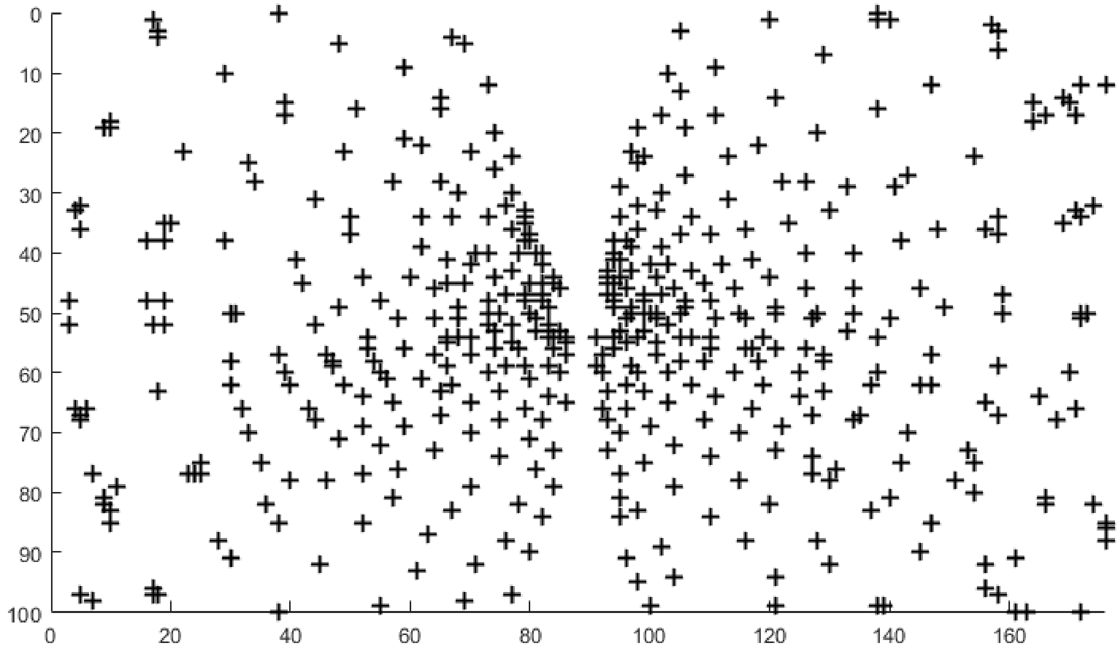


FIGURE 5-19: ORIGIN POINTS OF THE STIMULI IN THE VISUAL FIELD FOR THE RETINOTOPIC MAPPING DATASET.

As a result of the work presented in this section, the signals that were created by the retinal encoding module for specific pixels of the video frames could be mapped into the best matching input coordinate of the JNeuCube. This formalised relationship between pixels and coordinates is a major component of the visual processing system that was developed as part of this research.

5.4.5 RECEPTIVE FIELDS – BLOCK SUMMARISATION

Since the output of the basic retinal encoding algorithm presented in Sections 5.4.1 and 5.4.2 was based on pixel-wise frame comparison, the network model would need to be equipped with one input channel per pixel in the frame to handle the encoded signals. However, in accordance with the different network template sizes introduced in Section 5.2.3, these pixel-based signals had to be compressed to facilitate their entry into network models with reduced numbers of visual input neurons as described in Section 5.4.3. The approach for signal reduction that was used for the work presented in this thesis was inspired by **visual receptive fields** (VRF) found in the retina. This section introduces the biological background of VRF and the neural connectivity in the retina that forms them. It then explains how "blocks", which in this work represent the biological VRF, were defined and arranged on the video frames and the pixel-based data to effectively summarise the spikes based on their spatial organisation.

In general, receptive fields describe areas of perception that are processed in union by specialised areas in the brain (Wandell & Winawer, 2015). VRF, more specifically, are established through the arrangement of groups of neurons in the visual cortex processing light stimuli from a specific area in the visual field (Wandell & Winawer, 2015) as shown in Figure 5-20. The areas covered in the visual field are generally circular and while their size increases towards the periphery, their number decreases (Smith, Singh, Williams, & Greenlee, 2001). This phenomenon applies not only to the signal processing in the primary but also in higher visual cortices (Wandell & Winawer, 2015).

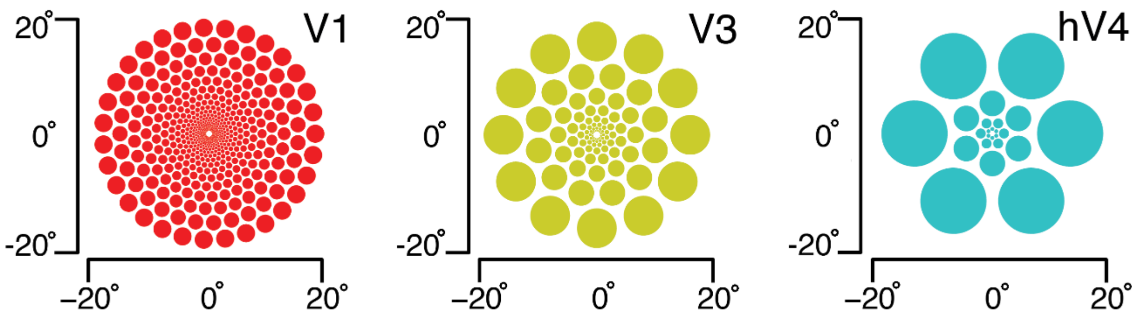


FIGURE 5-20: RECEPTIVE FIELDS IN DIFFERENT LEVELS OF THE VISUAL CORTEX.[44]

The principles of VRF were applied to the research described in this thesis in the form of "blocks" that summarised the signals created by the retinal encoding module, with some simplifications. The inverse relationship between the number and size of VRF/blocks with the increasing eccentricity of the visual field was kept as an integral part of the block design. The blocks were arranged in layers where the outermost layer covered the whole video frame, while subsequent layers covered successively less of it. At the same time, the number of blocks per layer increased with decreasing eccentricity. One introduced simplification was that, while VRF were generally found to be circular, the blocks created for this research were approximated as squares. Since the input video frames would be rectangular and blocks were not designed to overlap at this stage, a square shape ensured that all areas of the frame were covered by at least one block at all times.

Figure 5-21 shows an example of the block arrangement overlaid over a background frame image. The three outermost layers (blue, green, and yellow) represent the periphery, for which the block system was designed based on squared VRF. In order to retain the square shapes of the blocks, the ratio of the numbers of columns to rows was aligned with the ratio of the width to the height of the video frame. Note that each block was assigned a unique

[44] Adapted from *Population Receptive Field Size Estimates in 3 Human Retinotopic Maps* by Winawer and Horiguchi (2015). Reproduced with permission. **V3**, tertiary visual cortex; **hV4**, human quaternary visual cortex.

number for easier identification of the signal source in the later mapping step, where the location of the block in the visual field was important.

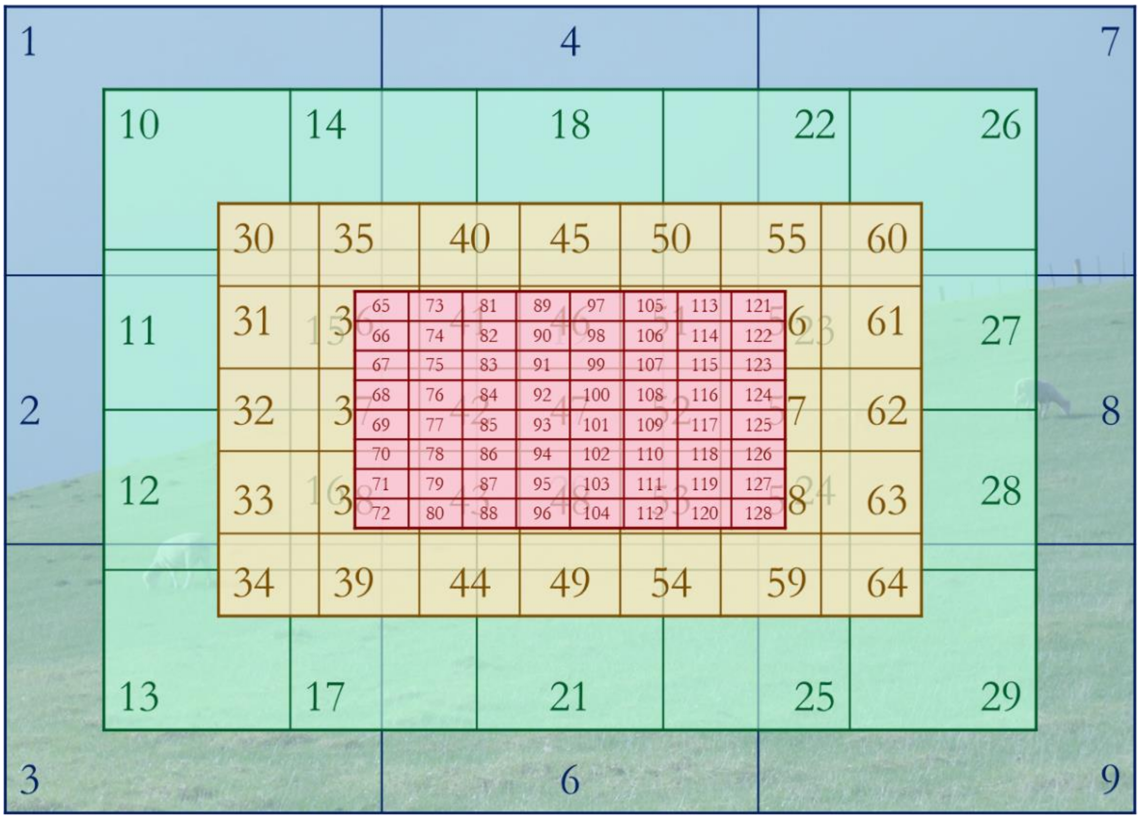


FIGURE 5-21: ARRANGEMENT OF THE RECEPTIVE FIELD BLOCKS OVER AN EXEMPLARY FRAME. THE INNERMOST LAYER OF BLOCKS (RED) REPRESENTS THE FOVEA, WHILE THE OTHER THREE LAYERS REPRESENT THE PERIPHERY.

The smallest block layer in Figure 5-21, shaded in red, marks the fovea. The blocks in this layer were purposefully not modelled as squares but instead attempted to mimic the ellipsoid shape of the fovea (Scheibe et al., 2014). Although the exact shape of the fovea varies widely (Wagner-Schuman et al., 2011), there is a noticeable length difference between its horizontal and vertical diameters. For example, Tick et al. (2011) compared 110 eyes and found that the diameter of the foveal pit was about 11% larger horizontally than vertically, while Bradley, Applegate, Zeffren, and van Heuven (1992) found this difference to be 17% for the 24 eyes they studied. Therefore, the blocks in the foveal layer were designed as oblong rectangles, with a width-to-height ratio of 1 to 0.9. This ratio was calculated as the weighted average of the two figures found in the literature.

At the same time as modelling VRF, the block approach for signal summarisation developed here was also founded on the biological principle of signal convergence across the neural layers of the retina. As described in Section 3.3.1, the **connectivity** of photoreceptors in the retina to subsequent layers of retinal neurons facilitates significant convergence (Hoon, Okawa, Della Santina, & Wong, 2014), with signals from around 120 million rods and six million cones feeding into about one million ganglion cells whose axons form the optic nerve (Goebel et al., 2012, pp. 1303-1305). In this process, the numbers of rods and cones converging into their respective ganglion cells differ by magnitudes. For example, in the very centre of the foveal dip, as few as only one of the here prominent colour-sensitive cones can be connected to a single ganglion cell, facilitating a very high spatial resolution in the centre of the visual field (Grünert & Martin, 2020; Kolb, 2012). Outside of this area, a ganglion cell typically receives input from multiple cones, although these numbers are still low (Curcio & Allen, 1990; Sterling, Freed, & Smith, 1988). On the other hand, the principle of rod convergence is that combining the signals of multiple rods into one ganglion cell leads to higher brightness sensitivity and facilitates edge and motion detection (Bruce et al., 2003, p. 29; Kolb, 2011a). Typically, several hundred rods feed into single ganglion cells via layers of bipolar and amacrine cells (Kolb, 2011a; S. C. S. Lee, Martin, & Grünert, 2019; Sterling et al., 1988).

The biological convergence of rods and cones was modelled in the block summarisation method presented here by designing the peripheral "rod blocks" to cover larger parts of the frame than the foveal "cone blocks". For computational simplification, the *divergence* of signals that is present in the biological eye was not included here. In this process, the information from rods and cones is decomposed into functional components, which are transmitted separately to several bipolar cells and then on to ganglion cells that process them in parallel (Kolb, 2011a; Masland, 2012; Sterling et al., 1988). However, since the data from the retinal encoding module had to be reduced instead of amplified, a divergence of signals was not included in the block summarisation module at this stage.

Using the general design principles of convergence and visual fields, the **number of blocks per layer** was dependent on the number of input neurons that had been identified as described in Section 5.4.3. There, it was found that the foveal blocks should contribute about half of the input signals, while in this section, their most biologically plausible shape was identified to be a rectangle with a width-to-height ratio of 1 to 0.9. When choosing the numbers of foveal block rows and columns, it was attempted to satisfy these two conditions, dubbed "half" and "ratio", as well as possible. For example, for the TAL_by_5 template with its 128 input neurons, 64 neurons were assigned to the foveal block level, which then contained eight columns and eight rows. While these did not meet the exact desired "ratio" condition, choosing slightly different numbers would impede the "half" condition to a greater extent than what would be gained by the improved ratio.

The conditions for choosing the numbers of blocks for the peripheral layers differed slightly. While the “half” condition was also expected to be met, the “ratio” condition set the desired width-to-height ratio to match that of the video frame in the input data to create square-shaped blocks. For the dataset that was used to evaluate the model, described in Chapter 7, this ratio was 16 to 9 with a width of 176 pixels and a height of 100 pixels in the frame.

Overarching over both the foveal and the peripheral block design was another condition, namely, to exactly match the calculated number of total input neurons. Meeting this target was considered most important, so slight adjustments to the other conditions were deemed acceptable. For example, for the MNI_by_3 template with its 93 input neurons, 42 were assigned to the foveal layer with seven columns and six rows, while the remaining 51 neurons were split between three peripheral layers. More exact matches for the “half” conditions of both the fovea and the periphery would have led to much worse ratios and also to a mismatched total number of required input neurons.

Finally, a plausible **number of block levels** had to found for each template. This was based on the number of input neurons. The smallest templates with less than 35 input neurons had two peripheral layers, those with up to 135 had three, those with up to 1,000 had four, those with up to 10,000 had five, and the largest templates had six peripheral layers. The final numbers for the blocks for each template are shown in Table 5-9.

TABLE 5-9: NUMBER OF BLOCKS IN EACH LAYER PER TEMPLATE. LAYER DIMENSIONS ARE GIVEN AS NUMBER OF ROWS BY NUMBER OF COLUMNS.

Template name	Total input	Foveal	Layer 0	Peripheral	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5	Layer 6
TAL_orig	16,034	**7,990**	85 × 94	**8,044**	56 × 84	36 × 55	23 × 35	15 × 23	10 × 15	6 × 10
TAL_by_2	2,021	**1,020**	30 × 34	**1,001**	20 × 30	13 × 17	9 × 12	6 × 8	4 × 6	
TAL_by_3	596	**288**	16 × 18	**308**	12 × 16	7 × 10	5 × 6	4 × 4		
TAL_by_4	247	**132**	11 × 12	**115**	7 × 9	5 × 6	4 × 4	2 × 3		
TAL_by_5	128	**64**	8 × 8	**64**	5 × 7	4 × 5	3 × 3			
TAL_by_6	76	**36**	6 × 6	**40**	4 × 6	3 × 4	2 × 2			
TAL_by_7	47	**25**	5 × 5	**22**	3 × 4	2 × 3	2 × 2			
TAL_by_8	31	**16**	4 × 4	**15**	3 × 3	2 × 3				
TAL_by_9	22	**12**	3 × 4	**10**	2 × 3	2 × 2				
TAL_by_10	16	**9**	3 × 3	**7**	2 × 2	1 × 3				
MNI_times_2	20,286	**10,176**	96 × 106	**10,110**	66 × 95	37 × 64	25 × 36	16 × 23	10 × 15	6 × 9
MNI_orig	2,536	**1,292**	34 × 38	**1,244**	22 × 33	15 × 21	9 × 15	6 × 8	4 × 5	
MNI_by_2	317	**156**	12 × 13	**161**	8 × 12	5 × 7	4 × 6	2 × 3		
MNI_by_3	93	**42**	6 × 7	**51**	5 × 6	3 × 5	2 × 3			
MNI_by_4	39	**20**	4 × 5	**19**	3 × 3	2 × 3	2 × 2			
MNI_by_5	20	**12**	3 × 4	**8**	2 × 2	2 × 2				

Following the idea that blocks represent visual receptive fields, the **size difference between subsequent block layers** was also determined in a biologically plausible way. This was based on the phenomenon of cortical magnification that already informed the number of input neurons for the different brain templates.[45] Besides the fixture that the foveal block layer should make up around ten degrees of the visual field based on the research by Jinglong Wu et al. (2012), it was also reported in the literature that cortical magnification is not linear but that it, instead, exhibits exponential size differences, which are larger for neighbouring areas further out in the periphery than for those closer to the centre of the visual field (Schira, Wade, & Tyler, 2007). This principle was applied in the development of the block summarisation method presented here. A "block scaling factor" was calculated as the *n*th root of 4, where *n* was the number of block levels minus 1.

$$block\ scaling\ factor = \sqrt[(number\ of\ block\ levels - 1)]{4}$$

Counting from the outermost layer beginning with zero as the "block level", the overall size of a block layer was then calculated depending on the block scaling factor, the frame size, and the number of blocks in the block layer:

$$layer\ width = \frac{frame\ width}{block\ scaling\ factor^{block\ level} * number\ of\ columns\ in\ layer}$$

and

$$layer\ height = \frac{frame\ height}{block\ scaling\ factor^{block\ level} * number\ of\ rows\ in\ layer}$$

which effectively led to

$$layer\ width = \frac{frame\ width}{4^{\frac{block\ level}{number\ of\ block\ layers - 1}} * number\ of\ colums\ in\ layer}$$

and

$$layer\ height = \frac{frame\ height}{4^{\frac{block\ level}{number\ of\ block\ layers - 1}} * number\ of\ rows\ in\ layer}$$

as equations to determine the size of each block level. The number 4 was chosen to emulate the exponential growth identified in cortical magnification, so it was set to the square of two since the frame is two-dimensional. Figure 5-22 shows an example of the block levels being applied to a video frame. There were three peripheral block layers (blue, green, yellow) and one foveal block layer (red). The size differences between the layers vary distinctly, with the

[45] These numbers can be found in Table 5-7 on page 150.

largest difference shown between the blue and the green layer and the smallest between the yellow and the red layer.

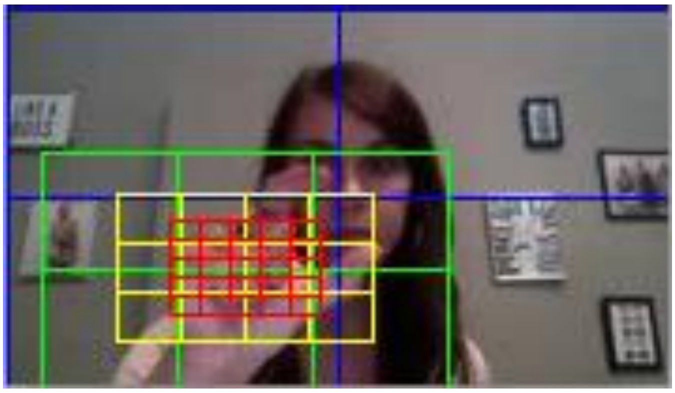


FIGURE 5-22: EXAMPLE OF FINAL BLOCK LAYERS APPLIED TO A VIDEO FRAME.

Contrary to what might be expected based on Figure 5-21, the foveal block in Figure 5-22 is not located in the centre of the visual field. This is because the focal processing centre moved towards the most active part of the frame after each frame change. The most active part of the frame was defined as the block with the highest activity, which was measured by calculating the block's spike rate, i.e., the number of spikes divided by the number of pixels in the block. The centre coordinate of the most active block was then used as the *focus* of the frame for the next frame comparison, around which all block levels were arranged. This moving frame focus was based on the biological principle of saccades, which are minuscule eye movements that direct the focus of the eye to a region of interest (Swanston & Wade, 2013, pp. 233-234).

The purpose of the block summarisation was to provide a method with which signals could be logically fed into the available **retinotopic mapping data** so that their position on the visual field determined the location of their corresponding input neuron. From the retinotopy data obtained and prepared as described in Section 5.4.4, the pixel coordinates of available retinotopy data were calculated and plotted into the visual field. The centres of the blocks were then overlaid over these data points and the spatially closest retinotopy point was found for each block centre. This process is visualised in Figure 5-23 and the corresponding source code can be found in Appendix A, Listing VIII. The best matching input coordinates for each block were found by calculating the Euclidian distance between the initial block centres and the available retinotopy coordinates. It is important to note that while the locations of the blocks moved due to the shift of the focus towards the most active part of the frame, the input neurons to which the blocks sent their signals stayed the same throughout the whole modelling process. This is because the JNeuCube implementation that was used for the research presented here allowed only fixed input neuron coordinates that had to be set up during the initialisation step of the model. For the same reason, all the videos that were chosen as input data had to have the same frame size. Changing frame sizes of video samples would influence the locations of the block centres, which would lead to different coordinates on the visual field and thus for the input coordinates in the network. The dynamic mapping of signals across V1 that is present in the human brain could hence not be modelled in the video processing system developed here, due to the restriction of the architecture that input neurons could not change their location in the network.

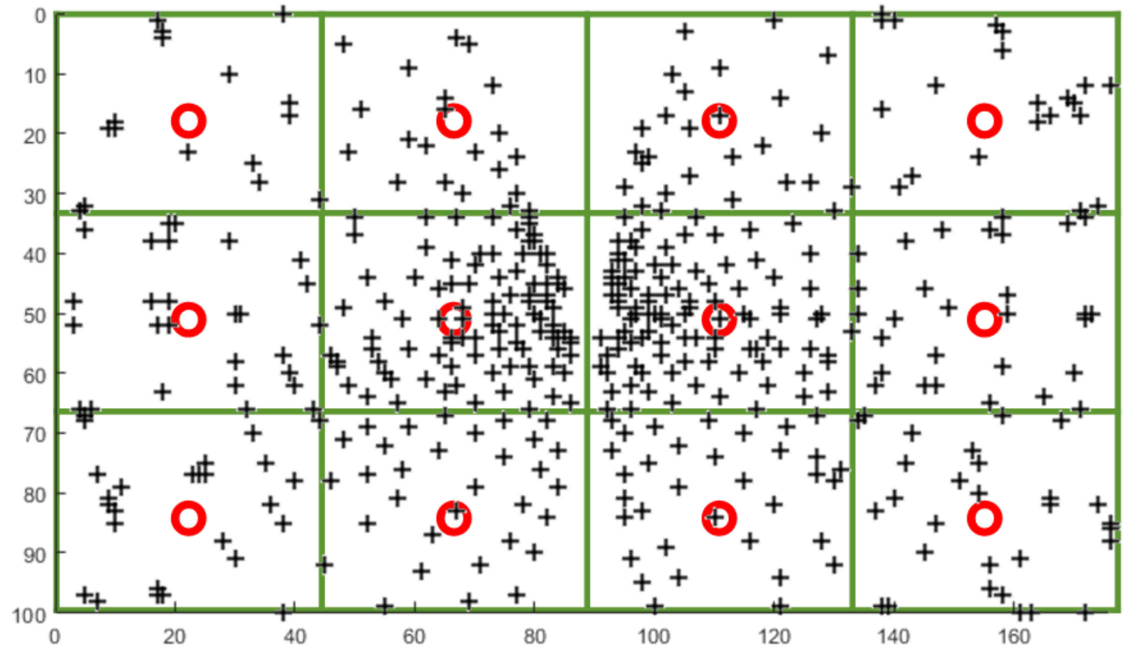


FIGURE 5-23: EXAMPLE OF THREE BY FOUR BLOCKS (GREEN) AND THEIR 12 CENTRES (RED) OVERLAID OVER THE POINTS OF THE VISUAL FIELD FOR WHICH RETINOTOPY COORDINATES WERE AVAILABLE.

The final design consideration for the block summarisation module was finding the optimum **block threshold**. This parameter describes the minimum spike rate that was required to create a spike for the block that would then be sent to its corresponding input neuron in the network. Just like cones and rods in the eye have different detection thresholds for incoming light stimuli, which resulted in separately chosen pixel thresholds as described in Sections 5.4.1 and 5.4.2, their connectivity to retinal ganglion cells differs widely (Grünert & Martin, 2020; Kolb, 2011b, 2012; S. C. S. Lee et al., 2019). This variety was reflected in the block summarisation method presented here by selecting different thresholds for the foveal and peripheral blocks. Further to the pixel thresholds that described by how much two pixels in the same location of subsequent frames had to differ to be recorded as a spike, the block threshold introduced another level of filtering that was inspired by the connectivity of the retinal neural layers (Masland, 2012). A similar block threshold had also been proposed for the auditory processing system in Section 5.3.4. For the auditory system, the block threshold was set to a value that was double the spike rate of the whole sample. Due to the architecture of the visual encoding, the overall spike rate of the sample was not available before the blocks were applied to each frame. Therefore, the threshold value was set to double the overall pixel spike rate of each frame.

Applying benchmark data to the video processing system (reported in Chapter 7) showed, however, that this initial threshold might not be the ideal choice. The search for optimum block threshold parameters is described in Chapter 7.

5.4.6 SECTION SUMMARY AND CONTRIBUTIONS

This section presented the design of a biologically inspired video processing system that replicated five aspects of the primary visual pathway:

1. **Peripheral greyscale encoding**. Based on the functioning of event-based cameras, this part of the encoding module compared the brightness levels of subsequent pixels in video frames. After converting the RGB values of the frame to a scale with 256 levels of grey, a minimum difference of 3 was set as the threshold that evoked the creation of a spike for the pixels in the frame.
2. **Foveal colour encoding**. Again looking at the differences of subsequent pixels in video frames, this module was a novel contribution that created additional spikes for the smaller focal area of the frame. After converting the RGB values to the more perceptually uniform CIE L* a* b* colour space, the colour difference between two pixels was quantified using the CIEDE2000 formula. A threshold of 5 colour difference units was set as the minimum difference that had to be reached before a spike would be created for the pixel.
3. **Number of input neurons**. It was explained that a biologically defined ratio of input to normal processing neurons had to be preserved to retain biological plausibility. This ratio was found by putting measured volumes of the visual cortex into relation to the overall brain size. Concrete figures for both hemispheres were calculated for the different network templates that had been developed for the computational model in Section 5.2.3.
4. **Location of input neurons**. A retinotopy dataset containing information about neural responses to visual stimuli of 181 people was obtained and transformed into a format that could be used to provide biologically plausible input locations for the visual signals. The relationship between the location of a pixel in the visual field and the location of the corresponding input neuron was formalised.

5. **Block summarisation**. In order to overcome the differences in signal numbers created by the encoding and the mapping steps, this aspect focused on simulating a novel way of grouping and merging signals based on visual receptive fields. An extension of previous work, the idea of a moving focal area was introduced. Block levels with foveal and peripheral blocks were defined for each of the neural template sizes depending on the number of input neurons.

As a result of these efforts, visual signals sent into the retinal encoding module could then be mapped into defined locations of the computational SNN model in a biologically plausible way. The algorithms and methods presented here partly answer Research Questions 2a and 2b that were asked in Section 1.3 about signal transformation and mapping of auditory and visual data. Detailed information and results of experiments conducted with this video processing system can be found in Chapter 7.

5.5 AUDIO-VISUAL PROCESSING SYSTEM

The final part of the system that was developed as part of this thesis was concerned with the combination of the auditory and visual processing pipelines, as was asked in Research Question 2c in Section 1.3. While sound and video data were encoded separately, their brain-inspired mapping into the same set of neural network templates facilitated a straightforward spatial integration of signals.[46] On the other hand, the temporal integration of the signals had to address differences in signal sampling rates, which meant that the encoded sound samples were much longer than their corresponding video samples and they had to be temporally aligned.[47] The audio-visual system was then evaluated using a newly created dataset that consisted of five signs from New Zealand Sign Language. The results of this evaluation are presented in Chapter 8.

5.5.1 SPATIAL INTEGRATION OF AUDIO-VISUAL DATA

The human brain is capable of combining information from multiple modalities effortlessly, although the processes of integrating these signals are very complex. As described in Section 3.4, signals from the auditory and visual pathway cross over multiple times before arriving at their respective cortices. Once there, both modalities follow a ventral and dorsal processing stream for object recognition and localisation, respectively. These multimodal interactions improve the brain's ability to interpret the signals from its surrounding environment in a holistic manner compared to only receiving input from a single modality (A. K. C. Lee & Wallace, 2019). Adding further complexity, the brain uses different neural coding schemes for auditory and visual data: while visual data are based on receptive fields and spatial encoding, auditory data are provided in a rate coding format that represents the characteristics of the sound in different firing rates (J. Lee & Groh, 2014). So-called cortical field maps then enable the brain to combine the information in a structured manner (Brewer & Barton, 2016).

For the audio-visual processing system developed as part of this research, the subcortical integration of signals along the auditory and visual pathways was not modelled at this stage since neither the unimodal auditory nor the visual processing system included a simulation of their respective pathways. This simplification was introduced because one objective of the research was to investigate concept formation, which in the human brain mainly happens

[46] This is described in detail in Section 5.5.1.
[47] An approach to overcome this discrepancy is described in Section 5.5.2.

after the signals have arrived at their cortical locations. The signal integration modelled in this research focused, therefore, on merging data at the cortical level.

The input signals were first encoded separately with their respective auditory and visual encoding modules and then fed into the same network at the same time.[48] It was hoped that by transforming the two signal types into the same spike-based format, any issues related to different coding schemes of stimulus characteristics could be averted. Since the neural network model that was used as a base for both the auditory and the visual processing models resembled the shape of the human brain, each modality could be mapped into their corresponding cortical location. While the locations of the input coordinates for both systems were based on separate tonotopic and retinotopic mapping datasets, the underlying network templates were derived from a common source, namely the Talairach and MNI templates as described in Section 5.2.3. This meant that in order to perform the spatial integration, both auditory and visual encoded signals simply had to be **mapped into the same network template** as shown in Figure 5-24. In the figure, grey dots represent the normal processing neurons in the reservoir, while orange and blue dots represent auditory and visual input neurons, respectively.

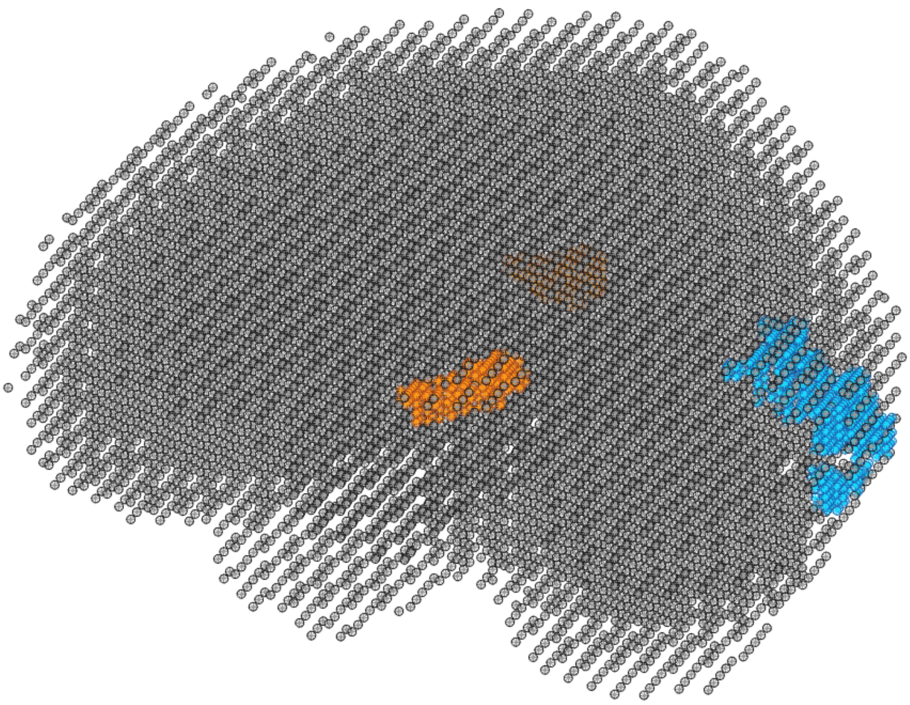


FIGURE 5-24: THE THREE-DIMENSIONAL NEURAL NETWORK WITH BOTH AUDITORY (ORANGE) AND VISUAL (BLUE) PROCESSING REGIONS FOR SIGNAL INPUT.

---

[48] The temporal aspect of this process is discussed in detail in Section 5.5.2.

In the human brain, the signals that have arrived at their respective cortices are passed on into the prefrontal cortex through two distinct processing streams, which are the ventral stream through the temporal lobes for object detection and the dorsal stream through the parietal lobe for object localisation (Milner & Goodale, 2008; Rauschecker, 2015). While these streams were not explicitly modelled here,[49] it was expected that by combining the two sets of input data into one model, they would form distinctive patterns in the network's connection weights. Investigating if these patterns resembled the biological processing streams to some degree was then seen as an interesting area of exploration for this research that was trialled on audio-visual data in Chapter 8.

5.5.2 EVENT TIME SYNCHRONISATION

Compared to the reasonably straightforward spatial combination of auditory and visual signals, their temporal alignment faced the challenge of overcoming large discrepancies between **sampling rates** of the two modalities. The cochlear encoding module required a sampling rate of 100,000 Hertz for the sound files, which was then compressed down to 1,000 Hertz by the summarisation approach described in Section 5.3.4. For a sound file with a length of one second, the encoded data file would contain 1,000 time points. The retinal encoding module, on the other hand, looked at differences between subsequent video frames, which meant it effectively replicated the frame sampling rate of the video files. For a video file with 30 frames per second and a length of one second, the encoded data file would contain 29 time points. The length of the encoded samples, therefore, differed by a factor of about 33.

In the brain, the auditory and visual pathways account for any temporal misalignments that are caused by sound waves travelling more slowly than light rays (Burr & Alais, 2006). Due to the different mechanisms by which the ears and eyes transform the stimuli into electrical signals, further discrepancies are introduced, and auditory signals are generally processed two to three times as fast as visual signals (Molholm et al., 2006). Stimuli that are temporally aligned by the brain are processed in connection with each other since close temporal proximity typically indicates a semantic relationship. The most important goal of the approach for audio-visual integration developed as part of this research was thus transferring this relationship to the network.

One way to overcome the timing discrepancies faced by the audio-visual processing system was to ensure their semantic connection was evident to the network model. This was

[49] An approach for modelling these streams would be to pre-train the network connections when initialising the model; this is briefly discussed as future work in Section 9.4.

achieved by **combining each pair** of auditory and visual encoded samples into a new audio-visual sample file. The code for this merging process is shown in Appendix A, Listing IX. Since the JNeuCube model processed each combined sample as one entity, the semantic relationship of both the auditory and the visual data in the sample was thus known to the network. However, the sample had to be provided in the shape of a rectangular matrix, which means that the length differences of the sound-video pairs had to be addressed.

Since the video samples were much shorter than the audio samples, several approaches were considered to **shrink the audio sample size** while at the same time increasing the video information length. First, all leading and trailing rows of zeros were removed from the audio files, since these represented time points at both ends of the file that did not contain any information. For the dataset that was used for the case study presented in Chapter 8, this process managed to reduce the average length of the audio sample files by about 30%. However, a visual inspection of the files from the dataset used in Chapter 8 showed that even after this step, a considerable number of files still contained leading and trailing empty rows that were not deleted due to occasional single spikes at the beginning and end of the sample. These spikes were considered irrelevant noise, so the algorithm was slightly altered to also remove rows *with* data that were immediately followed by an empty row. This process managed to reduce the average length of the audio files for the dataset described in Chapter 8 by a further 8.5%. While this process was also applied to the video files, it did not make any difference to their appearance since they contained very few or no empty rows.

Tackling the issue of the length discrepancies from another angle, three options of **prolonging the video samples** were also explored. The simplest way of increasing their length was to pad them with zeros. However, this approach was not ideal from a computational point of view. Given the still large length discrepancies between audio and video samples, the majority of the neurons in the network that were allocated to process the incoming video data would have received zeros for most of the time. This meant that instead of learning meaningful patterns, they would have learned that they were largely not needed and therefore produced only very weak connections. A second approach to "fill" the video file with meaningful data was to duplicate the whole sample and thus present the video signals to the network several times while the audio data would have been presented only once. This approach was rejected for being too biologically implausible since the brain would typically not receive duplicate visual input per singular audio stimulus.[50] The final approach that was considered was stretching the video data by a certain factor. Based on the observation by

[50] The work presented in this thesis generally assumed a sober and neurologically healthy brain for all biological comparisons.

neuroscientists that visual stimuli are about two to three times slower than auditory stimuli in eliciting a processing response in the brain (Molholm et al., 2006), this approach was deemed most promising. As a starting point for the experiments described in Chapter 8, the stretching factor was set to 3. This meant that every time point in the visual sample was repeated three times, essentially slowing down the videos while still maintaining their characteristic dynamic patterns. The stretching process of the video samples in combination with the removal of empty rows for the sound samples significantly improved the file length ratio for the dataset that was used in the case study presented in Chapter 8. Before the modifications, the audio files were on average about 14 times as long as the video files, while now, they were only about three times as long. The remaining two-thirds of the file were then filled with zeros. While this still introduced the semantic bias towards recognising zeros instead of patterns, the impact of this misalignment was less pronounced.

5.6 CHAPTER SUMMARY

This chapter described a novel computational architecture to model audio-visual data in a biologically plausible way. The design of the systems was informed by literature on biological and neurological processes, some of which were attempted to be modelled as part of this research. The system architecture followed four steps, encoding, mapping, learning, and analysis, of which the first three were explained in detail in this chapter. While the encoding and mapping steps were developed separately for auditory and visual data, the learning step involved data modelling and integration for both modalities. Finally, the analysis of the developed models was performed using benchmark datasets. These experiments are described in Chapters 6, 7, and 8.

The contributions of this chapter are:

1. **Encoding**. A set of biologically plausible parameters was found for an existing cochlear encoding module. A retinal encoding module was developed based on the functionality of existing hardware and enhanced by adding colour vision capabilities.
2. **Mapping**. For both the auditory and the visual system, the optimum number of input neurons was found depending on the size of the neural network. Furthermore, biologically plausible locations of these input neurons were determined based on tonotopy and retinotopy datasets. Finally, a novel compression algorithm for the encoded sound data was developed and an existing algorithm for the visual data was enhanced by adding a moving focal area and providing flexible summarisation boundaries.
3. **Learning**. An existing architecture for an SNN was used. A set of 16 brain-shaped network templates was developed based on existing brain atlases. A novel method for spatial and temporal integration of audio-visual data was introduced.

6 BENCHMARKING THE SOUND PROCESSING SYSTEM

"There is nothing like first-hand evidence."

– *Sherlock Holmes in A Study in Scarlet*

6.1 CHAPTER OVERVIEW

This chapter describes how the sound processing model introduced in Section 5.3 was applied to a standard benchmark dataset in the domain of speech recognition. The performed experiments aimed to evaluate the capabilities of the proposed model and to gain insight into potentially suitable model parameter configurations by comparing different setups using the same benchmark data and computer hardware. The chapter first describes the origin, content, and structure of the dataset, followed by a detailed explanation of how the dataset was analysed with the model. Subsequently, the results of the experiments are presented, and conclusions are made about the model and its optimal configuration. The chapter closes with a short discussion of the advantages and shortcomings of the model compared to other published work using the same or comparable datasets, in an effort to answer Research Question 3a that was asked in Section 1.3.

6.2 DATASET DESCRIPTION

Contemporary research into automated speech recognition began almost 70 years ago with the development of a "digit recogniser" by three then-employees of Bell Telephone Laboratories (Davis, Biddulph, & Balashek, 1952). While this circuit-based device had to be adjusted to individual speakers before it was fully functional, it could reach impressive recognition accuracies of up to 99%. The domain of spoken digit recognition has since been used widely as a study field for prototyping computational auditory and speech processing systems. Unlike trying to identify spoken phrases or sentences, where countless combinations of utterances can exist and contextual meaning might influence perception, digit recognition offers a well-defined set of words that can be represented as independent samples and transferred to any language where a counting system is known. For these reasons, it was decided to evaluate the auditory processing system described in this thesis on such a dataset.

The benchmark dataset used for this experiment was the Free Spoken Digits Dataset (FSDD) which originated as a GitHub project and was filled with content in a community effort (Jackson, Hereman, Walker, & Weveler, 2016). The raw sound files were published in WAV format and were downloaded from the GitHub repository on 17th October 2019 for the experiment described in this chapter. Contrary to the popular[51] proprietary TIDIGITS dataset, which was created as a benchmark dataset for spoken digit recognition (Leonard, 1984), the FSDD can be used by anyone for any purpose free of charge.

Like other spoken-digits datasets, the FSDD covers all ten digits from 0 to 9. Four speakers of English, who were identified in the dataset as Jackson, Nicolas, Theo, and Yweveler, recorded themselves pronouncing each of the ten digits 50 times each, creating 500 speech samples per speaker and 2,000 speech samples in total. Jackson and Theo were identified as speaking with a North American English accent, while Nicolas had a French accent and Yweveler had a German one. The samples were recorded at a sampling rate of 8,000 Hertz and were monophonic. They were between 0.14 and 2.28 seconds long, with an average length of 0.42 seconds. There was minimal background noise and all samples were clearly audible, although no information on the recording devices was available.

In this experiment, only the samples recorded by Jackson were used. This was expected to facilitate the classification of content rather than speaker voice and intonation. This pragmatic approach was considered sufficient for this initial testing of the sound processing model. Cross-speaker classification is a topic that should be considered in future research.

[51] According to the IEEE, the original paper describing the TIDIGITS dataset has been cited well over 200 times (https://ieeexplore.ieee.org/abstract/document/1172716/citations#citations).

6.3 EXPERIMENTAL SETUP

The goal of this experiment was to find the best parameter configuration for the model. Since the model's architecture offered a variety of parameters that were all to some extent related to each other and influenced the model's performance, it first had to be decided which of them to optimise based on the amount of knowledge that could be gained from studying them. This section first discusses the three steps required to enter the data into the model, followed by the parameters for the neural network itself and finally the experiment's training and testing procedure.

6.3.1 DATA PREPARATION

Preparing the sound data to be entered into the neural model required three steps that were discussed in detail in Section 5.3. The following paragraphs describe how these steps were applied in practice for this experiment.

In the first step, **converting the sound data** to electrical impulses, all parameters for the cochlear encoding module were determined based on findings from neurology research as described in Section 5.3.1. Since these parameters were founded on biological observations, they were assumed to be in their most biologically plausible state and left unchanged for this experiment. However, the software that was used to perform this step required all samples to have a sampling rate of 100,000 Hertz. This meant that the FSDD data had to be upsampled from their original sampling rate of 8,000 Hertz before any further processing could occur. Upsampling was performed using the *resample* function from MATLAB's Signal Processing Toolbox (The MathWorks Inc., 2019b), which was applied directly to the WAV files from the dataset. The code for this pre-processing is shown in Appendix A, Listing I. The resampled sound files were then converted to spike matrices using the cochlea.py Python software module by Zilany et al. (2014) with the parameters identified in Section 5.3.1. The code for the encoding step is shown in Appendix A, Listing II.

The second step, **mapping** these spikes into the brain-shaped model, was concerned with finding a biologically plausible way to insert the spikes into the network. This step covered two aspects, *number* and *location* of neurons. While the appropriate locations could be determined by looking into research results from tonotopy studies, as described in Section 5.3.3, the number of input neurons was set to be dependent on the total size of the neural network. All 16 network configurations described in Section 5.2.3 were used in this experiment with their respective number of auditory input neurons that was calculated as described in Section 5.3.2. Since the influence of the network size on the performance of the neural network model was found to be not very well understood in existing literature, it was chosen as a parameter for optimisation to be investigated in the research presented here. The available network configurations were based on two different brain atlases (MNI and Talairach) and contained between 1,525 and 1,932,848 neurons. For the varying numbers of input neurons, the frequency ranges to which the single neurons would respond had to be identified. This was done by uniformly selecting points from a list of available neurons for which frequency and coordinate information were known, as described in Section 5.3.4. While this selection process was technically part of the third step (compression), it was relevant here to set up the model. The code for this selection algorithm is shown in Appendix A, Listing III. The lists of frequency values that were created by this script were then copy-pasted into the parameter configuration of the Cochlea encoding module shown in Appendix A, Listing II.

In the third step, **compressing** the data along its two dimensions, mainly the time dimension had to be considered since the frequency dimension had already been determined by the configuration of input neurons for each model. As described in the previous paragraph, the number of channels in the encoded data was defined to match the number of input neurons in the studied template sizes. For the time dimension, a fixed compression rate of 100 was used for each model and tested as part of this experiment. This effectively created a final sampling rate of 1,000 Hertz for the data. Since this model is a first of its kind, no previous research existed that had investigated optimal values for this scaling factor. The factor of 100 was chosen because it provided a convenient way of scaling down the data to a manageable size while not losing too much information. The code for this compression step is part of the code shown in Appendix A, Listing II in the function called *scale_down*.

The three steps described above are mainly related to data encoding and input into the model. Equally important is the configuration of the network, which also had to be specified before the experiment could be run. The model parameters of the network are described in the following section.

6.3.2 NETWORK MODEL PARAMETERS

Apart from the network size that was considered as a variable in the experimental setup described here, several more parameters could be altered in JNeuCube. While a detailed discussion of the available parameters and their purpose is provided in Section 5.2.2, this section focuses on the specific values used for this experiment.

Firstly, the **neuron type** is an important factor to consider when designing the network. One of the most popular types of neurons is the LIF neuron. This very simplistic model has traded biological plausibility for computational efficiency, enabling larger networks with shorter training times (Izhikevich, 2004). Another rather simple neuron model that promises more biological plausibility than the LIF model was introduced by Izhikevich (2003). These two neuron models, LIF and Izhikevich, were taken into consideration for this experiment since they promised fast computation times, which was essential for the larger network sizes that were to be tested.

Although Izhikevich claims[52] that his 2004 models have a high level of biological plausibility due to them matching several biological features (Izhikevich, 2004), a set of preliminary experiments on the dataset studied here resulted in about 5 to 10% lower classification accuracy when compared to using the LIF neurons. These preliminary experiments were similar to the main experiment in terms of data preparation, network parameters and the model training and testing procedure, with the exception that only the smallest network template (TAL_by_10 with 1,525 neurons) was used and each configuration was run only ten times to reduce the computation time needed at this stage. While experiments using the 20 different neuron properties described by Izhikevich (2004) achieved between 52 and 64% classification accuracy, the LIF neurons tested with a set of firing voltage thresholds between 0.01 and 2 achieved between 64 and 68% classification accuracy[53]. Based on these results, it was decided to only investigate the LIF neurons further for the main experiment.

As a consequence of the findings of the preliminary neuron model experiments, the **LIF firing threshold** was chosen as a second dimension for parameter optimisation in the main experiment (besides the network template size). The values that were studied for the LIF firing threshold ranged from 0.01 to 0.5 with smaller distances between smaller values and larger distances between larger values. Previous experiments with similar experimental setups had shown that optimum LIF threshold values tended to be at the lower end of this range (Kasabov et al., 2016), so the step size between threshold values was chosen to be smaller for smaller threshold values and larger for larger threshold values. The following 18 threshold values were used:

0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50

As explained in detail in Section 5.2.2, the JNeuCube implementation contains more parameters than just the neuron type and the neural firing threshold. In order to decrease the

[52] See Section 5.2.2 for a more in-depth discussion of this claim.
[53] The detailed results of these experiments can be found in Appendix B, Table B-1 and Table B-2.

model's complexity for this experiment, however, the rest of the parameters were set to values that had proven successful in prior studies with the same system setup (Kasabov et al., 2016). An overview of these values is provided in Table 6-1, while an explanation of their meaning and rationale for default values can be found in the general description of the network parameters in Section 5.2.2.

TABLE 6-1: NETWORK MODEL PARAMETERS FOR THE EXPERIMENT ON SPOKEN DIGIT RECOGNITION.

Category	Parameter name	Value
Neuron model	LIF reset voltage	0 (simulating a full neural discharge)
	LIF refractory time	4
	LIF resistance	1
	LIF capacitance	10
Network initialisation	SWC weight range	Minimum -0.1 and maximum +0.1
	SWC radius	2.5 times the network size scaling factor[54]
	SWC positivity rate	0.7[55]
STDP	*A* positive/*A* negative	0.001
	τ positive/τ negative	10
	Weight boundaries	Upper bound +2 and lower bound -2
deSNN	Modulation factor	0.8
	Drift positive/negative	0.005
Classification	kNN *k*	3
	Number of folds	5
	Training rate	0.7

[54] For example, in the standard MNI template this would be set to 2.5, while in the "MNI_by_3" template it would be set to 2.5 × 3 = 7.5.
[55] Research by Hendry et al. (1987) showed that about 70% of neurons in monkeys' brains are excitatory.

6.3.3 MODEL TRAINING AND TESTING PROCEDURE

Besides the parameter configuration, the general setup of the experiments can greatly influence their outcomes. The procedure chosen here was intended to create a robust and replicable result by observing best-practice methods from applied machine learning (Brownlee, 2020).

First, the dataset was split into training and test data, with 70% of the samples randomly chosen from each class for the training subset and the remaining 30% being held back as the test subset. The purpose of this split was to provide an unbiased evaluation. While the training subset was used to find the best parameters for the model, the test subset was then applied only to the model with the best configuration. The classification accuracy of this final test run was then used to compare the model to other algorithms.

The training subset was further split into training and validation data during a five-fold cross-validation process. Using this approach, performance measures for the current model configuration could be acquired and compared with each other to assess the quality of the current set of parameters. Each of the five models built during the cross-validation process used four of the five folds to train the model and the remaining fold to validate the model by passing the unseen samples through the trained model and predicting their labels. For better comparability, all five models were initialised with the same connections created by the SWC algorithm. The average of the five results was then reported as the overall result for the current model configuration and set of parameters.

For each set of parameters, this cross-validation process was performed 30 times. Since the connections in the network were initialised randomly by the SWC algorithm, each of these 30 runs had a slightly different network setup. By calculating and reporting the algorithmic mean of all 30 runs, it was hoped to minimise the impact of outlier configurations that could occur in single runs. With five folds per run and 30 runs, the "averaged overall result" for each set of parameters was consequently based on the performance of a total of 150 models.

The averaged overall results for the different sets of parameters were then compared to each other. The configuration that had achieved the highest classification accuracy was chosen to perform a final experiment with the test dataset that had been held back until this point. For this, the model was trained on the whole training subset and then used to predict the labels for the unseen test subset. Like the validation process, this evaluation was also performed 30 times to create a reliable result. For better reproducibility, the 30 randomly initialised networks from the validation step were reused for the evaluation. The algorithmic mean of the classification accuracies of these 30 runs was then reported as the final overall result for the sound processing system introduced in this thesis. The result was then compared to the accuracies of existing models as described in the literature.

The execution times for each model configuration were also monitored as a performance metric. For completeness, and to enable future research, the following information should be noted: All experiments were run on a standard PC with an Intel® Core™ i7-8700 CPU with 3.20 GHz clock speed, 16 GB RAM, and the Windows 10 64-bit operating system. The JNeuCube software was run using the Apache NetBeans IDE 11.3.

6.4 RESULTS

The outcome of the experimental setup described in Section 6.3 were several sets of results for the different stages of the modelling process. Firstly, the overall average results of the 30 runs for the five-fold cross-validation were acquired. These are shown in Table 6-2 and visualised in Figure 6-1. The detailed results of the 30 runs for the different network sizes can be found in Appendix B, Table B-3 to Table B-10.

The result of the cross-validation process informed the choice of the best set of parameters for the model. Since the experiments described here could only provide an initial evaluation of the newly developed sound processing system, only two parameters were considered for the optimisation process at this stage. The first parameter was the network size, for which 16 templates were available, and the second parameter was the LIF threshold, for which 18 discrete values were chosen. These two parameter dimensions appear as columns and rows in Table 6-2. The last column and row show the average classification accuracy for the respective parameter. Cell background colours indicate how the values compare to each other, with green indicating better performance, yellow indicating average performance, and red indicating worse performance.

Figure 6-1 visualises the numbers from Table 6-2 in a line diagram. As is noticeable from both the table and the figure, the values for the classification accuracies range from about 75% to about 90%. The worst performing template size was "TAL_by_10", followed by "TAL_by_9" and "MNI_by_5". These were the smallest available template sizes. One possible explanation for their comparatively bad performance could be that the number of input neurons for these template sizes was also very low, which makes it likely that too much information was lost during the data compression step so that the meaningful features of the samples were not distinguishable anymore.

All the larger templates performed at a comparable level between 85 and 90% for LIF thresholds greater than 0.9. However, bigger discrepancies existed for lower LIF values. Notably, there was a steep increase in classification accuracy of about 7% for the MNI templates between the LIF thresholds 0.01 and 0.02, while a similar surge could be observed for the Talairach templates between the LIF values 0.05 and 0.06. Behaviourally, very low LIF thresholds would typically lead to more spiking activity in the network, since less postsynaptic potential has to be built up by incoming spikes before the neuron would fire. This "over-activity" could potentially reduce the distinguishability of the patterns that can be detected by the network. If all samples create an abundance of spikes, the significance of each of these spikes to identify an individual sample decreases.

The highest value for overall classification accuracy that was achieved during the parameter optimisation process was 89.79% for the "TAL_by_6" template with a LIF threshold of 0.3. However, the best-performing template overall was "MNI_by_3" with an average classification accuracy of 88.39% and the best-performing LIF threshold was 0.15 with an average accuracy of 86.02%. Since the purpose of the optimisation process was to identify the best model configuration, and the best accuracies were found to be very similar, it was decided to run the test dataset on the five best models instead of just on the best one. The five chosen values, as well as the best average values for the two parameter dimensions, are highlighted in bold font and with a cell border in Table 6-2. The classification results for these five model configurations using the test dataset are shown in Table 6-3.

TABLE 6-2: CLASSIFICATION RESULTS FOR THE EXPERIMENT ON THE SPOKEN DIGIT RECOGNITION DATASET. ALL VALUES ARE GIVEN IN %.

Size	**MNI**			**Talairach**					**Ø**
LIF	by 3	by 4	by 5	by 6	by 7	by 8	by 9	by 10	
0.01	80.03	79.89	83.96	76.97	78.05	78.36	76.98	78.05	79.04
0.02	87.75	86.69	85.06	79.51	79.07	78.62	76.56	75.62	81.11
0.03	88.85	87.11	84.93	77.65	77.60	78.89	77.46	80.41	81.61
0.04	89.25	87.11	85.33	79.01	77.58	79.17	75.57	82.52	81.94
0.05	88.46	87.62	85.53	78.71	81.45	81.21	82.28	78.92	83.02
0.06	88.63	86.92	85.78	84.74	85.88	84.64	85.52	77.88	85.00
0.07	89.12	87.45	85.47	86.09	85.06	86.12	85.81	78.39	85.44
0.08	88.98	88.14	85.33	86.44	85.95	86.72	85.14	79.00	85.71
0.09	88.69	88.56	85.11	87.64	86.60	87.34	82.28	80.07	85.79
0.10	88.30	88.65	84.45	87.42	86.84	86.75	81.22	79.55	85.40
0.15	**89.65**	87.79	84.17	88.25	87.60	89.11	84.36	77.22	**86.02**
0.20	**89.45**	87.31	83.67	88.33	**89.72**	88.47	83.50	76.44	85.86
0.25	**89.34**	87.19	83.51	89.33	88.55	87.36	83.45	75.37	85.51
0.30	89.10	86.70	83.41	**89.79**	87.46	87.15	83.64	75.53	85.35
0.35	88.67	87.22	83.66	89.19	87.49	87.78	83.17	76.00	85.40
0.40	89.12	86.53	83.56	88.91	87.46	87.33	83.46	75.33	85.21
0.45	88.80	86.52	83.52	88.72	87.49	87.43	83.68	75.36	85.19
0.50	88.76	86.73	83.43	88.90	87.08	87.25	83.56	75.09	85.10
Ø	**88.39**	86.90	84.44	85.31	84.83	84.98	82.09	77.60	84.32

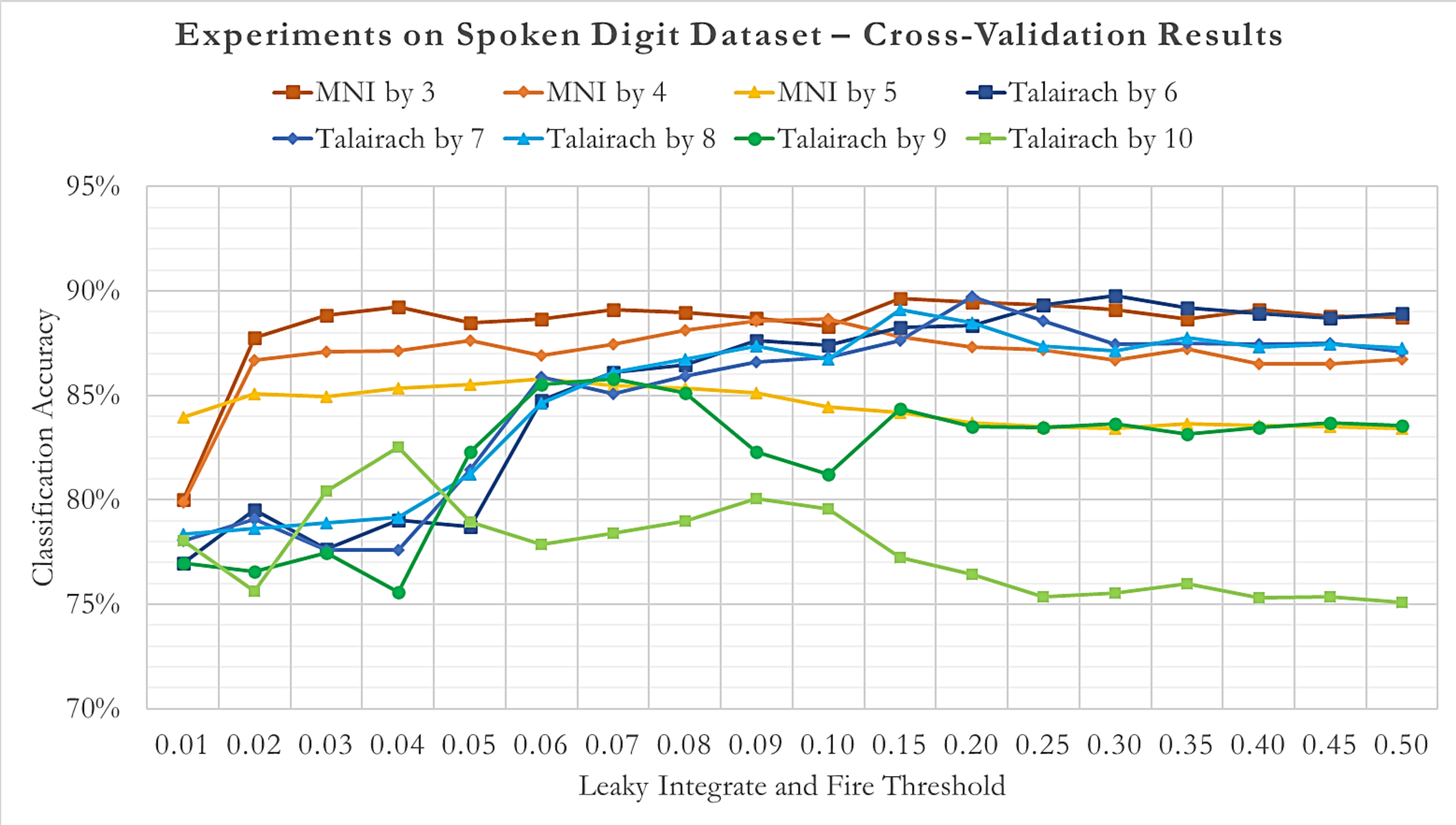


FIGURE 6-1: CLASSIFICATION RESULTS FOR THE EXPERIMENT ON THE SPOKEN DIGITS DATASET.

TABLE 6-3: CLASSIFICATION ACCURACIES FOR THE FIVE BEST SOUND PROCESSING MODELS ON VALIDATION AND TEST DATA.

Template name	LIF threshold	Validation accuracy	Test accuracy
MNI_by_3	0.15	89.65%	**90.52%**
MNI_by_3	0.2	89.45%	89.86%
MNI_by_3	0.25	89.34%	90.04%
TAL_by_6	0.3	**89.79%**	89.77%
TAL_by_7	0.2	89.72%	90.03%

Interestingly, the classification results for the test dataset were slightly better than those for the cross-validation. The reason for this could be that more training data were available for this final test of the model since 70% of the samples were used for training instead of just 4/5 of these 70% during the cross-validation. The best accuracy that could be achieved for the sound processing model was 90.52% for the "MNI_by_3" template with a LIF threshold of 0.15.

Besides the reported classification results, it is also evident from both Table 6-2 and Figure 6-1 that the results for the largest template sizes are missing. This is because running the respective experiments on the available hardware resulted in memory errors – there was simply not enough RAM space for the network data to be processed. Table 6-4 summarises the execution times for all experiments and notes which template sizes could not be run. The execution times are summarised by template size because batch processing the experiments meant that runtimes for single LIF values were not recorded. However, it was anecdotally noted that lower LIF values seemed to have an increased execution time. One explanation for this observation could be that lower thresholds led to increased spiking activity since fewer signals were needed to excite the neurons sufficiently to fire. This increased activity required more computational processing steps and hence more time.

As an additional note, the runtime for the cochlear encoding was, on average, 25 minutes and 7 seconds per sample. The encoding involved many calculations and thus required considerable CPU capacity so it was parallelised across the available four CPU cores to speed up the process. In contrast, the classification experiments required large memory capacity to store the current state of the network, which is why 16 GB RAM were not enough to run the larger templates even though encoded data were available for them.

TABLE 6-4: RUNTIMES OF EXPERIMENTS ON SPOKEN DIGITS DATASET BY BRAIN TEMPLATE.

Brain template	Number of neurons	Runtime
TAL_orig	1,527,747	Memory error
TAL_by_2	192,600	Memory error
TAL_by_3	56,770	Memory error
TAL_by_4	23,550	Memory error
TAL_by_5	12,150	Memory error
TAL_by_6	7,199	12 d 16 h 35 m 22 s
TAL_by_7	4,452	6 d 7 h 5 m 51 s
TAL_by_8	2,960	3 d 16 h 44 m 41 s
TAL_by_9	2,086	2 d 3 h 42 m 25 s
TAL_by_10	1,525	1 d 6 h 34 m 59 s
MNI_times_2	1,932,848	Memory error
MNI_orig	241,606	Memory error
MNI_by_2	30,182	Memory error
MNI_by_3	8,907	6 d 13 h 36 m 18 s
MNI_by_4	3,747	2 d 6 h 12 m 37 s
MNI_by_5	1,939	1 d 3 h 37 m 13 s

The execution times for the models were also considered when choosing and recommending the best model configuration. While larger templates generally performed better, they also required considerably more training time. Depending on the application area, slightly lower accuracies could be acceptable as a trade-off for much faster execution times.

Interestingly, the Talairach templates were much slower than the MNI templates, even if the number of neurons in the network was comparable. No feasible explanation for this discrepancy could be found so investigating this further was identified as a topic for future research.

6.5 DISCUSSION

In this section, the results of prior works in the field of sound recognition are compared to the results of the sound processing system presented in this thesis. Automated sound recognition encompasses a wide array of problems, from environmental sound classification over music recognition and speaker identification to transcribing continuous speech, to name a few. The sub-field of speech recognition, even if narrowed down to just spoken digit recognition, has been extensively researched and a vast amount of literature with reports of systems and results is available. This discussion, therefore, focused on only those works that were thematically close to the sound processing system presented here. This means that for the purpose of this discussion, papers had to:

- Use a biologically inspired algorithm and, preferably, an SNN;
- Report results on either the FSDD, the TIDIGITS dataset, or a comparable dataset; and
- Employ a similar experimental setup with separate validation and test data to ensure the reliability of the results;

to be considered for comparison with the work presented here. The restriction to focus on biologically inspired methods stemmed from the observation that a large amount of work had been done on more conventional methods like Hidden Markov Models (Afify et al., 2009; Rabiner et al., 1989; Spille, Kollmeier, & Meyer, 2017; Stowell, Benetos, & Gill, 2017), time series classification (Gee, Garcia-Olano, Ghosh, & Paydarfar, 2019), or Convolutional and Deep Neural Networks (Çakır, Parascandolo, Heittola, Huttunen, & Virtanen, 2017; Harshita & Adiga, 2018; Hinton et al., 2012; McLoughlin, Zhang, Xie, Song, & Xiao, 2015; Ren et al., 2018; Sharmin, Rahut, & Huq, 2020; Sinha, Awasthi, & Ajmera, 2020; C.-Y. Wang et al., 2020). These methods generally focused on achieving high classification accuracies and optimal parameter tuning. In contrast, the objective of the work presented in this thesis was to mimic parts of the biological hearing process and to assess if such a system was feasible, so a comparison with work that had similar objectives was considered more meaningful. Furthermore, several comparable algorithms and methods were identified that had been evaluated on datasets from domains other than spoken digit recognition (Cerezuela-Escudero et al., 2016; Graves et al., 2013; Higgins, Stringer, & Schnupp, 2018; Holmberg et al., 2007; Lyon, Ponte, & Chechik, 2011; Näger, Storck, & Deco, 2002; Namarvar, Liaw, & Berger, 2001; Storck, Jäkel, & Deco, 2001; Jibin Wu et al., 2020; Xiao & Weibei, 2016) or on only a subset of the digits (Higgins, Stringer, & Schnupp, 2017). While these methods were biologically inspired and thus fit the first criterion, it was difficult to draw conclusions from comparisons with application areas where the data differed greatly. Lastly, the classification approach described in the papers had to satisfy the minimum criterion of using separate training and test data. Encouragingly, no peer-reviewed papers were found that did not meet this requirement.

In total, 15 papers were identified that matched the three criteria outlined above. These papers are summarised in Table 6-5, which also contains a brief statement about the methods and datasets that were used in those works and what accuracies were achieved. Since only very few papers could be found that evaluated their methods on the FSDD, three more datasets were considered here that also contained spoken utterances of ten or eleven[56] digits: TIDIGITS (Leonard, 1984), the digits subset of the TI-46 dataset (Doddington & Schalk, 1981), and Aurora-2 (Hirsch & Pearce, 2000).

All of the included papers describe biologically inspired signal processing methods that were used to some extent to classify the spoken digit datasets. Some of these works used encoding methods that were inspired by the functioning of the cochlea. For example, three research groups used a passive ear model developed by Lyon (1982), while two used a silicon cochlea sensor that was developed as a piece of hardware simulating the cochlea (S.-C. Liu et al., 2014). However, none of the works used the cochlear model by Zilany et al. (2014) that was used in the sound processing system presented in this thesis.

Furthermore, none of the identified literature made use of a tonotopic mapping approach. The majority of the identified works described novel, biologically inspired signal classification methods that were based on the functioning of SNN or similar methods like Liquid State Machines (Maass, Natschläger, & Markram, 2002). In contrast, as an extension of conventional SNN, the architecture presented in this thesis was designed to be three-dimensional and brain-shaped to enable a signal mapping procedure that was based on tonotopy. The system, therefore, evaluated a unique approach for processing auditory signals.

With regards to the performance of the system presented here, its accuracy of 90.52% compares well with that of the papers in Table 6-5, although it is at the lower end of the range. While this approach could not outperform other methods in the domain of speech processing, its merit lies in its ability to easily integrate other modalities like visual signals as described in Chapter 8. The system's good performance on the FSDD data is an added benefit and a promising sign for an expected good performance of the combined audio-visual system.

[56] Datasets with eleven digits contain both "oh" and "zero" for the digit 0.

TABLE 6-5: OVERVIEW OF WORKS FROM THE LITERATURE THAT ARE COMPARABLE TO THE SOUND PROCESSING SYSTEM.

Reference	Method	Dataset	Accuracy
Dibazar, Song, Yamada, and Berger (2004)	Dynamic synapse neural network	TIDIGITS	85.10%
Graves, Eck, Beringer, and Schmidhuber (2004)	Long short-term memory recurrent neural network	TIDIGITS TI-46 digits	98.90% 98.00%
Verstraeten et al. (2005)	Lyon passive ear model and Liquid State Machine	TI-46 digits	99.50%
Schrauwen et al. (2007)	Lyon passive ear model and Liquid State Machine	TI-46 digits	95.00%
Wade, McDaid, Santos, and Sayers (2010)	SNN with LIF neurons	TI-46 digits	95.25%
Abdollahi and Liu (2011)	Silicon cochlea sensor and Support Vector Machine	TIDIGITS	95.58%
Schafer and Jin (2014)	Spike-based feature detector and Support Vector Machine	Aurora-2	~95%
Y. Zhang et al. (2015)	Lyon passive ear model and Liquid State Machine	TI-46 digits	99.79%
Tavanaei and Maida (2017a)	Multi-layer SNN with LIF neurons	Aurora-2	96.00%
Tavanaei and Maida (2017b)	SNN with Izhikevich neurons	Aurora-2	90.80%
Anumula, Neil, Delbrück, and Liu (2018)	Silicon cochlea sensor and Phased Liquid State Machine	TIDIGITS	91.25%
Dong, Huang, and Xu (2018)	SNN with LIF neurons and Support Vector Machine	TIDIGITS	97.50%
Li and Príncipe (2018)	LIF-based spike generator and Reproducing Kernel Hilbert Space	TI-46 digits	95.23%
Jibin Wu, Chua, Zhang, et al. (2018)	Self-organising map with SNN	TIDIGITS	97.40%
Iranmehr, Shouraki, Faraji, Bagheri, and Linares-Barranco (2019)	SNN with ionic liquid space	FSDD	~74 %

Interestingly, Y. Zhang et al. (2015) also investigated the influence of network size on the classification performance and found that larger networks achieved higher accuracies. While their network design was different from the one presented in this thesis, it is encouraging that the same trend could be identified in the experiments described in this chapter.

Another noteworthy finding of this brief literature review was that about half of the papers were published in the last five years. This indicates that the topic of using biologically inspired methods for sound processing systems is gaining traction and more development can be expected in the near future.

6.6 CHAPTER SUMMARY

This chapter presented an experiment on a novel biologically inspired sound processing system using the Free Spoken Digits Dataset for spoken digit recognition. Two parameters of the system were optimised during the experiment, LIF threshold and network size. The best performance was achieved on the "MNI_by_3" template with a LIF threshold of 0.15, which classified 90.52% of the test samples correctly. With regards to Research Question 3a in Section 1.3, which asked about the comparative performance of the model, it was found that more specialised sound processing typically performed only slightly better than the model tested here. Furthermore, Research Question 3c, which asked about the influence of network size on system performance, can be answered affirmatively. While it was found that templates with more neurons generally performed better than those with fewer neurons, they also required more processing time. This trade-off between accuracy and required computational resources must be considered when choosing the best processing system for the desired application.

In future work, one way to improve the system's classification accuracy could be to widen the parameter search and find a better configuration for the model. The two parameters that were optimised here, LIF threshold and network size, only formed a starting point for an initial evaluation of the model, and to the best of the thesis author's knowledge, no structured analysis has been performed on how the parameters of the JNeuCube influence its performance. It would then also be an interesting question to investigate if the model generalises well to datasets from other domains since the configuration that was found here might only be optimal for the FSDD.

Another means of improving the sound processing model in the future could be to include some form of simulation of the auditory pathway in the system. Higgins et al. (2018) found that for their setup, including the auditory pathway significantly improved their results. Since this step was skipped in the sound processing model presented here, it might also be a worthwhile aspect of future research.

From a computational standpoint, the system could be improved in the future by adding a perceptual filter before the encoding step. The chosen filter would also have to be based on relevant biological features so that it would add tangible value to the architecture.

7 BENCHMARKING THE VIDEO PROCESSING SYSTEM

"There should be no combination of events for which the wit of man cannot conceive an explanation."

– Sherlock Holmes in The Valley of Fear

7.1 CHAPTER OVERVIEW

This chapter describes how the video processing model introduced in Section 5.4 was applied to a benchmark dataset in the domain of gesture recognition. The performed experiment aimed to evaluate the capabilities of the proposed model and to gain insight into potentially suitable model parameter configurations by comparing different setups using the same benchmark data and computer hardware. The chapter first describes the criteria which were applied when choosing the dataset. This is followed by a detailed explanation of how the dataset was analysed with the model. Subsequently, the results of the experiments are presented, and conclusions are made about the model and its optimal configuration. The chapter closes with a short discussion of the advantages and shortcomings of the model compared to other published work using the same or similar datasets, in an effort to answer Research Question 3a that was asked in Section 1.3.

7.2 DATASET SELECTION AND DESCRIPTION

The dataset used for the experiments on the video processing model was carefully selected to provide the best possible evaluation of the model's capabilities with respect to the objectives of this thesis. There are hundreds of video datasets available in the public domain that cover different tasks, topic areas, collection modalities, and purposes. Therefore, the search was very quickly narrowed down using several criteria.

Firstly, the domain to which the dataset belonged should reflect natural processes that could easily be integrated with audio data. Since the dataset for the combined audio-visual system had already been chosen at this point and contained five sign language signs and their equivalent spoken words, a dataset for the visual experiments that resembled signed words or hand gestures was desirable. Secondly, to fully assess the capabilities of the model, the selected data should be both dynamic and in colour. Since the retinal encoding module transformed the differences of two subsequent frames into spikes, a dataset consisting of only static images, or videos with hardly any movement, would not be suitable here. Videos in colour instead of greyscale were also preferable so that the newly developed colour encoding algorithm could be evaluated. Finally, from a computational standpoint, the dataset should be usable for a benchmarking and classification task so that the results achieved with the video processing system presented in this thesis could be compared with other architectures and models from published works. Further to this point, it would be favourable if the dataset had been used in several previous studies to facilitate this comparison.

Applying all of these criteria meant that well-known, highly-cited datasets (Chaquet, Carmona, & Fernández-Caballero, 2013) like Weizmann, KTH, or CAVIAR could not be used for this experiment because they lacked critical features. For example, the Weizmann dataset (Gorelick, Blank, Shechtman, Irani, & Basri, 2007) shows full-body shots of people performing different actions, which was not deemed useful for studying hand gestures. The KTH dataset (Laptev, 2004) was rejected for this experiment because the videos were recorded in greyscale, and the CAVIAR dataset (Fisher, Santos-Victor, & Crowley, 2005) showed multiple people in the same frame, which also made it unsuitable for the experiments performed here.

Narrowing down the search to "video hand gesture recognition datasets" yielded far fewer potentially suitable dataset candidates. Table 7-1 shows the name, reference, special feature, number of samples, and number of classes of five datasets that were considered for the experiment presented in this chapter. They were shortlisted because they had received comparatively high numbers of citations (Ruffieux, Lalanne, Mugellini, & Abou Khaled, 2014) or seemed prominent in search results on Google Scholar and Scopus.

TABLE 7-1: OVERVIEW OF SHORTLISTED GESTURE RECOGNITION DATASETS.

Name	Reference	Special feature	Samples	Classes
ChaLearn	Guyon, Athitsos, Jangyodsuk, and Escalante (2014)	One-shot learning task	54,000	8-12
Cambridge	T.-K. Kim, Wong, and Cipolla (2007)	Varying illumination	900	9
EgoGesture	Yifan Zhang, Cao, Cheng, and Lu (2018)	First-person (egocentric) view	24,000	50
JESTER	Materzynska, Berger, Bax, and Memisevic (2019)	Crowd-sourced	150,000	27
Keck	Jiang, Lin, and Davis (2012)	Military gestures	294	14

The dataset that was identified as being most suitable for the experiment conducted on the video processing system was the JESTER dataset because it was expected to present a bigger challenge than the other four datasets. The JESTER dataset contains almost 150,000 videos of 25 gestures and two non-gesture classes that were collected by over 1,300 different crowd-sourced actors in their homes, resulting in a huge diversity of people and video backgrounds. According to its creators, this makes it the largest freely available gesture recognition dataset to date (Materzynska et al., 2019). The files were published as a series of JPG images that represented the frames of the collected videos. Therefore, they had to be reassembled into MP4 video files by the thesis author before they could be entered into the retinal encoding module. With a frame rate of 12 fps, each video was roughly three seconds long and contained about 35 frames. The files were downloaded from the project website (https://20bn.com/datasets/jester/v1) on 23rd June 2020 for the experiments described in this chapter.

In this experiment, only five of the available 25 gestures were used. Exemplary still images of each class are shown in Figure 7-1. These classes were selected because they most closely resembled signed words from New Zealand Sign Language, which was the study subject of the combined audio-visual experiment described in Chapter 8. The five gestures chosen here visually matched the five sign language words that were used in Chapter 8. The “drumming fingers” class of the JESTER dataset closely resembled the sign for “who”, the “stop sign” class the sign for “stop”, the “thumb up” class the sign for “up”, the “thumb down” class the sign for “down”, and the “zooming out with two fingers” class the sign for “bird”. The intent behind this selection was to provide a task from which plausible predictions could be inferred for the combined audio-visual model and to reduce the computational complexity of the experiments.

FIGURE 7-1: EXEMPLARY STILL IMAGES OF THE FIVE CLASSES FROM THE JESTER DATASET THAT WERE USED FOR THE EXPERIMENT.

Furthermore, only 100 samples were randomly chosen for each class to reduce the required computational effort and to create a more biologically realistic challenge for this initial evaluation of the video processing system. In nature, the human brain would not typically expect several thousand samples of the same gesture before being able to reliably recognise this gesture. However, the creators of the JESTER dataset noted that the classification accuracy for their model plummeted when the number of training samples was greatly reduced (Materzynska et al., 2019). This meant that the classification accuracy for the experiments described in this chapter was expected to be lower than that of comparable models reported in the literature that had used more training samples. On the other hand, a better-than-expected result would support the hypothesis that the computational model described in this thesis might exhibit biological, brain-like characteristics. In any case, the reduced computational complexity resulting from these constraints was expected to be beneficial in the search for the best parameter configuration of the model, since more experiments with different settings could be performed on the available hardware.

7.3 EXPERIMENTAL SETUP

As for the experiments with the sound processing system, the goal of this experiment was to find an optimum parameter configuration for the model so that the best results could be compared to other models that had used the same dataset. Since the model's architecture offered a variety of parameters that were all to some extent related to each other and influenced the model's performance, it first had to be decided which of them to optimise based on the amount of knowledge that could be gained from studying them. This section first discusses the three steps required to enter the data into the model, followed by the parameters for the neural network itself and finally the experiment's training and testing procedure.

7.3.1 DATA PREPARATION

Preparing the video data before it could be entered into the neural model required three steps that were discussed in detail in Section 5.4. The following paragraphs describe how these steps were applied in practice for this experiment.

In the first step, **converting the video data** to spikes, the retinal encoding module compared the pixel properties of two subsequent frames and tried to capture the video's dynamic characteristics in spike sequences. Figure 7-2 shows an example of the encoding module on a sample of the "drumming fingers" class. The code for the retinal encoding module can be found in Appendix A, Listing V. The frames were processed by a peripheral greyscale encoding algorithm (function *get_frame_diff_as_spikes* in the code) and the focal centres of the frames were also processed by a foveal colour encoding algorithm (function *get_fovea_spikes*). This focal centre was moved to the most active region of the frame after each frame change (as part of the function *get_block_spikes*).

The two algorithms first quantified the differences between the pixels in the two frames and then used a thresholding function to determine if a so-called "pixel spike" should be created. The pixel spikes were then summarised into blocks that were systematically arranged across the frame. Again using a thresholding function, a "block spike" was created for each block that had more spikes per pixel in its area than there were pixel spikes per pixels in the whole frame (functions *get_block_spikes* and *get_block_fovea*). This concept of a dynamic block threshold was adapted from the sound processing system, where it had led to good model performance. However, the results from the first set of experiments with the video processing system were not as good as expected (45% accuracy – for more details see

Section 7.4). Therefore, it was decided that a range of fixed block threshold values should also be tested, with separate parameters for the periphery and the fovea.

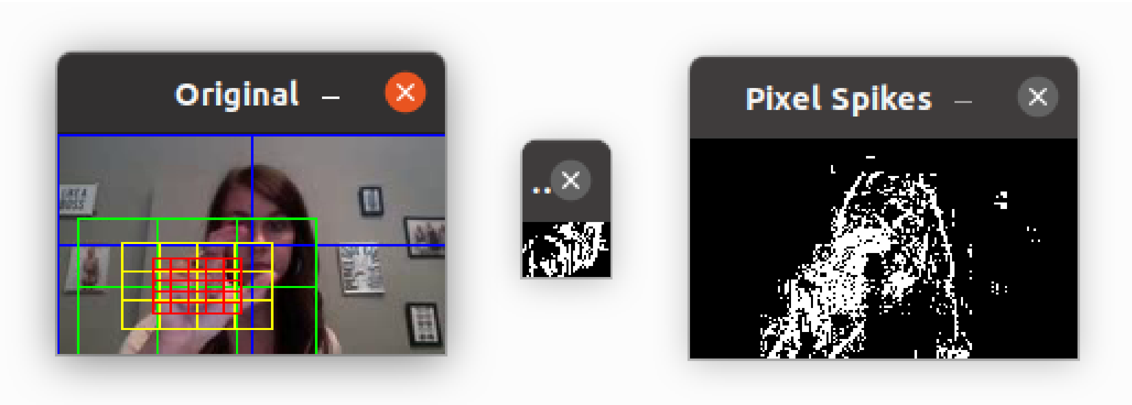


FIGURE 7-2: EXAMPLE SCREENSHOT OF THE ENCODING MODULE WITH A SAMPLE OF THE "DRUMMING FINGERS" CLASS. LEFT: THE ORIGINAL VIDEO WITH OVERLAID BLOCK BOUNDARIES. CENTRE: THE PIXEL SPIKES CREATED BY THE FOVEAL COLOUR ENCODING. RIGHT: THE PIXEL SPIKES CREATED BY THE PERIPHERAL GREYSCALE ENCODING.

An additional but minor computational restriction that stemmed from the arrangement of blocks across the frame was that all videos had to have the same frame ratio. Since all videos had a height of 100 pixels (Materzynska et al., 2019), and the majority of the videos had a width of 176 pixels, the desired frame ratio was fixed at 16:9. When selecting the 500 sample videos for the experiments described here, videos that did not match this ratio were discarded and a different video was randomly selected instead. The finally selected videos were then converted into spike matrices using the encoding module in Appendix A, Listing V.

In the second step of the data preparation, the **mapping** phase, these spikes were entered into the brain-shaped network model in a biologically plausible way based on data from retinotopy studies. This step covered two aspects, *number* and *location* of neurons. While the locations could be determined using the retinotopy data as described in Section 5.4.4, the number of neurons was dependent on the size of the SNN. All 16 template sizes described in Section 5.2.3 were used in this experiment with the required numbers of visual input neurons that were calculated as described in Section 5.4.3. As for the sound processing system, an examination of the relevant literature showed that the influence of network size on the performance of the SNN model is not very well understood. Therefore, network size was chosen as a parameter for optimisation for the experiments described in this chapter.

The available network configurations were based on two different brain atlases, MNI and Talairach, and contained between 1,525 and 1,932,848 neurons. For the varying numbers of input neurons, their locations were generated by combining the retinotopy data with the block summarisation algorithm as shown in the code in Appendix A, Listing VIII. The aim of this process was to map pixel coordinates in the video frames to the locations of processing neurons in the network.

The block summarisation algorithm was part of the third step, **compressing** the encoded spike data. This step was described in detail in Section 5.4.5 and was performed on the video dataset used in this experiment because the 100 × 176 pixels in the data could not be fed directly into the varying numbers of input neurons in the network templates. The block numbers and sizes that were defined for each template were, however, based on the frame ratio of the videos in this dataset. During this step, it was found that for the largest five template sizes, there were more block levels defined than could be applied to the videos in the dataset. Since the size of the blocks decreased towards the centre of the frame by the scaling factor described in Section 5.4.5, the final block levels contained blocks that were calculated to be smaller than one pixel in width or height. Therefore, no data could be encoded for the TAL_orig, TAL_by_2, TAL_by_3, MNI_times_2, and MNI_orig templates. However, based on the experience from Chapter 6, where it was found that the larger networks could not be simulated with the available hardware, this limitation was not anticipated to influence the outcome of the experiments described in this chapter.

The three steps described above are mainly related to data encoding and input into the model. Equally important is the configuration of the network, which also had to be specified before the experiment could be run. The model parameters of the network are described in the following section.

7.3.2 NETWORK MODEL PARAMETERS

In addition to the network size and the two block threshold values used for the encoding, which were already identified as variable parameters for the visual processing model, the JNeuCube software contained several more settings that could be altered to modify the model's behaviour. A detailed discussion of the available parameters and their purpose is provided in Section 5.2.2, while this section focuses on the values that were optimised here.

The first factor that was explored when designing the network was the **neuron type** and its **properties**. As for the sound processing system, the two types of neurons that were considered for the experiment here were LIF neurons (Brunel & van Rossum, 2007) and Izhikevich neurons (Izhikevich, 2003). For the LIF threshold, the same 18 values were used as for the experiments described in Chapter 6 for better comparability:

0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50

For the Izhikevich neurons, only the "F" property, which was described as being found in the "most common type of excitatory neuron in mammalian neocortex" (Izhikevich, 2004, p. 1064), was used for the experiments here. Since all types of Izhikevich neurons had performed poorly in the sound experiments, and the architecture and general setup of the experiments on the video data were very similar to those of the sound data, Izhikevich neurons were not expected to achieve very good results here either. Therefore, the experiments focused on exploring the LIF thresholds in addition to the network sizes.

Further to optimising the neuron type and properties that were used for the network, the video experiments had the added complexity of having to identify optimum values for the block threshold parameters in the retinal encoding module. These optimisation experiments were largely performed on only the MNI_by_5 template to reduce the computation times needed for these experiments. Based on the findings from the experiments on sound data described in Chapter 6, larger networks performed better than smaller networks when all other parameters were left unchanged, and it was hoped that this improvement could also be observed here. The MNI_by_5 template was chosen because, for the sound experiments, the group of MNI templates had achieved higher classification accuracies than the Talairach templates while only needing about half the training time for the same number of neurons.

The first set of experiments on the video processing system used a dynamic threshold for the two block threshold parameters of the retinal encoding module, based on a concept that was created for the signal compression step in the sound processing model. Further to this approach, ten fixed values were also tested for each of the two block thresholds:

0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9

As for the sound experiments, the rest of the model parameters that mainly concerned the network architecture were set to values that had proven successful in prior studies with the same system setup (Kasabov et al., 2016). This also facilitated better comparability of the two models. An overview of the chosen values is provided in Table 7-2, while an explanation of their meaning and rationale for default values can be found in the general description of the network parameters in Section 5.2.2.

TABLE 7-2: NETWORK MODEL PARAMETERS FOR THE EXPERIMENT ON GESTURE RECOGNITION.

Category	Parameter name	Value
Neuron model	LIF reset voltage	0 (simulating a full neural discharge)
	LIF refractory time	4
	LIF resistance	1
	LIF capacitance	10
Network initialisation	SWC weight range	Minimum -0.1 and maximum +0.1
	SWC radius	2.5 times the network size scaling factor
	SWC positivity rate	0.7
STDP	A positive/A negative	0.001
	τ positive/τ negative	10
	Weight boundaries	Upper bound +2 and lower bound -2
deSNN	Modulation factor	0.8
	Drift positive/negative	0.005
Classification	kNN k	3
	Number of folds	5
	Training rate	0.7

7.3.3 MODEL TRAINING AND TESTING PROCEDURE

The experimental procedure of the experiments using the video data largely followed that of the sound data, with the exception that additional experiments were performed on only the smallest network templates to find an optimal set of block threshold parameters for the retinal encoding. The employed procedure followed best-practice methods from applied machine learning research (Brownlee, 2020), with separate training, validation, and test data. Thirty per cent of the video samples were randomly chosen and held back as the final test data, while the remaining 70% were used to train and optimise the model using the five-fold cross-validation method. The best-performing model configurations were then used to classify the test data.

As for the sound experiments, all five models that were built for one run of the cross-validation were initialised with the same randomly created connections (but then fed samples from the varying four folds that were used for training). That way, the model results could be more easily averaged and compared. This cross-validation process was then performed 30 times for each set of parameters, where each of the 30 runs used a new arrangement of randomly created initial network connections so that the effects of outlier configurations could be minimised. The reported averaged cross-validation results were thus based on 150 different models.

There were **five stages of experimentation** for finding the optimum set of parameters, mainly focused on the retinal encoding module. Since there was no theoretical knowledge available on the newly developed algorithms and how their parameters would affect the performance of the model, it was one objective of the experiments described in this chapter to explore this area further. In the **first stage** of the experimentation, the initial configuration of the retinal encoding as described in Sections 5.4.1 and 5.4.2 was used. After these had not produced satisfactory results, in the **second stage** another set of parameters was tested, which was based on trying to match the spike rates that were found by Paulun et al. (2018) to achieve good results in their experiments. Since the system architecture by Paulun et al. was similar to the one proposed in this thesis, it was hoped that matching their system behaviour would improve the model's performance. The values chosen for the peripheral and foveal pixel thresholds were 25 and 5, respectively, while the two block thresholds were left at 0.0. While this did improve the performance of the model on the JESTER dataset studied here, the results were still lower than expected. Therefore, in the **third stage** of the experimentation, it was decided to reset the two pixel thresholds to the values that were originally identified as being most biologically plausible in Sections 5.4.1 and 5.4.2. A grid-search-like process was then employed to test a series of values for the two block threshold parameters since these could not be based on biological observations. During this stage, experiments were only performed on the smallest template size, MNI_by_5, to reduce the required hardware capacity and computation time. The third stage of experiments showed that there was likely no universally applicable best set of parameters that could be used going forward. The results (see Section 7.4) did not show a clear winner and the performance differences were not statistically significant. A compromise was found by selecting final parameters that were both biologically plausible and achieved comparatively good results. These values, 0.3 for the peripheral block threshold and 0.5 for the foveal block threshold, were then used in the final two stages of the experimentation. In the **fourth stage**, the encoded data were passed into network models that had been initialised with the Izhikevich Type F neurons as described in Section 7.3.2. For these experiments, all 16 network template sizes were used. The purpose of this set of experiments was to explore if any improvements could be achieved by changing the neuron type since all previous experiments had been performed on LIF neurons. However, this was not found to be the case. Therefore, the **fifth stage** again focused on experimenting with LIF neurons. In combination with the data encoded using the identified threshold parameters for the retinal encoding, all 16 network sizes were tested in this stage in an effort to investigate if the improvements seen with increasing network size in the sound experiment could be replicated with the video data. While the results were still not as good as had been hoped, the experimentation and parameter search were finalised at this point. Based on the extensive, yet ultimately unsuccessful procedure, it was expected that significant performance improvements could only be achieved through major redevelopments of the retinal encoding module that included more biologically plausible mechanisms.

The experiments from the final stage again followed the cross-validation procedure that had been used for the sound experiments described in Chapter 6. The five best-performing models were then chosen to be tested on the unseen 30% of the samples that had been held back from the beginning. For both the cross-validation and the final test, 30 experiments were run for each model configuration based on LIF threshold and network size. The results of this classification (and of those of the other stages of experimentation) are reported in Section 7.4.

As for the sound processing system, the execution times for each model configuration were monitored as a performance metric. All experiments were run on a standard PC with an Intel® Core™ i7-8700 CPU with 3.20 GHz clock speed, 16 GB RAM, and the Windows 10 64-bit operating system. The JNeuCube software was run using the Apache NetBeans IDE 11.3.

7.4 RESULTS

This section presents the results of the experiments on the JESTER gesture recognition dataset. While the focus here lies on the outcome of the five stages of experimentation and the overall averaged cross-validation accuracies, the detailed results for each of the models can be found in Appendix B, Table B-11 to Table B-60. As for the sound processing model, the experimental results in the diagrams and tables are usually shown dependent on the network template size and the LIF threshold. One exception to this are the experiments in the third stage, where only the MNI_by_5 template was used for optimisation and the results depended on a retinal encoding parameter. The second exception in the diagram layout is the set of experiments using the Izhikevich neurons, which were only performed on one type of neuron with no variable neuron parameters and, thus, only depended on the network size. All result diagrams used the same scale for the accuracy values for better comparability.

Since the retinal encoding module is an original contribution of this thesis, its behaviour with different sets of parameters was explored in detail during the experiments on the JESTER dataset. In the first set of experiments, the data samples were encoded using the pixel thresholds determined in Sections 5.4.1 and 5.4.2 in combination with dynamic block thresholds, copying an approach that had shown promising results in the sound processing model. However, as can be seen in Figure 7-3, the performance of this model was poor, with classification accuracies ranging from about 35 to about 45%.[57] With five classes of samples, this equates to roughly twice the accuracy of a by-chance selection. The best-performing model was TAL_by_5 with a LIF threshold of 0.5, which achieved 44.47% accuracy. Both TAL_by_5 and the threshold of 0.5 were also best-performing on average. However, the differences to the other network templates and thresholds were minimal. The distinguishably worst-performing template was MNI_by_3, but unlike what had been found in the sound experiments, no clear relationship between size and outcome could be identified. There was also no statistically significant difference between the accuracies of the LIF thresholds, as calculated using a t-test with $p = 0.05$.

[57] The detailed results of all experiments performed with these parameters can be found in Appendix B, Table B-11 to Table B-19.

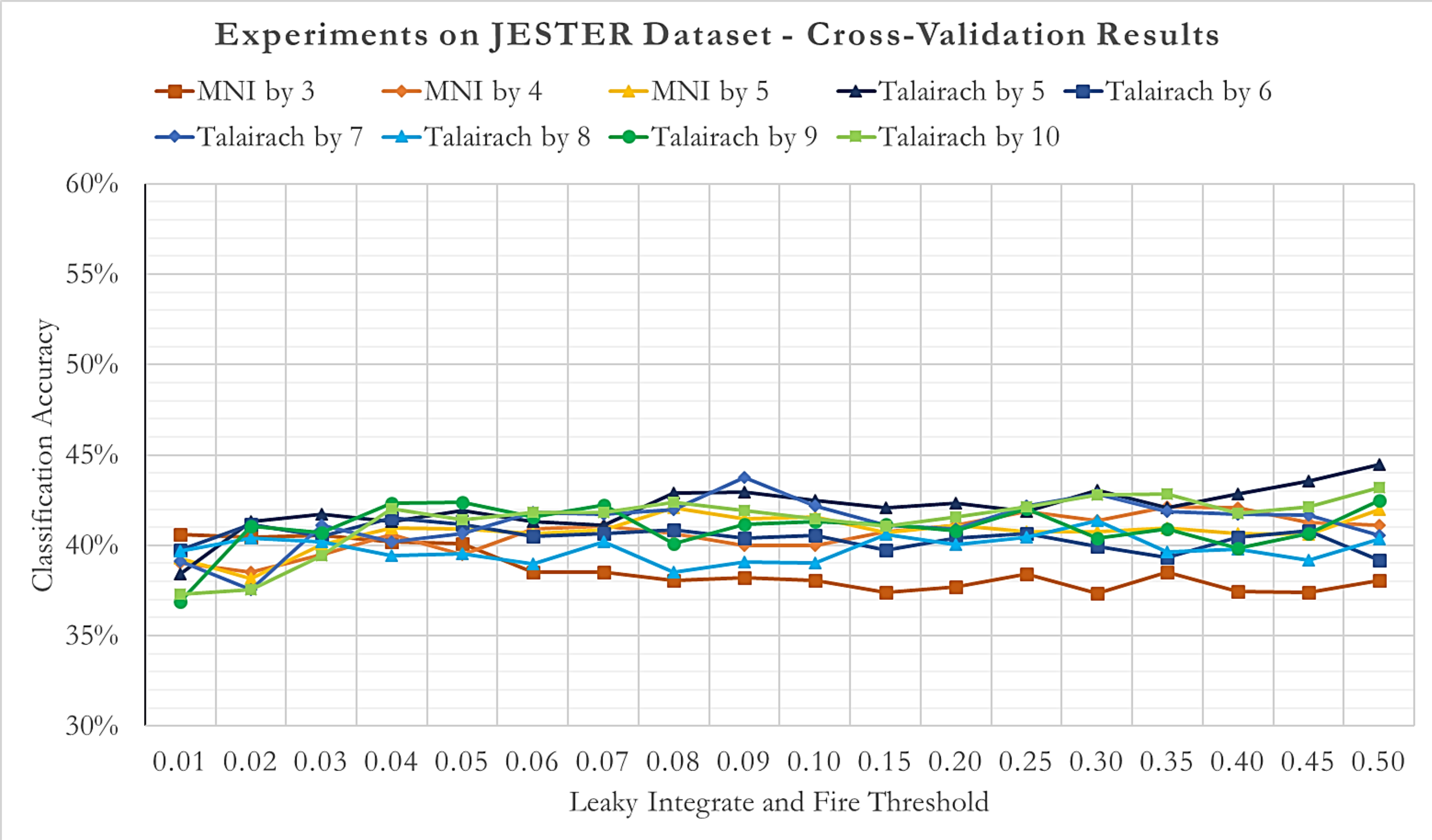


Figure 7-3: Classification results for the experiments on the gesture recognition dataset using dynamic retinal block thresholds.

In an effort to replicate the promising results that had been achieved by Paulun et al. (2018) using a similar system architecture, the thesis author decided to consider the spike rates of the retinal encoding module when choosing parameters. In the Paulun paper, the spike rates of the encoding module were reported to be directly related to the classification accuracy, with the best performance achieved when the spike rate was around 32% and the block threshold was set to 0. In both the Paulun system and the system developed here, a block threshold of 0 causes any spiking pixel in the block to also create a spike for the whole block in that time step. Therefore, in the second stage of experimentation on the JESTER dataset with the system presented in this thesis, this spiking behaviour was replicated to test if this would improve the performance of the model.

The two block thresholds were set to 0 and the peripheral pixel threshold was increased from its initial value of 3 to 25 to reduce the spike rate to an average of 41.2%. The foveal pixel threshold was left at 5, which produced an average spike rate of 54.2%. An overview of the spike rates, results, and the number of block levels per template size is shown in Figure 7-4.[58] The grey bars show the number of block levels in the retinal encoding for each template; their corresponding axis is on the right side of the diagram. The blue diamonds and orange circles show the peripheral and foveal spike rates, respectively, while the green triangles show the overall averaged classification accuracy for each template. Result values were not available for the largest three template sizes because these could not be modelled using the available hardware. However, the samples could be encoded into spike files so spike rates for these templates were included in the diagram. Since the retinal encoding algorithm presented in this thesis did not contain randomised variables, the spike rates would only change if the dataset was changed.

Interestingly, both the peripheral and the foveal spike rate increased when the number of block levels decreased. This was likely caused by the circumstance that with fewer block levels, the blocks were also larger and thus contained more pixels, which in combination with a block threshold of 0 meant that each pixel spike would directly create a block spike and less filtering was applied. At the same time, there was also a slight decrease in classification accuracy with increasing spike rate. This means that the dependencies found by Paulun et al. (2018) could not be confirmed in this experiment.

[58] The exact values can be found in Appendix B, Table B-20.

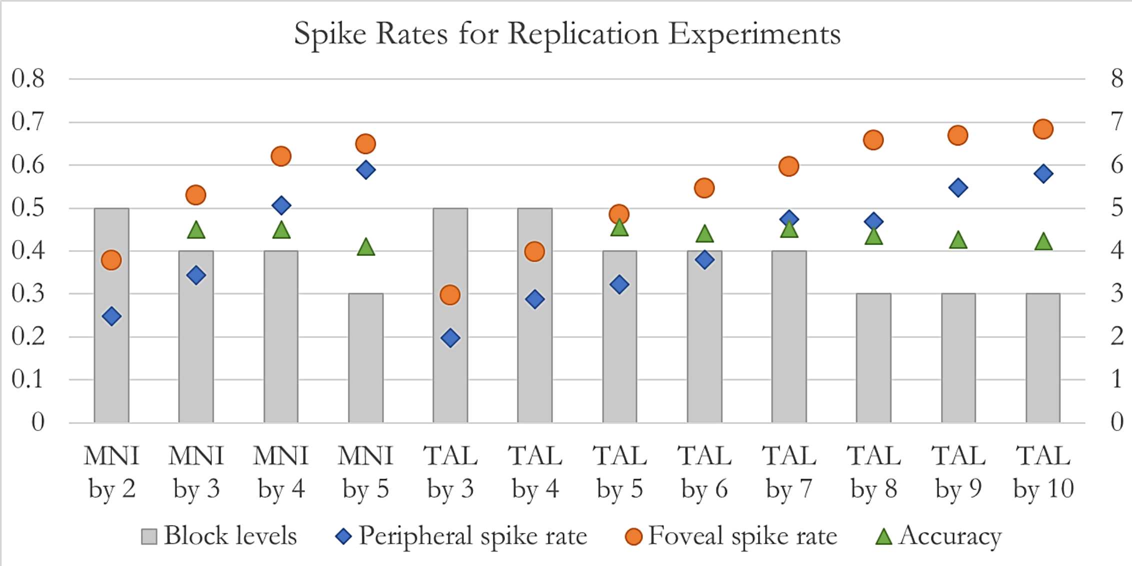


FIGURE 7-4: SPIKE RATES AND RESULTS FOR THE VIDEO PROCESSING SYSTEM WHEN TRYING TO REPLICATE THE BEHAVIOUR OF A SIMILAR SYSTEM. THE RIGHT AXIS APPLIES TO THE NUMBER OF BLOCK LEVELS AND THE LEFT AXIS APPLIES TO THE OTHER THREE DATA SERIES.

The averaged results per template and LIF threshold are shown in Figure 7-5 and the detailed results can be found in Appendix B, Table B-21 to Table B-29. As with the first set of experiments using the dynamic block thresholds, the accuracy differences between the models were not statistically significant (calculated using a t-test with p = 0.05). However, they were slightly better than the experiments using the dynamic thresholds, with most accuracy values ranging from 41% to 46%. The best-performing model was TAL_by_5 with a LIF threshold of 0.06, which achieved 48.78% accuracy. TAL_by_5 was also the best-performing network template on average, whereas the best-performing LIF threshold was 0.07. As for the sound experiments using the dynamic block threshold, the differences between network templates and LIF thresholds were minimal and not statistically significant (again using a t-test with p = 0.05). However, a slight downwards trend could be noticed in the lower threshold range that was also found in the sound experiments.

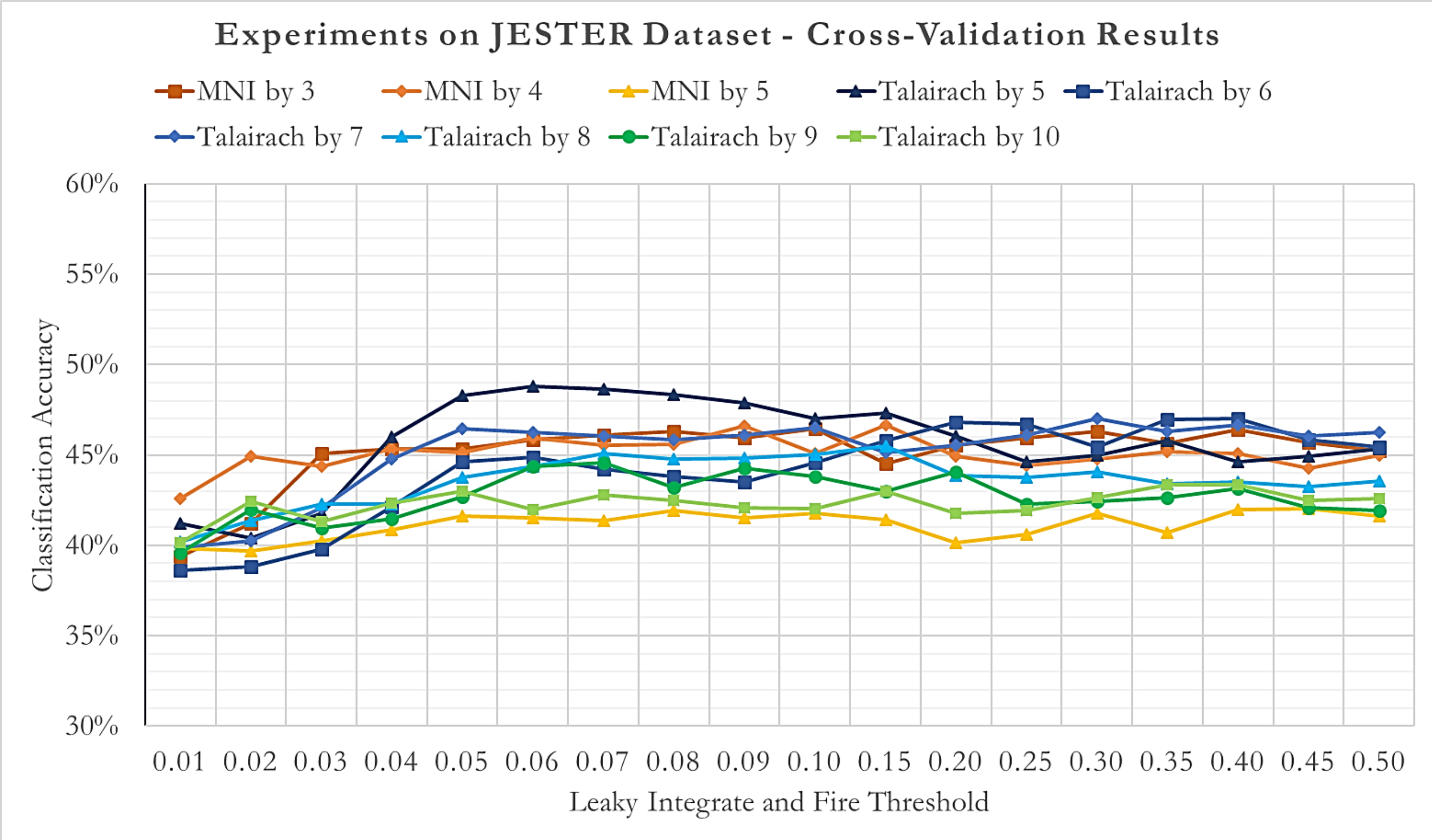


FIGURE 7-5: CLASSIFICATION RESULTS FOR THE EXPERIMENTS ON THE GESTURE RECOGNITION DATASET USING RETINAL ENCODING THRESHOLDS THAT ATTEMPTED TO MATCH THE SPIKING BEHAVIOUR OF THE SYSTEM INTRODUCED BY PAULUN ET AL. (2018).

As these results were still not as good as expected, the next stage of experimentation aimed at finding an optimum pair of block threshold values. A simple grid search mechanism was employed for this explorative search that looked at ten different values each for the peripheral and the foveal block thresholds. However, changing the peripheral block threshold did not yield noticeable performance improvements, so the initial plan of a full exploration of the quadratic parameter space was abandoned. Instead, both parameters were tested separately. In all experiments performed during this stage, the peripheral and foveal pixel thresholds were fixed at 3 and 5, respectively. These two values were identified in Sections 5.4.1 and 5.4.2 as being most biologically plausible. Setting biologically inspired parameters to values that were derived from neurological observations had already proven to be a good approach for the cochlear encoding as part of the sound processing system. Therefore, this principle was also followed here since no other source was available that could inform the choice of parameters for the video processing system.

The results of the two sets of block threshold experiments are shown in Figure 7-6 and Figure 7-7. Since all experiments were performed on the same network template, MNI_by_5, the data series in the diagram show the threshold values and not the template sizes. When optimising the peripheral block threshold, the foveal block threshold was fixed at 0.0, while for the foveal optimisation, the peripheral block threshold was fixed at 0.3. Detailed results for these experiments can be found in Appendix B, Table B-30 to Table B-50.

Looking at Figure 7-6, it becomes apparent that, unfortunately, the peripheral block threshold does not seem to have a big influence on the classification accuracy of the model. Most values range between 38% and 45%, with no statistically significant outliers (calculated using a t-test with $p = 0.05$). While the best-performing individual model with 45.23% classification accuracy had a peripheral block threshold of 0.7 and a LIF threshold of 0.5, the on average best-performing block threshold was 0.3 and the best LIF threshold was 0.3. Although the differences between peripheral block thresholds were minimal, it was decided to use the value 0.3 for further experimental exploration. No previous theoretical work could inform the choice of this parameter, so the anecdotally highest value was chosen.

The next set of experiments aimed at finding a good foveal block threshold. The results of these experiments are shown in Figure 7-7. They look slightly more promising than those of the peripheral optimisation, with most classification accuracies ranging from 40% to 50%. The best accuracy of 51.45% was achieved with a foveal block threshold of 0.9 and a LIF threshold of 0.35. On average, the highest accuracies could be seen for a foveal block threshold of 0.9 and a LIF threshold of 0.3. However, a t-test with $p = 0.05$ showed that there was no statistically significant difference between the accuracies of these experiments.

Interestingly, the highest value that was tested for the foveal block threshold performed best in this experiment. This indicated that higher accuracies could be achieved if most of the foveal pixel spikes were removed. This unexpected observation sparked the question about the usefulness of the colour encoding mechanism in the retinal encoding module. As an attempt to introduce more biological plausibility, does it also manage to improve the performance of the model? This question was answered with a set of experiments where the foveal encoding mechanism was deactivated in the source code of the retinal encoding. Samples were thus only transformed into spikes using the peripheral greyscale encoding. The models that were then trained on these new samples performed better than most other models where the samples had been encoded *with* the foveal colour encoding mechanism. The results of these new models are shown as a pink line in Figure 7-7. Their average classification accuracy was 48.80%.

Although a t-test showed that this result was not significantly ($p = 0.05$) better than that of the models including the foveal colour encoding, it does provide an anecdotal indication that the retinal encoding mechanism as it is used in the system presented in this thesis might work better if it only operated on greyscale videos. Including the foveal colour encoding when transforming the video samples does not improve the classification abilities of the final system. The findings of these two sets of experiments also indicated that the algorithms in the retinal encoding module would need to be reconsidered in future work, preferably by including more biologically plausible mechanisms. While tweaking the block threshold parameters could not achieve the expected performance improvements, revising the algorithm might.

For the last two stages of experimentation, a value of 0.5 was used for the foveal block threshold. This value had achieved the second-highest classification accuracy of 49.00%, after the accuracy of 49.06% for a threshold of 0.9. It was decided to keep the foveal encoding in the system at this point because it was a novel contribution of this thesis and further investigation intended to provide a more robust conclusion on its capabilities.

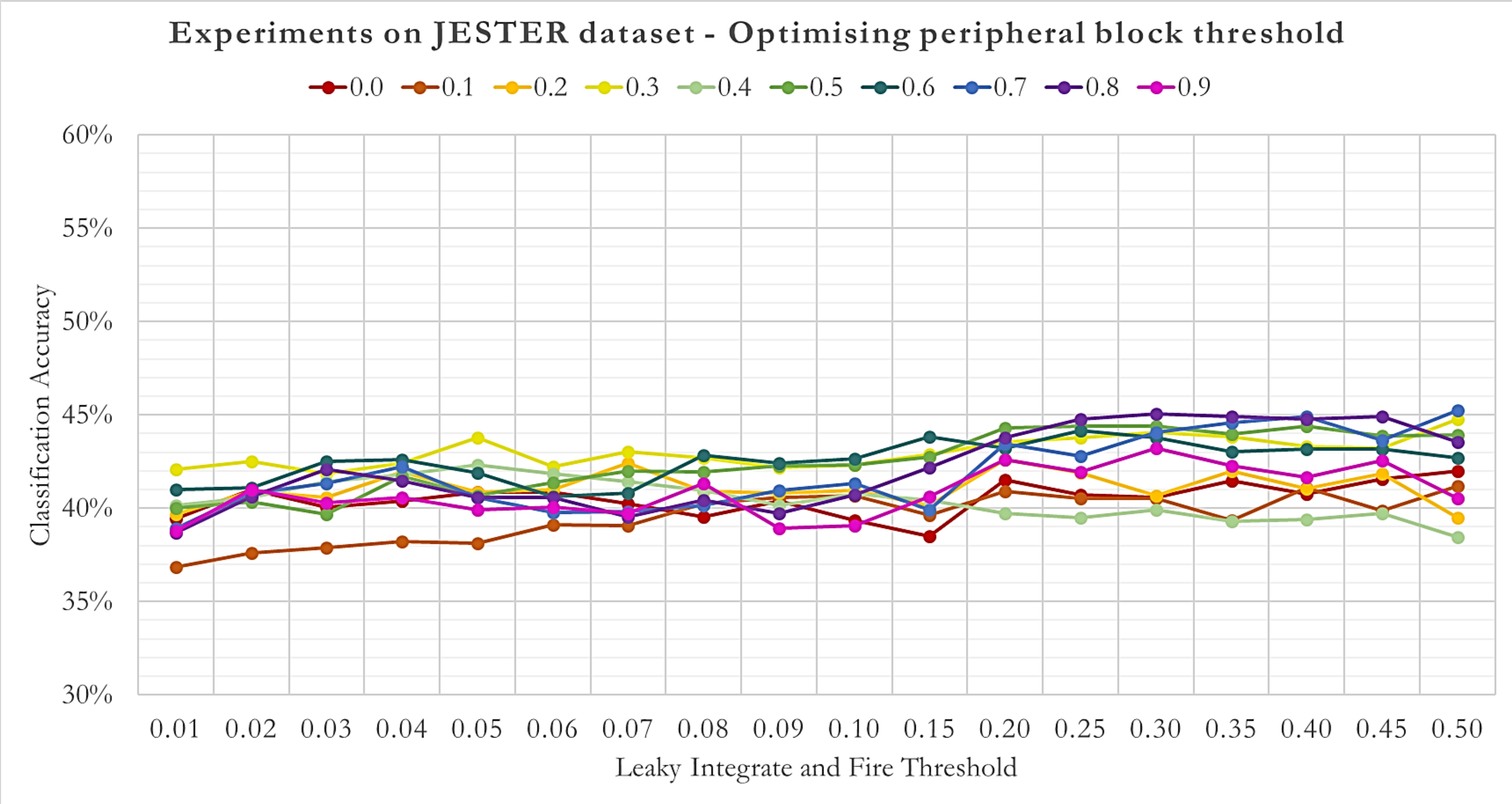


FIGURE 7-6: CLASSIFICATION RESULTS FOR THE EXPERIMENTS ON THE GESTURE RECOGNITION DATASET USING ONLY THE SMALLEST TEMPLATE SIZE AND DIFFERENT VALUES FOR THE PERIPHERAL BLOCK THRESHOLD.

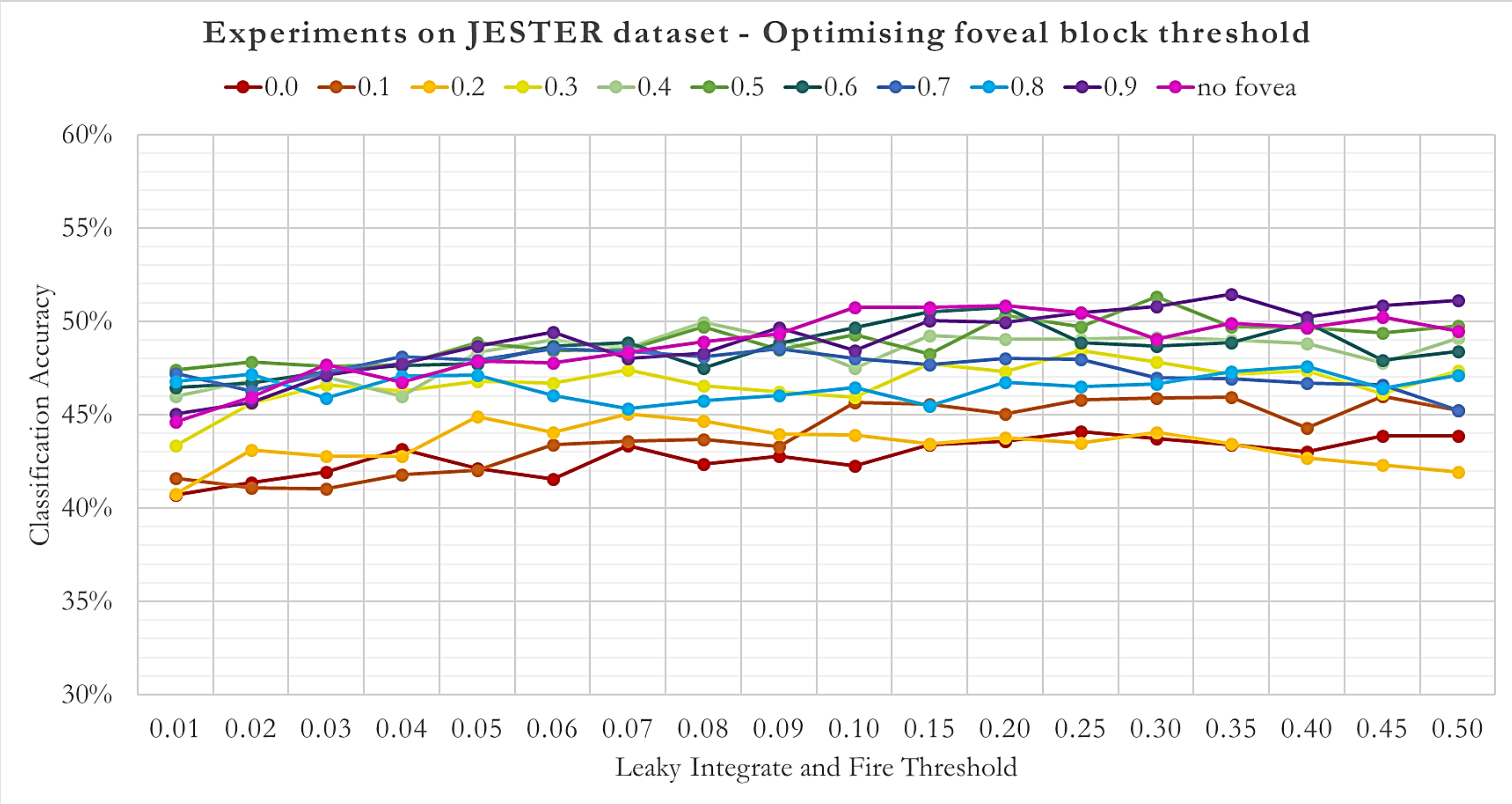


FIGURE 7-7: CLASSIFICATION RESULTS FOR THE EXPERIMENTS ON THE GESTURE RECOGNITION DATASET USING ONLY THE SMALLEST TEMPLATE SIZE AND DIFFERENT VALUES FOR THE FOVEAL BLOCK THRESHOLD.

Drawing from the findings in the previous experiments, the fourth stage of experimentation then deviated in the network's neuron type. Instead of using LIF neurons, the models in these experiments were created using Izhikevich neurons with the property labelled "F" by Izhikevich (2004). The four previously identified retinal encoding parameters were used as fixed values and models were built using all available network template sizes. The summarised results of these experiments are shown in Figure 7-8 and the detailed values can be found in Appendix B, Table B-51. They range from about 41% to about 47% and are hence similar to those achieved in the previous experiments using the LIF neurons. However, unlike the results of the first two stages of experimentation, it could be observed that smaller templates like MNI_by_5, TAL_by_8, and TAL_by_10 performed slightly but significantly (t-test with $p = 0.05$) better than their larger counterparts. This finding could indicate slight overfitting of retinal encoding parameters to smaller network sizes since they were optimised on only the MNI_by_5 template. Based on these findings in combination with what had been observed in the sound experiments described in Chapter 6, it was decided not to investigate the Izhikevich neurons further.

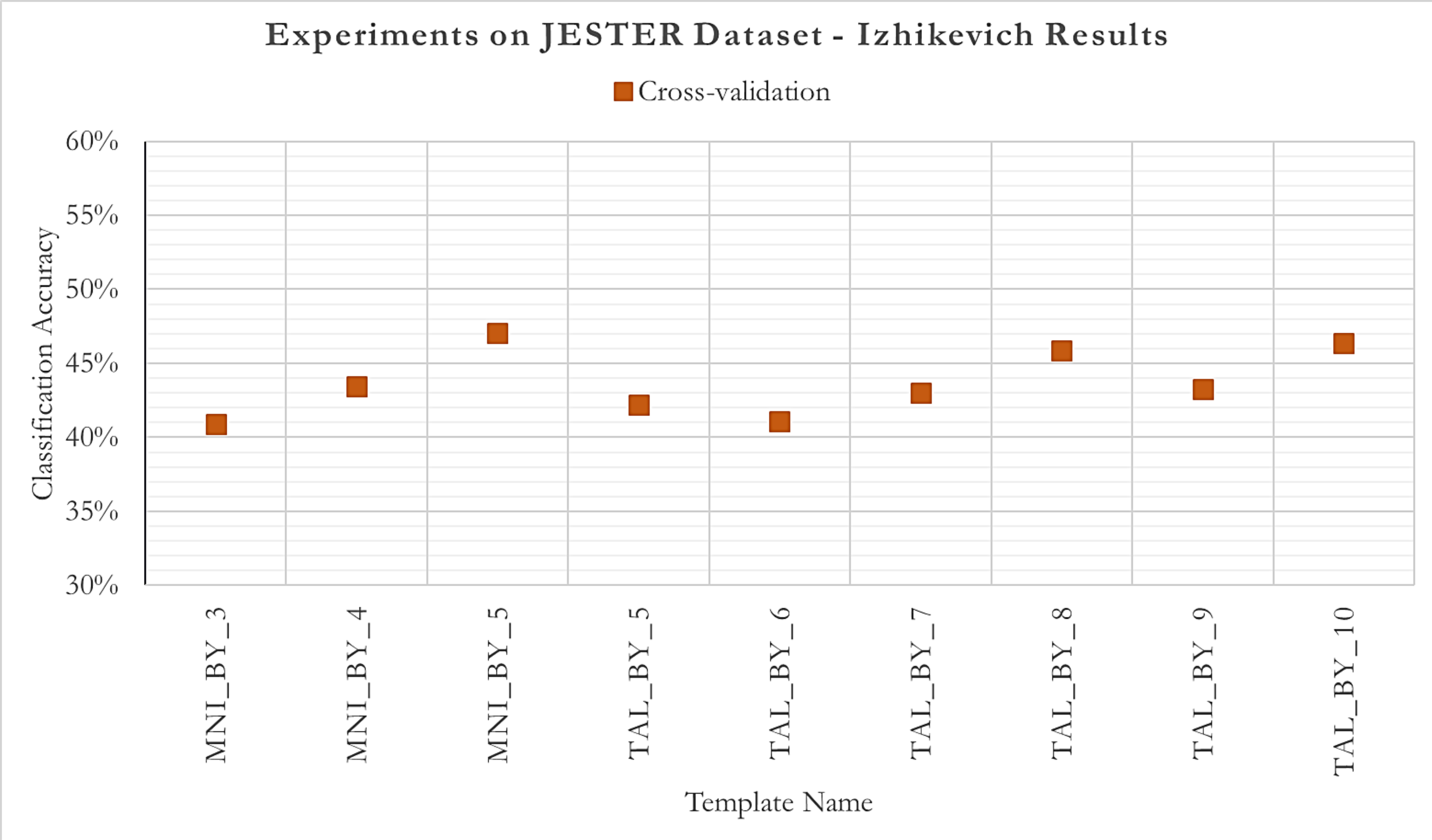


FIGURE 7-8: CLASSIFICATION RESULTS FOR THE EXPERIMENTS ON THE GESTURE RECOGNITION DATASET USING IZHIKEVICH NEURONS.

The final stage of experimentation tested if the optimisation results that were achieved on the MNI_by_5 template could be transferred to other template sizes. The data samples were encoded using the identified retinal threshold parameters: 3 and 5 for the peripheral and foveal pixel thresholds, respectively, and 0.3 and 0.5 for their respective block thresholds. The neurons in the models were again LIF neurons, so different firing thresholds were tested for the models in addition to all available template sizes. Figure 7-9 visualises the averaged cross-validation results of these experiments, which are also shown in Table 7-3. The detailed results of the 30 runs for each model configuration can be found in Appendix B, Table B-52 to Table B-60. The columns and rows in Table 7-3 show the template size and LIF threshold, respectively, with the last column and row containing the average classification accuracy for the respective parameter. Cell background colours show how the values compare to each other, with green indicating better performance, yellow indicating average performance, and red indicating worse performance.

Similar to the experiments using the Izhikevich neurons, it was observed that smaller templates performed slightly better than their larger counterparts. This behaviour, which stands in contrast to what was found in the sound experiments described in Chapter 6, could be caused by overfitting the "optimised" retinal encoding parameters to smaller template sizes. Furthermore, the issue of too few input neurons for smaller template sizes, which was found in the sound processing model, was not as pronounced in the video processing system, since the numbers of input neurons were much higher here. In general, however, the differences between template sizes were much smaller than those found in the sound experiments. Most accuracy values range from about 45% to 50%.

One similarity to the results of the sound experiments was that lower LIF thresholds seemed to be related to lower classification accuracy, at least for the Talairach templates in the experiments performed here. Furthermore, it was observed that the training times of these models were longer than those with higher LIF thresholds. One possible explanation for this could be that lower LIF thresholds lead to an "over-activity" in the network, which impeded the neural network's ability to extract meaningful information from single spikes and extended the required computational processing time.

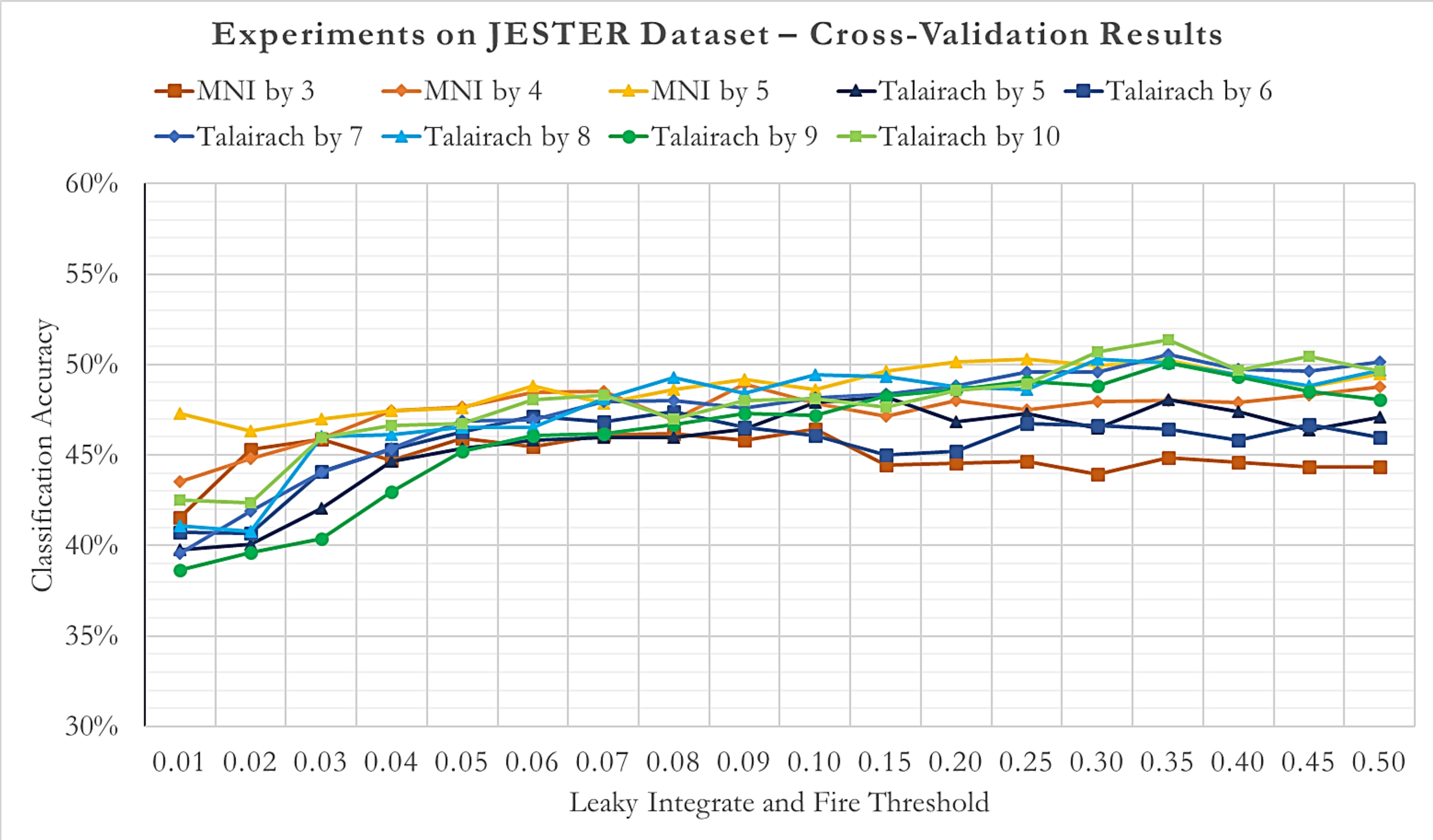


FIGURE 7-9: CLASSIFICATION RESULTS FOR THE EXPERIMENTS ON THE GESTURE RECOGNITION DATASET USING THE OPTIMISED RETINAL BLOCK THRESHOLDS.

TABLE 7-3: CLASSIFICATION RESULTS FOR THE FINAL EXPERIMENT ON THE GESTURE RECOGNITION DATASET. ALL VALUES ARE GIVEN IN %.

Size	**MNI**			**Talairach**						**Ø**
LIF	by 3	by 4	by 5	by 5	by 6	by 7	by 8	by 9	by 10	
0.01	41.53	43.53	47.28	39.79	40.73	39.55	41.07	38.62	42.51	41.62
0.02	45.33	44.80	46.32	40.08	40.66	41.92	40.79	39.64	42.36	42.43
0.03	45.88	45.90	47.00	42.06	44.07	44.06	46.00	40.40	45.97	44.59
0.04	44.69	47.44	47.45	44.65	45.29	45.36	46.10	42.99	46.63	45.62
0.05	45.94	47.64	47.61	45.38	46.27	46.88	46.55	45.19	46.74	46.47
0.06	45.45	48.45	48.84	45.80	47.13	46.93	46.53	46.08	48.05	47.03
0.07	46.12	48.49	47.87	45.97	46.84	47.95	48.12	46.19	48.33	47.32
0.08	46.15	46.94	48.61	45.98	47.40	48.03	49.26	46.68	46.99	47.34
0.09	45.80	48.91	49.16	46.41	46.55	47.62	48.39	47.29	48.02	47.57
0.10	46.41	47.85	48.62	47.92	46.07	48.18	49.41	47.17	48.11	47.75
0.15	44.45	47.15	49.64	48.20	45.02	48.37	49.31	48.25	47.63	47.56
0.20	44.56	48.00	50.16	46.85	45.21	48.80	48.79	48.61	48.56	47.73
0.25	44.66	47.50	50.29	47.30	46.76	49.60	48.59	49.08	48.92	48.08
0.30	43.94	47.95	49.93	46.50	46.63	49.57	**50.31**	48.81	**50.68**	48.26
0.35	44.87	48.02	50.22	48.07	46.41	**50.55**	50.07	50.08	**51.35**	**48.85**
0.40	44.60	47.90	49.41	47.38	45.80	49.75	49.42	49.32	49.69	48.14
0.45	44.34	48.32	48.76	46.38	46.67	49.65	48.83	48.50	**50.46**	47.99
0.50	44.32	48.79	49.47	47.10	45.97	50.15	49.70	48.08	49.65	48.14
Ø	44.95	47.42	**48.70**	45.66	45.53	47.38	47.63	46.17	47.81	46.80

All results of the five stages of experimentation described in this section were based on a five-fold cross-validation procedure, which provided an initial indication of the models' performance with a chosen set of parameters. As a next step, the best-performing models were then used to label the unseen samples that had been held back from the beginning. Since the results of the cross-validation were very similar, the best five models were chosen instead of just the best one. The accuracy values of the five chosen models, as well as the best average values for the two parameter dimensions, are highlighted in bold font and with cell borders in Table 7-3. The classification results that were achieved on the test dataset by these five models are shown in Table 7-4.

TABLE 7-4: CLASSIFICATION ACCURACIES FOR THE FIVE BEST VIDEO PROCESSING MODELS ON VALIDATION AND TEST DATA.

Template name	LIF threshold	Validation accuracy	Test accuracy
TAL_by_7	0.35	50.55%	49.17%
TAL_by_8	0.3	50.31%	49.35%
TAL_by_10	0.3	50.68%	49.33%
TAL_by_10	0.35	**51.35%**	49.42%
TAL_by_10	0.45	50.46%	**50.48%**

The best accuracy that could be achieved for the video processing model was 50.48% for the "TAL_by_10" template with a LIF threshold of 0.45. For all models except the last one, the classification results for the test data were slightly worse than those achieved during the cross-validation process. One possible explanation for this could be overfitting to the training data and, thus, limited generalisability of the models.

Running the retinal encoding software was substantially faster than the cochlear encoding process. Transforming all 500 samples for one template size took about five minutes. This circumstance facilitated and encouraged the parameter search process for the retinal encoding module. In contrast, the training times for the video models were about twice as long as that of the sound models, with larger network templates taking longer than smaller ones. However, no detailed runtimes were recorded since the video experiments were run in batches that were not separated and timed.

As for the sound experiments, the models based on the larger template sizes could not be run on the available hardware due to restrictions in RAM space, which resulted in memory errors. 16 GB were not enough to execute the current Java implementation of the NeuCube on network templates with more than 12,150 neurons. However, in contrast to the sound processing system, the video experiments could be run for the TAL_by_5 template, which added a set of models to the experiments.

The results for the video processing system were worse than expected and no best model configuration could be recommended at this stage. Instead, the results suggest that further work needs to be done to make the retinal encoding more robust and find better values for its parameters.

7.5 DISCUSSION

This section critically evaluates the results that were achieved with the visual processing system developed in this research on the selected JESTER gesture recognition dataset. It first gives an overview of works from the literature that used the same dataset to evaluate their models and then discusses briefly approaches for improving the system presented here in the future.

The JESTER dataset was set up as a benchmark by its creators, who made an effort to provide an easily accessible infrastructure to keep track of achieved results. On their website (https://20bn.com/datasets/jester/v1) anyone who has trained a model on the available data using the provided training and validation sets can upload their labels for specially provided test data. These labels are then assessed automatically and the results are published on the JESTER website. Some of the earlier benchmark attempts by other researchers were already included in the original publication by Materzynska et al. (2019). While the website has proven to be very useful in providing verified third-party results, detailed information about the employed system architectures is, unfortunately, only included in a minority of entries. Further published results on the JESTER dataset were identified by searching for works that had cited the paper by Materzynska et al. (2019). Pre-prints were excluded from this search because they were not peer-reviewed. Table 7-5 summarises the nine models that were found, ordered by ascending accuracy. Entries in the Accuracy column that are written in italics were self-reported by the respective authors but not verified by the tool on the JESTER website.

All works made use of some form of neural network to classify the JESTER data and the majority of these were convolutional neural networks. One group used a *Residual Network* (ResNet), first introduced by He, Zhang, Ren, and Sun (2016), which is an architecture that facilitates the creation and improves the efficiency of very deep neural networks. An extension of this architecture, ResNeXt (Xie, Girshick, Dollár, Tu, & He, 2017) was used in another of the identified benchmark papers. According to the system design descriptions of the selected works, all but one of them focused on capturing the dynamic features of the data in their respective system architecture. The exception was the method used by the authors of the original paper that introduced the dataset. In this, the authors ran some initial experiments using a standard three-dimensional convolutional neural network to set a benchmark for other researchers. Three of the identified works created a data encoding algorithm that focused on capturing movements. For the remaining six, feature extraction and data interpretation were executed jointly. Three papers specifically mentioned using a three-dimensional network architecture.

TABLE 7-5: OVERVIEW OF WORKS FROM THE LITERATURE THAT REPORTED RESULTS ON THE JESTER DATASET.

Reference	**Method**	**Accuracy**
Materzynska et al. (2019)	Three-dimensional (3D) convolutional neural network (CNN)	93.87%
Zhou, Andonian, Oliva, and Torralba (2018)	Temporal relation network	94.78%
K. Yang et al. (2018)	Temporal pyramid relation network	95.34%
B. Yu, Luo, Wu, and Li (2020)	Attentive feature fusion framework	*95.77%*
M. Lee, Lee, Son, Park, and Kwak (2018)	Motion feature network for encoding and ResNet CNN for classification	96.22%
Köpüklü, Köse, and Rigoll (2018)	Motion fused frames for encoding and deep CNN for classification	96.28%
Jingran Zhang, Shen, Xu, and Shen (2020)	Temporal reasoning graph	*96.9%*
Köpüklü, Gunduz, Kose, and Rigoll (2020)	ResNeXt-101 3D CNN with specialised feature detector	*96.99%*
Y. Zhang et al. (2020)	Deformable ResNeXt 3D CNN	*97.1%*

All reported accuracy values were well above 90%, which is a remarkably high correct classification rate given the difficulty of the data. However, the highest three reported values were not verified by the JESTER website. Still, these values show that the problem presented by the dataset is solvable, which naturally raises the question of why the system presented in this thesis performed so poorly.

As the creators of the JESTER dataset, Materzynska et al. (2019) ran a series of experiments in which they investigated the relationship between the number of training samples and the achieved classification accuracy. They found a direct causality and noted that when they reduced the number of training samples to 100 samples per class, their accuracy dropped by about a third to 62.4%. Their explanation for this finding was that the huge diversity of video backgrounds made it hard for their CNN to generalise from the data and extract the most important features. This means that by reducing the number of training samples in the experiments described in this chapter, the difficulty for the classification system was inadvertently increased, even though the number of gesture classes was reduced at the same time. Therefore, one simple solution that would probably lead to a better performance of the system introduced here could be to increase the number of training samples.

But would this setup still be considered biologically plausible, as was the original goal of the work presented in this thesis? In nature, the human brain tends to require low numbers of "training samples" before being able to grasp the meaning of a presented stimulus and then generalise from those few examples. Arguably, a human would be able to classify all samples in the JESTER test dataset correctly after seeing only a handful of videos from the training data. The models listed in Table 7-5, although displaying some biologically inspired properties like a 3D network design, did not have the declared aim of being biologically plausible. It is hence only of limited use to draw inspiration from those systems. Instead, the promising but initial characteristics of the video processing architecture presented in this thesis should be improved in future work.

The greatest leverage for this kind of improvement can be expected from modifying and enhancing the retinal encoding algorithm for both the signal transformation (modelled after the biological process of phototransduction) and the grouping of signals into blocks (modelled after receptive fields). Looking at the spikes that were created by the encoding module provided a hint as to why the SNN could not classify the encoded data well. From visual inspection, there seems to be a lot of "noise" in the spikes, i.e., surplus spikes that are likely not directly related to meaningful information in the stimulus, even when using the optimised encoding parameters. This can be seen in Figure 7-10, where four spike samples are visualised. Black dots in the images represent spikes. The samples were created for the TAL_by_5 template and hence contain 128 columns for the input neurons. The first 64 columns were filled by the peripheral greyscale encoding and the remaining 64 columns were filled by the foveal colour encoding. This distinction is highlighted by the red line. The time points of the samples are shown as rows in the images.

While the top two samples belong to the "Drumming Fingers" class (Class 1), the bottom two samples belong to the "Stop" class (Class 2). From visual inspection of the classes, it seems that there are no obvious commonalities between samples of one class and no distinctive variations between samples of different classes. This absence of discriminative features in the spike patterns likely impacted the performance of the video processing model. The SNN simply could not find patterns because there existed none or only very weak ones in the encoded data, even if they did exist in the original stimuli, as is evident by the benchmark results reported in the literature.

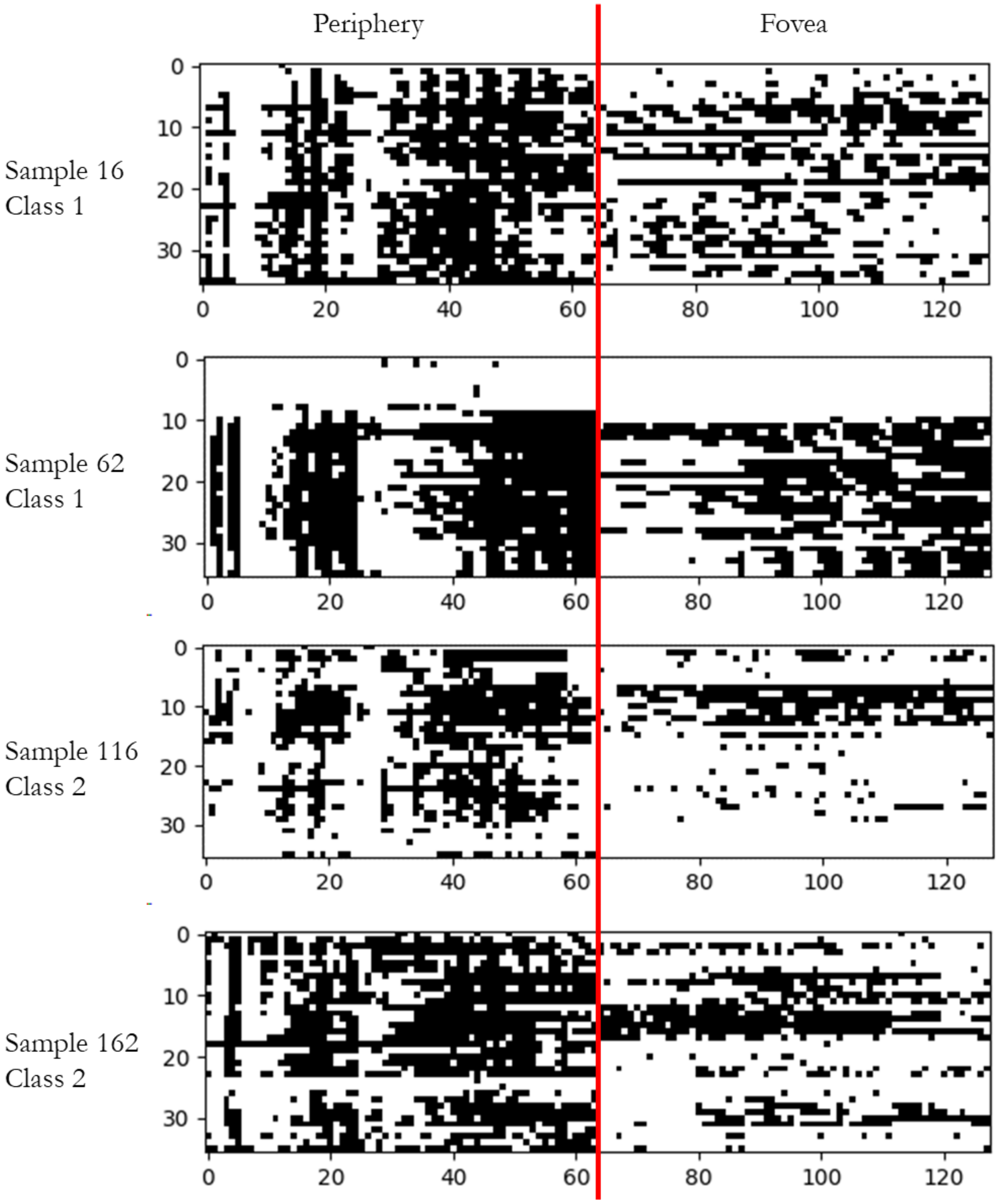


FIGURE 7-10: VISUALISATION OF THE SPIKES THAT WERE CREATED BY THE RETINAL ENCODING MODULE FROM FOUR SAMPLES OF THE JESTER DATASET FOR THE TAL_BY_5 TEMPLATE.

As identified above, one possible way to address the shortcomings of the model's performance could be to increase the biological plausibility of the encoding and blocking algorithms. If the spikes captured the discriminative features of the videos better, a higher classification accuracy could be expected. The current retinal encoding module draws on the functionality of the Dynamic Vision Sensor, which detects and encodes temporal changes of brightness. However, this is just one property of the multiple functionalities that have been observed in retinal ganglion cells (Goebel et al., 2012, p. 1305; Swanston & Wade, 2013, pp. 137-138).[59] For example, the algorithm presented in this thesis does not distinguish between magno and parvo retinal cells, even though they play an important role in spatial processing (Bruce et al., 2003, pp. 45-47). Furthermore, the encoding algorithm used here has no equivalent to the centre-surround receptive fields observed in the retina (Kuffler, 1953). Although the blocking algorithm attempted to summarise spatially close signals, the distinctive patterns that are created by this mechanism in biology could not be replicated here. Lastly, the dynamic nature of the video data might not have been captured as intended. Based on suggestions by Paulun et al. (2018), the blocking algorithm attempted to simulate saccades by shifting the focus area of the encoding to the most active part of the frame. However, the locations of the associated input neurons had to stay static due to implementation restrictions of the JNeuCube software, effectively eliminating the gained improvement. This issue could be addressed in the future by either choosing a different SNN model implementation or by connecting more than the required number of blocks to the static input neurons and then only sending signals through the dynamically activated channels.

[59] See Section 3.3.1 for a more detailed discussion.

7.6 CHAPTER SUMMARY

This chapter presented an experiment on a novel biologically inspired video processing system using the JESTER gesture recognition dataset. Several system parameters were optimised during the five stages of experimentation. The best performance was achieved on the "TAL_by_10" template with a LIF threshold of 0.45, which classified 50.48% of unseen test samples correctly. With regards to Research Question 3a in Section 1.3, which asked about the comparative performance of the model, it was found that all other studied models from literature performed much better than the model tested here. Research Question 3c, which asked about the influence of network size on model performance, could also not be answered affirmatively using the experimental setup in this chapter. It was found that most model configurations performed within a similar accuracy range, which was only about twice as good as by-chance selection. Even though the visual processing model did not perform as expected, it still constitutes an important part and necessary first step of the combined audio-visual model, which is evaluated in Chapter 8. It was concluded that the retinal encoding module should be revised in future work to include more biologically plausible mechanisms.

8 PROTOTYPING THE AUDIO-VISUAL SYSTEM

"Life is infinitely stranger than anything which the mind of man could invent."

– Sherlock Holmes in A Case of Identity

8.1 CHAPTER OVERVIEW

The final set of experiments that were performed for the work presented in this thesis ties back to its initial intention of studying the perception of language and semantic concepts as described in Section 1.2. In the experiments described in this chapter, language data consisting of speech files and gesture videos were analysed using the combined audio-visual processing system. The dataset that was used here contained five signs from **New Zealand Sign Language** (NZSL) and their equivalent spoken words. The data were encoded into spikes using the cochlear and retinal encoding modules described in Sections 5.3.1, 5.4.1, and 5.4.2. The spikes were then mapped into their respective tonotopic and retinotopic locations, as described in Sections 5.3.3 and 5.4.4, but this time, both sound and video data were entered into the same network model and analysed together.

As for the two unimodal sets of experiments described in Chapters 6 and 7, a classification task was performed on the dataset and different parameter settings were compared. This chapter first describes how the dataset was created and how it was analysed with the model. In addition to classifying the combined data, the speech and sign samples were also modelled separately to evaluate if their combination improved the unimodal classification accuracies. The results of these three experiments are presented and conclusions are made about the model and its optimal configuration. Furthermore, the trained networks are visualised to assess which neurons and connections in the SNN were activated by the data and if any insights could be drawn from the observed patterns. The chapter closes with a discussion of the advantages and limitations of the model and an outlook on its potential future use cases.

8.2 DATASET CREATION AND DESCRIPTION

No dataset was publicly available that combined signed videos of NZSL with their spoken word equivalents. Therefore, the dataset that was used for the experiments described in this chapter was created as part of this research from videos of five signs that were recorded by the thesis author. The corresponding speech samples were taken from the Tensorflow Speech Commands dataset (Warden, 2018), which was downloaded from download.tensorflow.org/data/speech_commands_v0.02.tar.gz on 21st October 2019.

Sign languages are a group of languages that are used by the Deaf[60] community and contain only visual components. These include manual signs and fingerspelling, but also lip movements, facial expressions (used, for example, to distinguish between a statement and a question), and body movement. Sign languages are usually based on the spoken language in the region in which they were developed – evident, for example, in lip movements that support the manual signs and follow the mouth shape of the spoken word. However, the grammar and sentence structure of sign languages is different from that of the related spoken language and no direct conversion is possible.

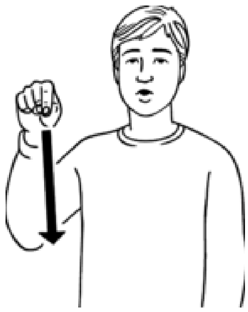
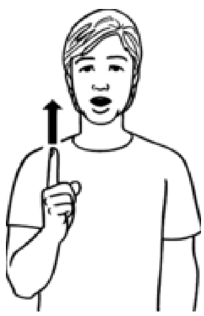
As part of her PhD programme, the thesis author attended a two-semester course called "New Zealand Sign Language and Deaf Culture" to immerse herself more in the topic area and gain a deeper understanding of the functioning and usage of NZSL. Based on British Sign Language, NZSL was developed by hearing-impaired New Zealanders with European heritage and later enriched by Māori concepts. Due to its history of suppression in the country's classrooms until the 1980s, several regional variants formed across the country. In 2006, NZSL became the third official language of New Zealand, which has improved the accessibility of the Deaf community to government and community services. As part of this recognition, an interactive NZSL dictionary was created and is maintained by Victoria University of Wellington that today contains over 4,500 entries (McKee, McKee, Pivac Alexander, Pivac, & Vale, 2011).

The five signs that were used for the experiment described in this chapter were chosen in conjunction with the gestures included in the JESTER dataset studied in Chapter 7. The classes in the JESTER dataset were compared to signs from NZSL and five matching gestures were found. The initially chosen words for the experiment described in this chapter were "bird", "down", "stop", "up", and "who". However, the Tensorflow Speech Commands dataset, which was the source of the equivalent spoken words, did not contain

[60] While the capitalised term *Deaf* is used to refer to the cultural identity of hearing-impaired people and other sign language users, the lower-case term *deaf* is used to describe the hearing ability.

the word “who”. Therefore, this class was replaced with the sign and the spoken word “tree” as both “tree” and “who” are signed in a similar position of the body frame and both contain movement of the whole hand. The final chosen signs, “bird”, “down”, “stop”, “tree”, and “up”, were then signed and recorded by the thesis author based on the videos from the NZSL dictionary. Pictograms of these five signs and links to their dictionary videos are shown in Table 8-1. Each sign was repeated 50 times during the recording and the layout of the videos (plain, light background while wearing a plain, dark t-shirt) was intended to match that of the videos in the dictionary. The frame size of 320 × 180 pixels was chosen to match the frame size ratio of 16:9 that was already used in the video experiments described in Chapter 7. This meant that the developed block numbers and layouts could be reused. The dataset was made publicly available on GitHub at https://github.com/AnneWendt/PhD-thesis/NZSL/videos.

Table 8-1: The five words selected as a dataset for the sign language experiment.

Word	Sign pictogram	Link to video
Bird		https://www.nzsl.nz/signs/4508
Down		https://www.nzsl.nz/signs/5737
Stop		https://www.nzsl.nz/signs/425
Tree		https://www.nzsl.nz/signs/1299
Up		https://www.nzsl.nz/signs/5767

8.3 EXPERIMENTAL SETUP

The main goal of the exploratory evaluation study described in this chapter was to test if combining the auditory and visual modalities was plausible and if the combination could improve the classification accuracy of the system compared to its two unimodal forms. For this, three experiments were performed: one on just the sound data, one on just the video data, and one on the combined audio-visual data. As a secondary objective, the best parameter configuration for the combined model was sought. This section first describes how the data were combined and entered into the three models. This is followed by an overview of the fixed and variable model parameters and, finally, the training and testing procedure of the experiments.

8.3.1 DATA PREPARATION

The procedures for preparing the sound and video data largely followed those of the unimodal sound and video experiments as described in Sections 6.3.1 and 7.3.1. The sound and video files were first converted into spikes using their respective encoding methods and identified parameters. However, for the combined model, the encoded signals had to be temporally aligned before they could be entered into the network. The theoretical design considerations of this process were described in detail in Section 5.5.2 and the corresponding code can be found in Appendix A, Listing IX.

The different setup of two encoding algorithms caused the sampling rate of the sound files to be much higher than that of the video files. Since this discrepancy could lead to a dominance of the auditory signals in the network, attempts were made to both shorten the sound files and elongate the video files. Using the method described in Section 5.5.2, the length of the sound files was reduced by about 38%, while each spike in the video files was repeated three times, effectively slowing down the presentation speed of the visual data to the network. Two corresponding files were then combined into one sample that was entered into the network as a semantic unit. An example of such a combined spike file is visualised in Figure 8-1. The vertical axis represents the index of the input neuron into which the signals were entered. Since the sample shown in the figure was taken from data that was used for a model built with the MNI_by_4 template, the first ten rows show auditory input and the remaining 39 rows show visual input.[61] The horizontal axis shows the time steps (i.e., data points) of the sample and each black dot in the diagram represents a spike.

[61] The numbers of auditory and visual input neurons were calculated based on volumetric ratios as described in Sections 5.3.2 and 5.4.3.

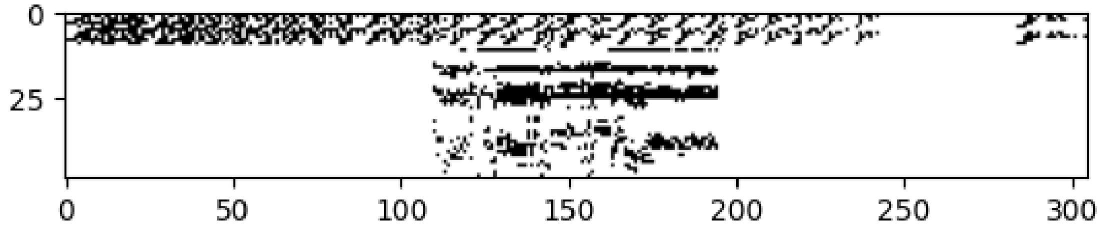


FIGURE 8-1: COMBINED SAMPLE OF SOUND AND VIDEO DATA FOR A NETWORK MODEL USING THE MNI_BY_4 TEMPLATE.

After all data for the three experiments were encoded, they had to be mapped into the neural network models. For the two unimodal experiments, the same procedure as described in Sections 6.3.1 and 7.3.1 was used. However, for the combined files, the locations had to be merged since only *one* file containing the locations of the input coordinates could be passed into the JNeuCube software. This was done by simply appending the location file used for the visual system to the end of that used for the auditory system. The given coordinates in the combined location file then corresponded to the neuron indices shown in the vertical axis of Figure 8-1. No further processing was needed in this step because the brain shape of the network facilitated a straightforward, biologically plausible spatial integration of the two modalities.

8.3.2 MODEL PARAMETERS

As for the unimodal auditory and visual processing models, there were a variety of parameters that could be adjusted in the experiments described here. For the encoding of the sound data, the parameters that were identified as being most biologically plausible were used. These can be found with a detailed explanation in Section 5.3.1. For the video encoding, the experiments in Chapter 7 showed that more work needed to be done to find the most suitable set of pixel and block threshold parameters. However, a set of encoding parameters was identified that was deemed best for the investigated data. For better comparability, the experiments described in this chapter used the same values as for the final experiment in Chapter 7. These values are summarised in Table 8-2.

Also, for better comparability and because little is known about their influence on the model performance, all parameters related to the network model were kept the same as for the experiments described in Chapters 6 and 7. These values are summarised in Table 8-3.

TABLE 8-2: RETINAL ENCODING PARAMETERS FOR THE EXPERIMENT ON THE SIGN LANGUAGE DATASET.

Parameter	**Value**
Peripheral pixel threshold	3
Peripheral block threshold	0.3
Foveal pixel threshold	5
Foveal block threshold	0.5

TABLE 8-3: NETWORK MODEL PARAMETERS FOR THE EXPERIMENT ON THE SIGN LANGUAGE DATASET.

Category	**Parameter name**	**Value**
Neuron model	LIF reset voltage	0 (simulating a full neural discharge)
	LIF refractory time	4
	LIF resistance	1
	LIF capacitance	10
Network initialisation	SWC weight range	Minimum -0.1 and maximum +0.1
	SWC radius	2.5 times the network size scaling factor
	SWC positivity rate	0.7
STDP	*A* positive/*A* negative	0.001
	τ positive/τ negative	10
	Weight boundaries	Upper bound +2 and lower bound -2
deSNN	Modulation factor	0.8
	Drift positive/negative	0.005
Classification	kNN *k*	3
	Number of folds	5
	Training rate	0.7

Another conclusion that was drawn from both the sound and the video experiments was that Izhikevich neurons (Izhikevich, 2003) performed poorly in the given system setup and that the threshold of the LIF neurons (Brunel & van Rossum, 2007) typically influenced the classification accuracy of the model. For these reasons, the models for the experiments described in this chapter contained only LIF neurons and the same 18 threshold values were used for parameter optimisation that had been used for the sound and video experiments described in Chapters 6 and 7:

0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50

One major deviation of the system setup that had been used for both the sound and the video experiments was that here, only two template sizes were used, MNI_by_4 and TAL_by_8. Since parameter optimisation was only a secondary objective for this study and the influence of network size had been investigated in the unimodal experiments described in Chapters 6 and 7, the focus here was shifted to comparing results of sound, video, and combined data. The choice of network template was informed by averaging and comparing the results of the two unimodal experiments. Since MNI and Talairach templates had been found to behave differently in the unimodal experiments, the on average best-performing template was chosen for each type. The best-performing MNI template was MNI_by_4 with 67.16% accuracy and the best-performing Talairach template was TAL_by_8 with 66.30% accuracy.[62] Therefore, only these two templates were studied here. Coincidentally, both were medium-sized, with MNI_by_4 containing 3,747 neurons and TAL_by_8 containing 2,960 neurons. Since it was found that the size of the network influenced the model's performance, this circumstance meant that the classification results for the experiments here were more easily comparable with each other.

8.3.3 MODEL TRAINING AND TESTING PROCEDURE

As for the studies on sound and video data, in the experiments performed here, a five-fold cross-validation method was first applied to 70% of the data for the parameter optimisation process. Each parameter set was run 30 times and all five models for each run of the cross-validation were initialised with the same set of connections. The best-performing models were then tested on the unseen 30% of samples to provide a robust verification of the results (Brownlee, 2020).

[62] The detailed values for the averaged results can be found in Appendix B, Table B-61.

For the optimisation process, three experiments were performed on each of the two network templates: one on just the sound data, one on just the video data, and one on the combined data. That way, the three setups could be compared and conclusions could be drawn about how the modality influenced the results. The unimodal experiments described in this chapter also acted as an extension to the findings from Chapters 6 and 7. Similar to those experiments, the network parameter that was optimised here was the LIF threshold as described in Section 8.3.2.

The final step in the experiments described here was visualising the trained network models to investigate if the five words in the combined dataset activated visibly different patterns. For this, a network model that had been trained on all samples of the dataset was replicated five times. All samples from only one class were then passed into the model for processing and the neural activity in the network was recorded. A "pruning" algorithm that had recently been implemented in the JNeuCube software was employed to remove those neurons and connections in the network that had not exhibited any activity when presented with these selected samples. This meant that only those neurons and connections that were relevant for processing a particular word could be visualised. The code for the three-dimensional visualisation of the pruned network models can be found in Appendix A, Listing X.

8.4 RESULTS

The averaged cross-validation results of the models built for this study are shown in Table 8-4. While rows correspond to the LIF threshold, the columns show the six combinations of template size and input data. The last column and row show the average classification accuracy for the respective parameter. Cell background colours indicate how the values compare to each other, with green indicating better performance, yellow indicating average performance, and red indicating worse performance. The detailed results of the 30 runs for these experiments can be found in Appendix B, Table B-62 to Table B-67. Figure 8-2 visualises the values from Table 8-4 in a line diagram. The blue-coloured lines with the triangles show the results for the TAL_by_8 models, while the orange-coloured lines with the diamonds show those for the MNI_by_4 models.

TABLE 8-4: CLASSIFICATION RESULTS FOR THE EXPERIMENT ON THE SIGN LANGUAGE DATASET. ALL VALUES ARE GIVEN IN %.

Size	**MNI_by_4**			**TAL_by_8**			**Ø**
LIF	audio	video	combined	audio	video	combined	
0.01	46.79	71.00	55.00	45.94	71.66	50.10	56.75
0.02	47.77	70.58	62.47	47.44	68.30	48.54	57.52
0.03	49.91	74.00	62.54	46.60	80.88	49.15	60.51
0.04	51.65	74.67	65.60	43.37	76.09	48.57	59.99
0.05	53.16	74.62	67.24	48.86	74.80	61.96	63.44
0.06	54.79	75.88	68.03	50.62	80.50	65.82	65.94
0.07	54.49	75.91	70.16	52.52	83.45	68.88	67.57
0.08	54.82	75.10	69.95	53.62	78.95	69.61	67.01
0.09	57.04	75.44	71.06	56.20	77.74	69.57	67.84
0.10	57.75	74.87	72.39	57.93	83.09	71.02	69.51
0.15	56.59	72.73	**73.02**	60.19	76.92	75.59	69.18
0.20	50.59	71.42	71.04	55.52	78.95	74.50	67.00
0.25	48.57	70.26	**71.24**	54.10	78.55	75.90	66.44
0.30	48.19	68.64	**69.69**	52.34	75.66	**76.28**	65.13
0.35	45.60	68.71	**69.05**	50.13	66.27	**76.29**	62.67
0.40	43.27	68.67	67.73	48.70	70.72	**74.02**	62.19
0.45	44.19	69.03	67.78	47.11	70.25	**73.52**	61.98
0.50	43.46	68.29	66.40	44.62	74.80	72.32	61.65
Ø	50.48	72.21	67.80	50.88	75.98	66.76	64.02

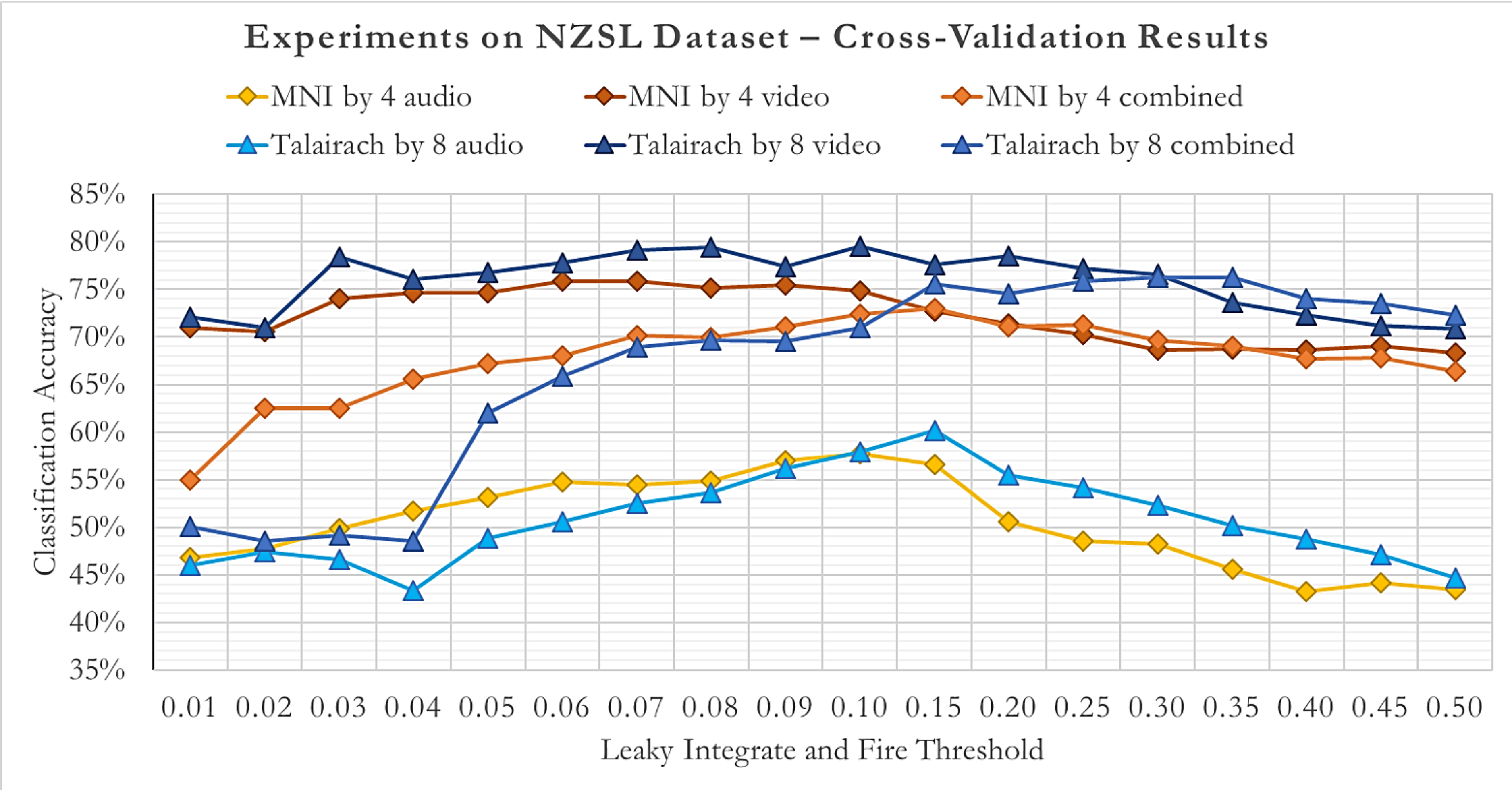


FIGURE 8-2: CLASSIFICATION RESULTS FOR THE EXPERIMENTS ON SPEECH, VIDEO, AND COMBINED DATA FROM THE SIGN LANGUAGE DATASET.

As can be seen in both the table and the figure, the classification accuracies varied widely from around 40% to just over 80%. Surprisingly, the models trained on only the sound data performed much worse than those trained on only the video data. For the experiments described in Chapters 6 and 7, where two different datasets had been used, this relationship between the two modalities was reversed.

The highest accuracy of all experiments was 83.45%, which was achieved using the TAL_by_8 template on just the video data with a LIF threshold of 0.07. The models using this template and input data were also the on average best-performing, with a mean classification accuracy of 75.98%. In contrast, the lowest accuracy of all experiments was 43.27%, which was achieved on the MNI_by_4 template trained on just the sound data with a LIF threshold of 0.4. This combination of template and input data was also the worst on average with a mean classification accuracy of 50.48%.

The performance difference between sound and video data in the experiments performed here can be explained by the quality of the input data. The Tensorflow speech dataset consisted of recordings of numerous people with varying voices, accents, and background noise. On the other hand, the video sign dataset was collected by the thesis author with an emphasis on good perceptibility, modelled after the videos from the NZSL dictionary. This meant that the sound data were relatively "dirty", while the video data were relatively "clean", which influenced their distinguishability by the models.

The range of the average LIF thresholds was found to be much smaller than that of the three input data types. On average, the best LIF threshold was 0.1 with a mean classification accuracy of 69.51% across all six model setups, while the worst LIF threshold was 0.01 with an accuracy of 56.75%. While the two models trained on only the sound data had a distinctive peak at a threshold of 0.1 and 0.15, the models trained on the video data did not exhibit this behaviour and instead stayed comparatively constant. The results of the two combined models are characterised by a steady increase in classification accuracy until a LIF threshold of about 0.15, where they start to plateau. As briefly discussed in Chapters 6 and 7, lower LIF thresholds lead to more spiking activity in the network, which in turn decreases the impact of single spikes on the final classification outcome. Higher thresholds, on the other hand, support the SNN's ability to filter out noise and can thus improve the distinguishability of patterns and the performance of the model.

A promising finding of this study was that for some LIF thresholds, the results for the models trained on the combined dataset surpassed those of the models that had been trained on unimodal data. For each template, four such instances were identified. These are highlighted in Table 8-4 by bold font and a cell frame. While the differences for the MNI template are very small (≤ 1%), three of the four improvements on the Talairach template were found to be statistically significant in a t-test with $p = 0.05$. These models were better capable to identify the patterns in the data when they were trained on both modalities. This indicates that in these models, the combination of auditory and visual input data created some level of synthesis in which the two modalities supported each other.

On the other hand, the models where this synthesis could not be observed might not have been able to extract the most useful bits of information. An explanation for this behaviour could be that too much noise was introduced by the relatively unclean sound data, which then adversely affected the distinguishability of the combined samples. The performance of the combined models was worst for LIF thresholds smaller than 0.07, which indicated that for larger thresholds, the noise that was introduced by the sound component could be filtered out by the network. In future work, the experiments combining sound and video data should be repeated with datasets that are in comparable states of pre-processing to investigate if the models will behave similarly.

The next step of the result analysis was to identify the two best-performing models for each template that were trained on the combined data and use these to label the test samples that had been held back until this point. This test accuracy could then be reported as the final results of the audio-visual processing system presented in this thesis and compared with similar models found in the literature. Since the best classification accuracies were quite similar, two models of each template were chosen for this final verification instead of just one. The two best MNI_by_4 models were those with LIF thresholds of 0.1 and 0.15 and the two best TAL_by_8 models were those with LIF thresholds 0.3 and 0.35. Table 8-5 shows the classification accuracies of these four models for the cross-validation and test processes. The best accuracy that could be achieved with the audio-visual processing model on the combined sign language data was 76.98% for the TAL_by_8 template with a LIF threshold of 0.35.

TABLE 8-5: VALIDATION AND TEST ACCURACIES FOR THE TWO BEST MODELS OF EACH TEMPLATE THAT WERE TRAINED ON THE COMBINED SIGN LANGUAGE DATA.

Template name	LIF threshold	Validation accuracy	Test accuracy
MNI_by_4	0.1	72.39%	74.78 %
MNI_by_4	0.15	73.02%	76.11 %
TAL_by_8	0.3	76.28%	75.53 %
TAL_by_8	0.35	**76.29%**	**76.98 %**

Especially for the models built with the MNI_by_4 template, the test accuracies were better than those achieved during the cross-validation. While a reason for this could be that more training data were available for the test (all 70% training samples instead of just 56% during the cross-validation), this could not conclusively explain why the same improvement could not be observed for the TAL_by_8 models.

Across the three types of input data that were used for the experiments described in this chapter, namely auditory, visual, and combined audio-visual, comparing the accuracies of the two brain templates identified further differences. It could be observed that the MNI_by_4 models performed better with smaller LIF thresholds than the TAL_by_8 models, except for the model that had been trained on just video data. For LIF thresholds larger than 0.1, this relationship was reversed and the Talairach models performed better than the MNI models.

Further to this accuracy comparison, it was observed that for the combined data, the training times of the Talairach models were almost three times as long as those of the MNI models. The model execution times, summarised over all tested LIF thresholds for each template and type of input data, are shown in Table 8-6. A similar time difference between the MNI and Talairach atlases had already been observed in the sound experiments described in Chapter 6, where no feasible explanation for this phenomenon could be found. While this difference in training times could also be observed in the experiments presented here for the models trained on just one data type, it was much more pronounced when using the combined data.

TABLE 8-6: RUNTIMES OF THE EXPERIMENTS ON THE SIGN LANGUAGE DATASET, SUMMARISED OVER ALL LIF THRESHOLDS.

Brain template	**Input modality**	**Number of neurons**	**Runtime**
TAL_by_8	Audio	2,960	3 d 0 h 30 m 31 s
TAL_by_8	Video	2,960	6 h 20 m 23 s
TAL_by_8	Combined	2,960	3 d 22 h 54 m 42 s
MNI_by_4	Audio	3,747	2 d 20 h 22 m 3 s
MNI_by_4	Video	3,747	3 h 56 m 43 s
MNI_by_4	Combined	3,747	1 d 10 h 12 m 35 s

The final step of the result analysis was a visualisation of the differences between the five signs that comprised the dataset. For this, the best parameter configuration for each template was selected to create two models, which were then trained on the full dataset. The MNI_by_4 template was used with a LIF threshold of 0.15 and the TAL_by_8 template was used with a LIF threshold of 0.35. After the training was completed and the dynamic patterns of the data had modified the connections in the network, all samples belonging to one of the five classes were passed through the model again for processing. The activity in the network that could be observed during this one-time pass of data was recorded for each combination of model and class and is visualised in Figure 8-3. All models in Figure 8-3 are shown from the left back side of the brain. However, since the SNN models are three-dimensional, they are difficult to visualise from just one angle, so more figures viewed from the right, top, and back sides of the brain can be found in Appendix B, Figure B-1, Figure B-2, and Figure B-3, respectively. The grey dots in the figures represent all neurons in the network template. Blue circles around them indicate that they were activated at least once during the one-pass presentation, while red crosses mark the input neurons. The blue lines highlight those connections that were activated.

When comparing the visualised activity for the different classes, only very few differences could be identified. For example, the class "bird" was the only one that did not evoke a connection from the left auditory cortex to the frontal cortex in the MNI_by_4 template. In the TAL_by_8 template, the class "tree" was the only one that did not exhibit activity in the right auditory cortex. However, even though the activated patterns of neurons and connections in the networks looked very similar, they were sufficiently unique for the model to identify the correct class label for about three out of four unseen samples.

While the visible differences between classes were found to be minimal, there is a noticeable disparity between the two templates. For the MNI template, neural activity extends across all parts of the network, whereas for the Talairach template, it seems to be confined to the area around the visual cortex. This observation is likely related to the previous finding that the two templates achieved different classification results. The Talairach model might have performed better because it put more emphasis on the information in the visual input data that was easier to classify. The MNI model, on the other hand, extended the connections from the input neurons to other areas of the network. However, no direct route between auditory and visual cortices could be found.

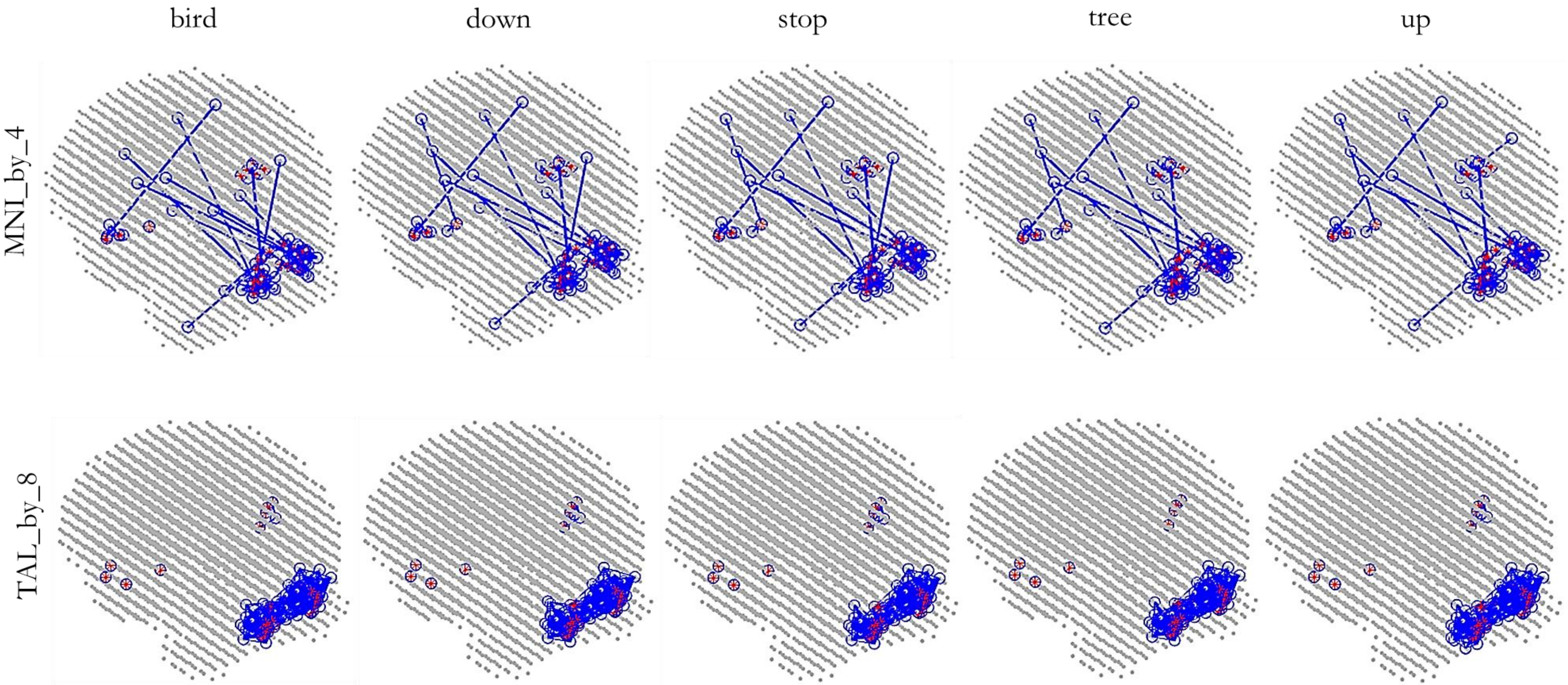


FIGURE 8-3: VISUALISATION OF THE NEURONS AND CONNECTIONS THAT WERE ACTIVATED FOR THE DIFFERENT CLASSES OF THE SIGN LANGUAGE DATASET IN FULLY TRAINED MODELS. THE VIEWING ANGLE IS FROM THE LEFT BACK OF THE BRAIN.

8.5 INSIGHTS AND ANALYSIS

The results of the experiments described in this chapter were both surprising and promising. Surprising, because the video processing system unexpectedly performed much better than the sound processing system, and promising, because combining the two modalities led to equal or better classification accuracy compared to their unimodal counterparts. This is an encouraging observation that spurs further interest to investigate potential future enhancements and use cases that can have an impact on people's lives.

Since the NZSL dataset that was used in this experiment was created as part of this research, no direct benchmarking comparisons, like those in Chapters 6 and 7, could be performed. However, it was possible to evaluate parts of the models with each other and with existing works.

The TensorFlow speech command dataset that formed the sound component of the newly created NZSL dataset has been used by more than 300 research groups according to Google Scholar, with varying results. The winning group of a Kaggle competition on the TensorFlow dataset achieved 91.06% classification accuracy on all 30 words (Kaggle Inc., 2018). In contrast, the unimodal sound processing system presented in this thesis achieved a maximum accuracy of 60.19% on the reduced dataset with five words. Since the algorithms that were used by the Kaggle competition winners were not made public, it can only be speculated that they might have been more specialised in the task of speech recognition than the brain-inspired methods that were used here. Compared to the promising results reported in Chapter 6, the decreased performance of the sound processing model that was seen in the experiments described in this chapter was possibly caused by the higher difficulty of the TensorFlow dataset, which was crowd-sourced by multiple speakers with a large variety of accents and background noise levels. In contrast, the data from FSDD, which had been used in the sound benchmark experiments described in Chapter 6, were all recorded by the same speaker in a noise-free environment.

Likewise, the improved classification accuracy of the unimodal video processing system compared to the results reported in Chapter 7 can probably be attributed to a lower difficulty level of the videos that were collected for the NZSL dataset. Their static, light background in combination with the signer's plain, dark clothes as well as the fixed video angle and size of the signer meant that these videos were less challenging than those from the JESTER dataset used in Chapter 7.

As a general conclusion from these observations, both the unimodal sound and video processing systems will have to be made more robust before they can be applied in real-life settings that naturally exhibit a great stimulus variability.

On the topic of combining the two modalities, the observed improvements in the recognition rate of the bimodal models can be regarded as a biologically plausible feature. Although the low accuracy of the auditory model likely impacted the performance of the visual model, the classification accuracies of the combined model were closer to those of the better performing video model. The combined system learned to put more emphasis on the modality that provided more reliable signals for identifying a stimulus. In the same manner, the human brain chooses to focus on data from those sensory organs that can deliver more reliable information, although a bias towards vision has been observed (Meijer, Veselič, Calafiore, & Noppeney, 2019; Witten & Knudsen, 2005). This enhancement of systems that employ multimodal cross-over was also reported by other research groups in previous work, mainly in the area of speech recognition that was supported by images or videos of lip movements (Noda et al., 2015; Rathi & Roy, 2021; Sun, Harwath, & Glass, 2016; Thermos & Potamianos, 2016).

Multimodal signal integration has also been applied to the specific area of sign language recognition, which was chosen as the application domain for the experiments described in this chapter. Several researchers found that combining information from different sensors enhanced their systems' recognition abilities. The underlying architectures used in those systems were usually Hidden Markov Models, Liquid State Machines, and Convolutional Neural Networks. No brain-inspired or SNN-based approach like the one presented in this thesis could be found in the literature. Furthermore, in contrast to the audio-visual system presented here, the modalities that were used in the models reported in the literature typically included a combination of visual and motion sensors that captured joint positions, movement trajectories, and/or 3D depth data. For example, Brashear, Starner, Lukowicz, and Junker (2003) developed a wearable translation system for American Sign Language that included a hat-mounted camera and a wrist band with an accelerometer, which significantly improved the recognition accuracy when combined. The same observations were made for Russian (Kagirov, Ryumin, & Axyonov, 2019; Kagirov, Ryumin, Axyonov, & Karpov, 2020), Chinese (Kamal, Chen, Li, Shi, & Zheng, 2019; Jihai Zhang, Zhou, Xie, Pu, & Li, 2016), and Arabic (Hassan, Assaleh, & Shanableh, 2019) sign languages where data from videos, depth sensors, and accelerometers were used. Due to the brain-inspired signal mapping approach that was introduced in this thesis, more modalities can be added to the system in the future.

For example, information from motion sensors could be mapped into the motor cortex. The challenge of this endeavour would then be to find an appropriate encoding mechanism.

Although the related works that were found in the literature report good results for "sign language recognition", the tasks that were used to assess their models largely consisted of "single sign classification". From a usability point-of-view, these systems are, unfortunately, not very well suited to the communication needs of Deaf people, since they are usually approached from an engineering perspective with a focus on evaluating algorithms (Bragg et al., 2019; Hill, 2020). In this regard, the work by Brashear et al. (2003) is noteworthy because it included Deaf people in the design and development process and because the researchers tried to classify sequences of signs rather than single signs. There are further efforts to address the unique challenges of continuous sign language recognition (Hassan et al., 2019; Koller, Forster, & Ney, 2015), which is a more promising approach to real-life usefulness. However, these works need to consider linguistic characteristics, which significantly increases the difficulty of the problem. While hand gestures, which have conventionally formed the datasets used in the more engineering-focused systems, largely represent the vocabulary of a given sign language, its grammar and syntax are often coded in non-manual signs such as facial expressions, eye gaze, and body position (Caridakis, Asteriadis, & Karpouzis, 2014). Therefore, a system for sign language recognition should also be able to capture and interpret these linguistic features. In the best case, the final system includes a separate linguistic translation unit.

The audio-visual processing system presented in this thesis can so far create associations between spoken words and front-view videos of signers. This means that the system automatically captures manual and non-manual signs since the whole upper body is observed. Including motion data in future work as suggested above could further improve the recognition ability of the visual processing system, for example by capturing more precise finger movements. However, to be fully applicable in real-life scenarios, the "audio-visual-motor system" would need to also contain a forward-backwards language translation module. Without this, it can merely connect the observed NZSL signs to words from a related spoken language such as English or Te Reo Māori. On the other hand, these observed connections suggest that the introduced system architecture can be a promising approach for solving this task if the required components are added.

8.6 CHAPTER SUMMARY

This chapter presented an experiment on a novel biologically inspired audio-visual processing system using a newly created New Zealand Sign Language dataset with auditory and visual components. The performance of the models when trained on just one modality was compared to that of models trained on the combined audio-visual data. For each of the two studied templates, there were four LIF thresholds with which the combined models outperformed their singular counterparts. The best-performing combined model was based on the TAL_by_8 template with a LIF threshold of 0.35, which classified 76.98% of the unseen test samples correctly. The findings of this exploratory proof-of-concept study suggest that the capabilities and boundaries of the combined approach should be investigated further.

The research questions that were asked in Section 1.3 in relation to the combination of auditory and visual data were attempted to be answered in this chapter. Question 1b asked if neurological pathways could also be observed in the computational model. This could only be shown partially. While the models built with the TAL_by_8 template did not show any activity outside of the areas where the input signals were entered, more "long-distance" connections, as well as active connections into other areas of the network, could be observed in the models built with the MNI_by_4 template. Question 1c asked which aspects of human audio-visual processing could enhance the analysis of sound and video data. The experiments in this chapter showed that combining modalities in a biologically plausible way improved the classification accuracy of the computational system compared to their unimodal counterparts.

9 CONCLUSION

"Education never ends, Watson. It is a series of lessons with the greatest for the last."

– Sherlock Holmes in The Adventure of the Red Circle

9.1 THESIS SUMMARY

This thesis described a biologically inspired, computational sound and video processing system that enables biologically plausible signal combination in a three-dimensional SNN model. Chapter 1 introduced the topic area and presented the motivation for this work. In Chapter 2, an early pilot study was described that had been conducted at the beginning of the candidate's PhD studies. The conclusion of this pilot study was that the research direction should be adjusted and the focus should be shifted to modelling auditory and visual processing of the ears, eyes, and brain. Chapter 3, therefore, looked in detail at the underlying biological mechanisms of the hearing and vision processes as well as at the integration of multimodal signals in the brain. Chapter 4 then reviewed existing systems for audio-visual processing that had attempted to use biologically inspired algorithms. Based on the findings from Chapters 3 and 4, a sound and video processing system was developed to include multiple biologically inspired features and biologically plausible parameters. The full system design is described in detail in Chapter 5. The system was then evaluated on sound, video, and audio-visual data, which was described in Chapters 6, 7, and 8, respectively. Finally, this chapter concludes the thesis, answers the research questions, points out the key advantages and limitations of the proposed system, and discusses potential future directions for this research.

9.2 ANSWERING THE RESEARCH QUESTIONS

This section revisits the research questions that were formulated in Section 1.3. The overarching hypothesis that was postulated in this work was that copying biological mechanisms in a computational system would result in the demonstration of new behaviour and improved performance in audio-visual processing and classification. The main question that was asked was:

How can a computational model of audio-visual information processing be created that uses brain-inspired mechanisms to analyse those data, and what can be learned from such a model?

Three focus areas with three sub-questions each were derived to support answering this question in a systematic manner.

(1) The first set of questions looked at the **biological inspiration** of the system:

a. How can the biological background of audio-visual information perception and processing inform the design of an audio-visual computational model?

Approaches to answering this question were investigated in Chapter 3 by studying the relevant literature from biology and neuroscience. The chapter contains detailed explanations of the biological processes involved in the transformation of auditory and visual data, their pathways into higher cortical areas, and the extraction of signal characteristics by the auditory and visual cortices. The learnings from this literature collection informed the design of the model as described in Chapter 5. Several aspects of the biological processes served as inspiration for the functionality of the computational model, such as signal encoding, signal mapping, and unsupervised learning.

b. Can neurological pathways of audio-visual information that are observed in the brain also emerge in a brain-inspired computational model?

As part of the description of biological processes in Chapter 3, Sections 3.2.2 and 3.3.2 also look at the pre- and post-cortical signal processing pathways of auditory and visual signals, respectively. Section 3.4 then explains how the two modalities are combined through the ventral and dorsal processing streams in which the signals are sent from their respective cortices to other areas of the brain. These pathways could not be observed in the visualisation of the bimodal models that were presented in Chapter 8. While the models built with the TAL_by_8 template only exhibited local activity around the input regions of the signals, the models built with the MNI_by_4 template showed some neural extensions into other areas of the network. It remains an open question and further investigation is needed to see if these connections can reliably retain their strength in different network setups or when more diverse data is used.

c. What aspects of the human audio-visual processing system can enhance the analysis of audio and video data?

It was found in the literature summarised in Chapter 3 that both the cochlea and the retina employ complex feature extraction mechanisms that are comparable to the pre-processing steps in machine learning systems. These enhance the capabilities of the brain to interpret the perceived signals. Furthermore, the integration of information from different modalities is a key aspect of the brain's deduction and understanding abilities. Therefore, these characteristics were considered to be important for the design of the computational system as described in Chapter 5. The feature extraction mechanisms can be found in the encoding algorithms, while multimodal operations were facilitated through the network's spatial layout. The experiments described in Chapter 8 showed that the system learned to pick the better performing modality when auditory and visual signals were combined. This led to an increased classification accuracy when compared to the unimodal systems.

(2) The second set of questions aimed at creating a suitable **system design**, which was heavily influenced by the findings of the previous answers:

a. How can audio-visual data be transformed (encoded) into electrical impulses for use in a spiking neural network?

The answer to this question was given in Section 5.3.1 for auditory input data and in Sections 5.4.1 and 5.4.2 for visual input data. Both the cochlear and the retinal encoding approaches were based on the corresponding biological mechanisms of signal transformation. The developed encoding algorithms expect sound and video files as input and can convert them into spikes that represent the dynamic features of the input data.

b. What is a biologically plausibly way to input (to map) sound and visual stimuli into the model?

The network templates that were used to create the SNN models were three-dimensional and brain-shaped, which means that locations in the network could be linked back to specific areas of the brain. Based on the concepts of tonotopy and retinotopy that are present in the brain, auditory and visual signals are reliably mapped into their respective cortical locations in a predefined, systematic manner. For the models developed as part of this research, this mapping was replicated using data from tonotopy and retinotopy studies, as described in Sections 5.3.3 and 5.4.4. Feature channels in the data (frequencies in the sound data and frame segments in the video data) could thus be entered into neurons of the network that were located and arranged in biologically plausible parts of the model.

c. How can the use of both auditory and visual data in one combined model be facilitated in a biologically plausible, yet computationally feasible way?

The shape of the network also facilitated the straightforward combination of the two modalities as described in Section 5.5. Since both data types could be entered into the model at the same time, only one round of training was necessary to connect the information.

(3) The final set of research questions was related to the aspect **system evaluation**:

a. How does the brain-inspired model perform on sound and video benchmark datasets compared to conventional approaches?

This question was answered in Chapters 6 and 7. The highest classification accuracy that could be achieved on a sound benchmarking dataset was 90.52% and the highest accuracy for a video benchmarking dataset was 50.48%. While the results of the sound processing system were almost as good as those of comparable systems, the video processing system performed unexpectedly poorly. The reasons that were identified for this were shortcomings in the retinal encoding module and too high a degree of reduction of the training data.

b. What are the advantages and disadvantages of using biologically plausible encoding and mapping approaches for processing audio-visual data?

The biggest advantage of the proposed system architecture was the straightforward integration of multimodal data, which was enabled by the novel, brain-inspired mapping approach. The biologically plausible encoding facilitated the extraction of spike-based features from the input data, which could then be entered into the spiking neural network. Since this work aimed at biological plausibility, limitations in performance on classical benchmarking datasets were expected and observed. The system's classification accuracy on the speech dataset was comparatively good but the video data could not be reliably labelled. However, the system showed promising results for the combined audio-visual data that can, at least in part, be attributed to the chosen signal mapping approach. A more general overview of the limitations of the proposed system architecture is given in Section 9.3 and potential future directions and applications are summarised in Section 9.4.

c. Does the size of the neural network influence the learning processes and performance of the model?

This question was investigated as part of the benchmark studies on sound and video data described in Chapters 6 and 7, respectively. While a clear relationship between template size, network training time, and classification accuracy could be observed in the sound processing system, this was not found in the video system, likely due to its generally poor performance. However, for the sound experiment, it was concluded that the choice of network template size should be informed by an analysis of the studied data and the expected outcome. There needs to be a problem-specific evaluation to verify if the improvement in classification accuracy is significant enough to justify the increased training time of the model. Within the limitations of the two experiments described in Chapters 6 and 7, it seems that the reduced signal information available to the smaller networks is still sufficient for a reasonable classification performance of those models. It is unclear why the smaller networks perform well in comparison to larger networks or if feature complexity plays a role in this network size effect observed herein. Moreover, there is insufficient evidence in this work to speculate on the cause. Further work is required that explores the impact of data cleanliness and information density as well as the specific shape of the network and its effect. Retaining the different levels of complexity and connectivity used in this research would allow direct comparison of the findings from this future work.

9.3 LIMITATIONS OF THE PROPOSED SYSTEM

While the limitations of the three systems have been discussed in the respective chapters, this section looks at the more general limitations of this study and its setup. Two main aspects affected the system design and experimental results.

Firstly, a level of **simplification** had to be used, which introduced biological inaccuracies. Since the brain's functionality has not yet been fully explored and understood by neuroscientists, the system design could only attempt to simulate biological processes within the confines of what is known to-date. This also involved skipping certain aspects of the sensory processing pathways and only including those parts that were deemed most important for the learning process. These choices most certainly influenced the model's performance and, thus, can only be considered a starting point for this exploratory research. Likewise, while the accompanying experimental studies gave initial insights and directions, they can not provide comprehensive empirical support for or against certain choices. In general, it can be noted that less biological plausibility seemed to negatively affect the model's classification accuracy. This could be observed when comparing the results of the unimodal experiments in Chapters 6 and 7. The auditory system with the highly biologically plausible cochlear encoding module performed significantly better than the visual system with the relatively implausible retinal encoding module. However, the performance also seemed to be largely affected by the cleanliness of the data as was shown by the experiments described in Chapter 8. More research is necessary to determine how more or less biological choices influence the model's learning and classification processes.

Simplifications were also necessary due to **limited computational resources**. While it was initially planned to make use of neuromorphic hardware to simulate the biological processes, this could not be executed due to financial constraints. Running the experiments on a standard PC for the most part prevented large-scale exploration of the model parameters. The NeuCube architecture used for this work is still relatively new, so only limited knowledge exists about the behaviour of its parameters in complex models. Choosing the best set of parameters for each model was, therefore, difficult and likely affected the results.

9.4 FUTURE DIRECTIONS

The best part about research is that there is never a lack of questions that are waiting to be answered. This is especially true for exploratory studies like the one that was presented in this thesis. The main question that remains at this point is:

How can the system be improved and made more useful for real-world applications?

One angle to investigate this question could be to test the system on more sound, video, and combined audio-visual data so that weak points can be identified and fixed. Especially for the NeuCube, but to a lesser extent also for the encoding and data compression algorithms, there were a lot of parameters whose behaviour could be studied in large-scale experiments. The retinal encoding module could draw more inspiration from its natural counterpart to make use of expertise that has been maturing in nature for millions of years. From a computational point of view, the runtimes of the experiments could be improved by using an implementation of the NeuCube that makes use of Graphical Processing Units or neuromorphic hardware as suggested by Scott (2015).

Apart from modifying the existing components of the model, its scope could be extended by more brain-inspired features. For example, more modalities, like olfactory, gustatory, and haptic perception, could be added as input data and mapped into their respective cortical processing areas. Another approach could be to integrate brain data collected by EEG or fMRI as suggested by Fong et al. (2018). The initial connection weights in the network, for example along the ventral and dorsal processing streams, could be pre-trained with actual brain data instead of being initialised randomly. This could help the system in identifying and connecting perceptual concepts.

9.5 CONCLUDING REMARKS

The work presented in this thesis originated from the researcher's ambition to replicate the brain's powerful abilities to process, connect, and understand information. Observing that promising AI and machine learning methods were originally based on biological mechanisms but existed as disjointed pieces, the researcher set out to design a model that would connect these pieces in a biologically meaningful way. This research work had an exploratory nature in that it investigated a novel synthesis of components and assessed the resulting model's potential use cases. While the novelty of the presented approach necessitated a focus on a limited number of components, the system's modular design enables independent improvements in future work.

The results achieved during the experimentation look promising for future investigation. Discrepancies observed between the performance of the auditory and the visual system were likely caused by a lack of biological plausibility and are expected to be minimised if a more nature-inspired approach is taken. Combining both modalities showed that the designed model is capable of connecting information. The network displayed emergent behaviour in which the simple functioning of the LIF neurons created observable patterns in the connections and neural activity.

Researchers around the world continually uncover more secrets of the human brain. It is my firm belief that the journey of AI, on which we have only just embarked in the grand scale of time, can only benefit from the knowledge that can be found in this treasure cove of wisdom.

REFERENCES

Abbott, A., Sengupta, N., & Kasabov, N. K. (2016, 24-29 Jul). *Which method to use for optimal structure and function representation of large spiking neural networks: A case study on the NeuCube architecture*. Presented at the International Joint Conference on Neural Networks (IJCNN), Vancouver, Canada. https://doi.org/10.1109/IJCNN.2016.7727357

Abdel-Hamid, O., Mohamed, A.-r., Jiang, H., Deng, J., Penn, G., & Yu, D. (2014). Convolutional Neural Networks for Speech Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22*(10), 1533-1545. https://doi.org/10.1109/TASLP.2014.2339736

Abdelaziz, A. H. (2018). Comparing Fusion Models for DNN-Based Audiovisual Continuous Speech Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26*(3), 475-484. https://doi.org/10.1109/TASLP.2017.2783545

Abdollahi, M., & Liu, S.-C. (2011, 10-12 Nov). *Speaker-independent isolated digit recognition using an AER silicon cochlea*. Presented at the IEEE Biomedical Circuits and Systems Conference (BioCAS), San Diego, CA. https://doi.org/10.1109/BioCAS.2011.6107779

Afify, M., Cui, X., & Gao, Y. (2009). Stereo-Based Stochastic Mapping for Robust Speech Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 17*(7), 1325-1334. https://doi.org/10.1109/TASL.2009.2018017

Akbas, E., & Eckstein, M. P. (2017). Object detection through search with a foveated visual system. *PLOS Computational Biology, 13*(10), e1005743. https://doi.org/10.1371/journal.pcbi.1005743

Alais, D., & Burr, D. (2019). Cue Combination Within a Bayesian Framework. In A. K. C. Lee, M. T. Wallace, A. B. Coffin, A. N. Popper, & R. R. Fay (Eds.), *Multisensory Processes: The Auditory Perspective* (pp. 9-31). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-030-10461-0_2

Allen, J. S., Damasio, H., & Grabowski, T. J. (2002). Normal neuroanatomical variation in the human brain: An MRI-volumetric study. *American Journal of Physical Anthropology, 118*(4), 341-358. https://doi.org/10.1002/ajpa.10092

Amunts, K., Armstrong, E., Malikovic, A., Hömke, L., Mohlberg, H., Schleicher, A., & Zilles, K. (2007). Gender-Specific Left-Right Asymmetries in Human Visual Cortex. *The Journal of Neuroscience, 27*(6), 1356-1364. https://doi.org/10.1523/jneurosci.4753-06.2007

Amunts, K., Morosan, P., Hilbig, H., & Zilles, K. (2012). Chapter 36 - Auditory System. In J. K. Mai & G. Paxinos (Eds.), *The Human Nervous System* (3rd ed., pp. 1270-1300). San Diego, CA: Academic Press. https://doi.org/10.1016/B978-0-12-374236-0.10036-7

Andersen, R. A., Snyder, L. H., Batista, A. P., Buneo, C. A., & Cohen, Y. E. (1998). Posterior Parietal Areas Specialized for Eye Movements (Lip) and Reach (PRR) Using a Common Coordinate Frame. In G. R. Bock & J. A. Goode (Eds.), *Novartis Foundation Symposium 218 - Sensory Guidance of Movement* (pp. 109-128). Chichester, United Kingdom: John Wiley & Sons Ltd. https://doi.org/10.1002/9780470515563.ch7

Andreopoulos, A., & Tsotsos, J. K. (2013). 50 Years of object recognition: Directions forward. *Computer Vision and Image Understanding, 117*(8), 827-891. https://doi.org/10.1016/j.cviu.2013.04.005

Andrews, T. J., Halpern, S. D., & Purves, D. (1997). Correlated Size Variations in Human Visual Cortex, Lateral Geniculate Nucleus, and Optic Tract. *The Journal of Neuroscience, 17*(8), 2859-2868. https://doi.org/10.1523/jneurosci.17-08-02859.1997

Anumula, J., Neil, D., Delbrück, T., & Liu, S.-C. (2018). Feature Representations for Neuromorphic Audio Spike Streams. *Frontiers in Neuroscience, 12*(23). https://doi.org/10.3389/fnins.2018.00023

Arandjelovic, R., & Zisserman, A. (2017, 22-29 Oct). *Look, Listen and Learn*. Presented at the IEEE International Conference on Computer Vision (ICCV), Venice, Italy. https://doi.org/10.1109/ICCV.2017.73

Asutay, E., & Västfjäll, D. (2012). Perception of Loudness Is Influenced by Emotion. *PLOS ONE, 7*(6), e38660. https://doi.org/10.1371/journal.pone.0038660

Atilgan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A. K. C., & Bizley, J. K. (2018). Integration of Visual Information in Auditory Cortex Promotes Auditory Scene Analysis through Multisensory Binding. *Neuron, 97*(3), 640-655.e644. https://doi.org/10.1016/j.neuron.2017.12.034

Azevedo, F. A. C., Carvalho, L. R. B., Grinberg, L. T., Farfel, J. M., Ferretti, R. E. L., Leite, R. E. P., . . . Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology, 513*(5), 532-541. https://doi.org/10.1002/cne.21974

Azizian, A., Freitas, A. L., Watson, T. D., & Squires, N. K. (2006). Electrophysiological correlates of categorization: P300 amplitude as index of target similarity. *Biological Psychology, 71*(3), 278-288. https://doi.org/10.1016/j.biopsycho.2005.05.002

Barlow, H. B. (1972). Dark and Light Adaptation: Psychophysics. In D. Jameson & L. M. Hurvich (Eds.), *Visual Psychophysics* (pp. 1-28). Berlin, Germany: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-88658-4_1

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of Auditory and Visual Information about Objects in Superior Temporal Sulcus. *Neuron, 41*(5), 809-823. https://doi.org/10.1016/S0896-6273(04)00070-4

Benson, N. C., Butt, O. H., Brainard, D. H., & Aguirre, G. K. (2014). Correction of Distortion in Flattened Representations of the Cortical Surface Allows Prediction of V1-V3 Functional Organization from Anatomy. *PLOS Computational Biology, 10*(3), e1003538. https://doi.org/10.1371/journal.pcbi.1003538

Benson, N. C., Jamison, K. W., Arcaro, M. J., Vu, A. T., Glasser, M. F., Coalson, T. S., . . . Kay, K. (2018a). *The HCP 7T Retinotopy Dataset*. Retrieved on 30 Sep 2020 from https://osf.io/bw9ec/

Benson, N. C., Jamison, K. W., Arcaro, M. J., Vu, A. T., Glasser, M. F., Coalson, T. S., . . . Kay, K. (2018b). The Human Connectome Project 7 Tesla retinotopy dataset: Description and population receptive field analysis. *Journal of Vision, 18*(13), 23-23. https://doi.org/10.1167/18.13.23

Berg, E. (1996). Forensic Image Processing - An Introduction to Image Enhancement. Retrieved from https://web.archive.org/web/20070206182200/http://www.imagingforensics.com/forensic.pdf

Bergmann, J., Genç, E., Kohler, A., Singer, W., & Pearson, J. (2014). Neural Anatomy of Primary Visual Cortex Limits Visual Working Memory. *Cerebral Cortex, 26*(1), 43-50. https://doi.org/10.1093/cercor/bhu168

Bergua, A. (2017). Retina. In A. Bergua (Ed.), *Das menschliche Auge in Zahlen* (pp. 99-103). Berlin, Germany: Springer. https://doi.org/10.1007/978-3-662-47284-2_19

Berner, R., & Delbrück, T. (2011). Event-Based Pixel Sensitive to Changes of Color and Brightness. *IEEE Transactions on Circuits and Systems I: Regular Papers, 58*(7), 1581-1590. https://doi.org/10.1109/TCSI.2011.2157770

Beyeler, M., Dutt, N. D., & Krichmar, J. L. (2013). Categorization and decision-making in a neurobiologically plausible spiking network using a STDP-like learning rule. *Neural Networks, 48*, 109-124. https://doi.org/10.1016/j.neunet.2013.07.012

Boddapati, V., Petef, A., Rasmusson, J., & Lundberg, L. (2017). Classifying environmental sounds using image recognition networks. *Procedia Computer Science, 112*, 2048-2056. https://doi.org/10.1016/j.procs.2017.08.250

Bradley, A., Applegate, R. A., Zeffren, B. S., & van Heuven, W. A. J. (1992). Psychophysical measurement of the size and shape of the human foveal avascular zone. *Ophthalmic and Physiological Optics, 12*(1), 18-23. https://doi.org/10.1111/j.1475-1313.1992.tb00248.x

Bradski, G. (2000). The OpenCV Library. Retrieved from https://opencv.org/

Bragg, D., Koller, O., Bellard, M., Berke, L., Boudreault, P., Braffort, A., . . . Ringel Morris, M. (2019, 28-30 Oct). *Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective*. Presented at the International ACM SIGACCESS Conference on Computers and Accessibility, Pittsburgh, PA. https://doi.org/10.1145/3308561.3353774

Brashear, H., Starner, T., Lukowicz, P., & Junker, H. (2003, 21-23 Oct). *Using Multiple Sensors for Mobile Sign Language Recognition*. Presented at the IEEE International Symposium on Wearable Computers (ISWC), White Plains, NY. Retrieved from http://hdl.handle.net/1853/28997

Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience, 3*(3), 243-249. https://doi.org/10.1038/nrn756

Brewer, A. A., & Barton, B. (2016). Maps of the Auditory Cortex. *Annual Review of Neuroscience, 39*(1), 385-407. https://doi.org/10.1146/annurev-neuro-070815-014045

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues* [Comparative localisation studies of the cerebral cortex in its principles described on the basis of cytoarchitecture]. Leipzig, Germany: Johann Ambrosius Barth. Retrieved from https://archive.org/details/b28062449

Brown, P. K., & Wald, G. (1964). Visual Pigments in Single Rods and Cones of the Human Retina. *Science, 144*(3614), 45-52. https://doi.org/10.1126/science.144.3614.45

Brownlee, J. (2020). *What is the Difference Between Test and Validation Datasets?* Retrieved on 24 Nov 2020 from https://machinelearningmastery.com/difference-test-validation-datasets/

Bruce, V., Georgeson, M. A., & Green, P. R. (2003). *Visual Perception* (4th ed.). Hove, Canada: Psychology Press. https://doi.org/10.4324/9780203427248

Brugge, J. F., & Geisler, C. D. (1978). Auditory Mechanisms of the Lower Brainstem. *Annual Review of Neuroscience, 1*(1), 363-394. https://doi.org/10.1146/annurev.ne.01.030178.002051

Brunel, N., & van Rossum, M. C. W. (2007). Quantitative investigations of electrical nerve excitation treated as polarization. *Biological Cybernetics, 97*(5), 341-349. https://doi.org/10.1007/s00422-007-0189-6

Bruns, P., & Röder, B. (2019). Cross-Modal Learning in the Auditory System. In A. K. C. Lee, M. T. Wallace, A. B. Coffin, A. N. Popper, & R. R. Fay (Eds.), *Multisensory Processes: The Auditory Perspective* (pp. 221-242). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-030-10461-0_11

Buchweitz, A., Shinkareva, S. V., Mason, R. A., Mitchell, T. M., & Just, M. A. (2012). Identifying bilingual semantic neural representations across languages. *Brain and Language, 120*(3), 282-289. https://doi.org/10.1016/j.bandl.2011.09.003

Burr, D., & Alais, D. (2006). Combining visual and auditory information. In S. Martinez-Conde, S. L. Macknik, L. M. Martinez, J. M. Alonso, & P. U. Tse (Eds.), *Progress in Brain Research* (Vol. 155, pp. 243-258). Amsterdam, The Netherlands: Elsevier. https://doi.org/10.1016/S0079-6123(06)55014-9

Bush, E. C., & Allman, J. M. (2004). Three-dimensional structure and evolution of primate primary visual cortex. *The Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology, 281A*(1), 1088-1094. https://doi.org/10.1002/ar.a.20114

Çakır, E., Adavanne, S., Parascandolo, G., Drossos, K., & Virtanen, T. (2017, 28 Aug-2 Sep). *Convolutional recurrent neural networks for bird audio detection*. Presented at the European Signal Processing Conference (EUSIPCO), Kos, Greece. https://doi.org/10.23919/EUSIPCO.2017.8081508

Çakır, E., Parascandolo, G., Heittola, T., Huttunen, H., & Virtanen, T. (2017). Convolutional Recurrent Neural Networks for Polyphonic Sound Event Detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25*(6), 1291-1303. https://doi.org/10.1109/TASLP.2017.2690575

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology, 10*(11), 649-657. https://doi.org/10.1016/S0960-9822(00)00513-3

Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine, 9*(2), 48-57. https://doi.org/10.1109/MCI.2014.2307227

Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1493), 1001-1010. https://doi.org/10.1098/rstb.2007.2155

Caridakis, G., Asteriadis, S., & Karpouzis, K. (2014). Non-manual cues in automatic sign language recognition. *Personal and Ubiquitous Computing, 18*(1), 37-46. https://doi.org/10.1007/s00779-012-0615-1

Caron-Desrochers, L., Schönwiesner, M., Focke, K., & Lehmann, A. (2018). Assessing visual modulation along the human subcortical auditory pathway. *Neuroscience Letters, 685*, 12-17. https://doi.org/10.1016/j.neulet.2018.07.020

Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences, 18*(2), 90-98. https://doi.org/10.1016/j.tics.2013.11.005

Cassidy, A. S., Alvarez-Icaza, R., Akopyan, F., Sawada, J., Arthur, J. V., Merolla, P. A., . . . Modha, D. S. (2014, 16-21 Nov). *Real-Time Scalable Cortical Computing at 46 Giga-Synaptic OPS/Watt with ~100× Speedup in Time-to-Solution and ~100,000× Reduction in Energy-to-Solution*. Presented at the International Conference for High Performance Computing, Networking, Storage and Analysis, New Orleans, LA. https://doi.org/10.1109/SC.2014.8

Cerezuela-Escudero, E., Jiménez-Fernández, Á., Paz-Vicente, R., Dominguez-Morales, J. P., Domínguez-Morales, M. J., & Linares-Barranco, A. (2016, 6-9 Sep). *Sound Recognition System Using Spiking and MLP Neural Networks*. Presented at the International Conference on Artificial Neural Networks (ICANN), Barcelona, Spain. https://doi.org/10.1007/978-3-319-44781-0_43

Čertický, M., Churchill, D., Kim, K.-J., Čertický, M., & Kelly, R. L. (2019). StarCraft AI Competitions, Bots, and Tournament Manager Software. *IEEE Transactions on Games, 11*(3), 227-237. https://doi.org/10.1109/TG.2018.2883499

Cessac, B., Kornprobst, P., Kraria, S., Nasser, H., Pamplona, D., Portelli, G., & Viéville, T. (2017). PRANAS: A New Platform for Retinal Analysis and Simulation. *Frontiers in Neuroinformatics, 11*(49). https://doi.org/10.3389/fninf.2017.00049

Chan, A. M., Halgren, E., Marinkovic, K., & Cash, S. S. (2011). Decoding word and category-specific spatiotemporal representations from MEG and EEG. *NeuroImage, 54*(4), 3028-3039. https://doi.org/10.1016/j.neuroimage.2010.10.073

Chaplin, T. A., Rosa, M. G. P., & Lui, L. L. (2018). Auditory and Visual Motion Processing and Integration in the Primate Cerebral Cortex. *Frontiers in Neural Circuits, 12*(93). https://doi.org/10.3389/fncir.2018.00093

Chaquet, J. M., Carmona, E. J., & Fernández-Caballero, A. (2013). A survey of video datasets for human action and activity recognition. *Computer Vision and Image Understanding, 117*(6), 633-659. https://doi.org/10.1016/j.cviu.2013.01.013

Chevreul, M. E. (1855). *The principles of harmony and contrast of colours, and their applications to the arts* (2nd ed.). Retrieved from https://books.google.co.nz/books?id=l91OAAAAMAAJ

Choi, K., Fazekas, G., Sandler, M. B., & Cho, K. (2017, 5-9 Mar). *Convolutional recurrent neural networks for music classification*. Presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA. https://doi.org/10.1109/ICASSP.2017.7952585

Clark, N. R., Brown, G. J., Jürgens, T., & Meddis, R. (2012). A frequency-selective feedback model of auditory efferent suppression and its implications for the recognition of speech in noise. *The Journal of the Acoustical Society of America, 132*(3), 1535-1541. https://doi.org/10.1121/1.4742745

Correia, J. M., Formisano, E., Valente, G., Hausfeld, L., Jansma, B., & Bonte, M. (2014). Brain-based translation: fMRI decoding of spoken words in bilinguals reveals language-independent semantic representations in anterior temporal lobe. *The Journal of Neuroscience, 34*(1), 332-338. https://doi.org/10.1523/jneurosci.1302-13.2014

Correia, J. M., Jansma, B., Hausfeld, L., Kikkert, S., & Bonte, M. (2015). EEG decoding of spoken words in bilingual listeners: from words to language invariant semantic-conceptual representations. *Frontiers in Psychology, 6*, 71. https://doi.org/10.3389/fpsyg.2015.00071

Costa, Y. M. G., Oliveira, L. S., & Silla, C. N. (2017). An evaluation of Convolutional Neural Networks for music classification using spectrograms. *Applied Soft Computing, 52*, 28-38. https://doi.org/10.1016/j.asoc.2016.12.024

Cote, R. H. (2006). Photoreceptor Phosphodiesterase (PDE6): A G-Protein-Activated PDE Regulating Visual Excitation in Rod and Cone Photoreceptor Cells. In J. A. Beavo, S. H. Francis, & M. D. Houslay (Eds.), *Cyclic Nucleotide Phosphodiesterases in Health and Disease*. Boca Raton, FL: CRC Press. https://doi.org/10.1201/9781420020847-8

Cruz, F., Parisi, G. I., Twiefel, J., & Wermter, S. (2016, 9-14 Oct). *Multi-modal integration of dynamic audiovisual patterns for an interactive reinforcement learning scenario*. Presented at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, South Korea. https://doi.org/10.1109/IROS.2016.7759137

Cui, G., Luo, M. R., Rigg, B., Roesler, G., & Witt, K. (2002). Uniform colour spaces based on the DIN99 colour-difference formula. *Color Research & Application, 27*(4), 282-290. https://doi.org/10.1002/col.10066

Cui, X. (n.d.). *Documentation | xjView*. Retrieved on 7 Oct 2020 from https://www.alivelearn.net/xjview/documentation/

Curcio, C. A., & Allen, K. A. (1990). Topography of ganglion cells in human retina. *Journal of Comparative Neurology, 300*(1), 5-25. https://doi.org/10.1002/cne.903000103

Da Costa, S., van der Zwaag, W., Marques, J. P., Frackowiak, R. S. J., Clarke, S., & Saenz, M. (2011). Human Primary Auditory Cortex Follows the Shape of Heschl's Gyrus. *The Journal of Neuroscience, 31*(40), 14067-14075. https://doi.org/10.1523/jneurosci.2000-11.2011

Dahmen, J. C., & King, A. J. (2007). Learning to hear: plasticity of auditory cortical processing. *Current Opinion in Neurobiology, 17*(4), 456-464. https://doi.org/10.1016/j.conb.2007.07.004

Dambrosio, F., Amy, D., & Colombo, A. (1995). B-mode color sonographic images in obstetrics and gynecology: preliminary report. *Ultrasound in Obstetrics & Gynecology, 6*(3), 208-215. https://doi.org/10.1046/j.1469-0705.1995.06030208.x

Davis, K. H., Biddulph, R., & Balashek, S. (1952). Automatic Recognition of Spoken Digits. *The Journal of the Acoustical Society of America, 24*(6), 637-642. https://doi.org/10.1121/1.1906946

de Sousa, A. A., Sherwood, C. C., Mohlberg, H., Amunts, K., Schleicher, A., MacLeod, C. E., . . . Zilles, K. (2010). Hominoid visual brain structure volumes and the position of the lunate sulcus. *Journal of Human Evolution, 58*(4), 281-292. https://doi.org/10.1016/j.jhevol.2009.11.011

de Villers-Sidani, E., Chang, E. F., Bao, S., & Merzenich, M. M. (2007). Critical Period Window for Spectral Tuning Defined in the Primary Auditory Cortex (A1) in the Rat. *The Journal of Neuroscience, 27*(1), 180-189. https://doi.org/10.1523/jneurosci.3227-06.2007

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences, 15*(6), 254-262. https://doi.org/10.1016/j.tics.2011.04.003

Delgutte, B. (1997). Auditory neural processing of speech. In W. J. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (pp. 507-538). Oxford, United Kingdom: Blackwell. Retrieved from http://research.meei.harvard.edu/NeuralCoding/Papers/HandbookChapter.pdf

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Li, F.-F. (2009, 20-25 Jun). *ImageNet: A large-scale hierarchical image database*. Presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL. https://doi.org/10.1109/CVPR.2009.5206848

Deng, L., Hinton, G. E., & Kingsbury, B. (2013, 26-31 May). *New types of deep neural network learning for speech recognition and related applications: an overview*. Presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, Canada. https://doi.org/10.1109/ICASSP.2013.6639344

Deniz, F., Nunez-Elizalde, A. O., Huth, A. G., & Gallant, J. L. (2019). The Representation of Semantic Information Across Human Cerebral Cortex During Listening Versus Reading Is Invariant to Stimulus Modality. *The Journal of Neuroscience, 39*(39), 7722-7736. https://doi.org/10.1523/jneurosci.0675-19.2019

DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences, 109*(8), E505-E514. https://doi.org/10.1073/pnas.1113427109

DeYoe, E. A., Ulmer, J. L., Mueller, W., Hacein-Bey, L., Szeder, V., Maciejewski, M. J., . . . Mathis, J. (2012). fMRI of Human Visual Pathways. In S. H. Faro, F. B. Mohamed, M. Law, & J. T. Ulmer (Eds.), *Functional Neuroradiology: Principles and Clinical Applications* (pp. 485-511). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4419-0345-7_26

Dibazar, A. A., Song, D., Yamada, W., & Berger, T. W. (2004, 25-29 Jul). *Speech recognition based on fundamental functional principles of the brain*. Presented at the International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary. https://doi.org/10.1109/IJCNN.2004.4620175

Doborjeh, M. G., Wang, G. Y., Kasabov, N. K., Kydd, R., & Russell, B. (2016). A Spiking Neural Network Methodology and System for Learning and Comparative Analysis of EEG Data From Healthy Versus Addiction Treated Versus Addiction Not Treated Subjects. *IEEE Transactions on Biomedical Engineering, 63*(9), 1830-1841. https://doi.org/10.1109/TBME.2015.2503400

Doddington, G. R., & Schalk, T. B. (1981). Computers: Speech recognition: Turning theory to practice: New ICs have brought the requisite computer power to speech technology; an evaluation of equipment shows where it stands today. *IEEE Spectrum, 18*(9), 26-32. https://doi.org/10.1109/MSPEC.1981.6369809

Dominguez-Morales, J. P., Jiménez-Fernández, Á., Domínguez-Morales, M. J., & Jiménez-Moreno, G. (2018). Deep Neural Networks for the Recognition and Classification of Heart Murmurs Using Neuromorphic Auditory Sensors. *IEEE Transactions on Biomedical Circuits and Systems, 12*(1), 24-34. https://doi.org/10.1109/TBCAS.2017.2751545

Dominguez-Morales, J. P., Jiménez-Fernández, Á., Rios-Navarro, A., Cerezuela-Escudero, E., Gutierrez-Galan, D., Domínguez-Morales, M. J., & Jiménez-Moreno, G. (2016, 6-9 Sep). *Multilayer Spiking Neural Network for Audio Samples Classification Using SpiNNaker*. Presented at the International Conference on Artificial Neural Networks (ICANN), Barcelona, Spain. https://doi.org/10.1007/978-3-319-44778-0_6

Dominguez-Morales, J. P., Liu, Q., James, R., Gutierrez-Galan, D., Jiménez-Fernández, Á., Davidson, S., & Furber, S. B. (2018, 8-13 Jul). *Deep Spiking Neural Network model for time-variant signals classification: a real-time speech recognition approach*. Presented at the International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil. https://doi.org/10.1109/IJCNN.2018.8489381

Domínguez-Morales, M. J., Dominguez-Morales, J. P., Ríos-Navarro, A., Cascado-Caballero, D., Jiménez-Fernández, Á., & Linares-Barranco, A. (2019). Neuronal Specialization for Fine-Grained Distance Estimation Using a Real-Time Bio-Inspired Stereo Vision System. *Electronics, 8*(12), 1502. https://doi.org/10.3390/electronics8121502

Dong, M., Huang, X., & Xu, B. (2018). Unsupervised speech recognition through spike-timing-dependent plasticity in a convolutional spiking neural network. *PLOS ONE, 13*(11), e0204596. https://doi.org/10.1371/journal.pone.0204596

Dörfler, M., Bammer, R., & Grill, T. (2017, 3-7 Jul). *Inside the spectrogram: Convolutional Neural Networks in audio processing*. Presented at the International Conference on Sampling Theory and Applications (SampTA), Tallinn, Estonia. https://doi.org/10.1109/SAMPTA.2017.8024472

Dorph-Petersen, K.-A., Pierri, J. N., Wu, Q., Sampson, A. R., & Lewis, D. A. (2007). Primary visual cortex volume and total neuron number are reduced in schizophrenia. *Journal of Comparative Neurology, 501*(2), 290-301. https://doi.org/10.1002/cne.21243

Doucet, M. E., Bergeron, F., Lassonde, M., Ferron, P., & Lepore, F. (2006). Cross-modal reorganization and speech perception in cochlear implant users. *Brain, 129*(12), 3376-3383. https://doi.org/10.1093/brain/awl264

Dougherty, R. F., Koch, V. M., Brewer, A. A., Fischer, B., Modersitzki, J., & Wandell, B. A. (2003). Visual field representations and locations of visual areas V1/2/3 in human visual cortex. *Journal of Vision, 3*(10), 1-1. https://doi.org/10.1167/3.10.1

Douglas, R. D., & Brewer, J. D. (1998). Acceptability of shade differences in metal ceramic crowns. *The Journal of Prosthetic Dentistry, 79*(3), 254-260. https://doi.org/10.1016/S0022-3913(98)70233-1

Drake, R. L., Vogl, W., & Mitchell, A. W. M. (2020). *Gray's Anatomy for Students* (4th ed.). Retrieved from https://www.clinicalkey.com.au/dura/browse/bookChapter/3-s2.0-C20150000041

Dupont, S., & Luettin, J. (2000). Audio-visual speech modeling for continuous speech recognition. *IEEE Transactions on Multimedia, 2*(3), 141-151. https://doi.org/10.1109/6046.865479

Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E.-J., & Shadlen, M. N. (1994). fMRI of human visual cortex. *Nature, 369*(6481), 525-525. https://doi.org/10.1038/369525a0

Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., & Peters, T. M. (1993, 31 Oct-6 Nov). *3D statistical neuroanatomical models from 305 MRI volumes*. Presented at the IEEE Nuclear Science Symposium and Medical Imaging Conference, San Francisco, CA. https://doi.org/10.1109/NSSMIC.1993.373602

Ferrucci, D. (2010, 11-15 Sep). *Build Watson: An Overview of DeepQA for the Jeopardy! Challenge*. Presented at the International Conference on Parallel Architectures and Compilation Techniques (PACT), Vienna, Austria. Retrieved from https://ieeexplore.ieee.org/document/7851501

Fetterly, K. A., Blume, H. R., Flynn, M. J., & Samei, E. (2008). Introduction to Grayscale Calibration and Related Aspects of Medical Imaging Grade Liquid Crystal Displays.

Journal of Digital Imaging, 21(2), 193-207. https://doi.org/10.1007/s10278-007-9022-y
Finney, E. M., Fine, I., & Dobkins, K. R. (2001). Visual stimuli activate auditory cortex in the deaf. *Nature Neuroscience, 4*(12), 1171-1173. https://doi.org/10.1038/nn763
Fisher, R., Santos-Victor, J., & Crowley, J. (2005). *CAVIAR: Context Aware Vision using Image-based Active Recognition*. Retrieved on 10 Jun 2020 from http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/
Flechsig, P. (1905). Hirnphysiologie und Willenstheorien [Brain physiology and mind theories]. *Annalen der Naturphilosophie, 4*(1), 475-498. Retrieved from https://zs.thulb.uni-jena.de/receive/jportal_jparticle_00281474
Flechsig, P. (1908). Bemerkungen über die Hörsphäre des menschlichen Gehirns [Remarks on the hearing sphere of the human brain]. *Neurologisches Centralblatt, 27*(1-2), 1-7, 50-57. Retrieved from https://archive.org/details/neurologischesc08unkngoog
Flores, Emma N., Duggan, A., Madathany, T., Hogan, Ann K., Márquez, Freddie G., Kumar, G., . . . García-Añoveros, J. (2015). A Non-canonical Pathway from Cochlea to Brain Signals Tissue-Damaging Noise. *Current Biology, 25*(5), 606-612. https://doi.org/10.1016/j.cub.2015.01.009
Fong, R. C., Scheirer, W. J., & Cox, D. D. (2018). Using human brain activity to guide machine learning. *Scientific Reports, 8*(1), 5397. https://doi.org/10.1038/s41598-018-23618-6
Fontaine, B., Goodman, D. F. M., Benichoux, V., & Brette, R. (2011). Brian Hears: Online Auditory Processing Using Vectorization Over Channels. *Frontiers in Neuroinformatics, 5*(9). https://doi.org/10.3389/fninf.2011.00009
Freeman, L. C. A., Wood, K. C., & Bizley, J. K. (2018). Multisensory stimuli improve relative localisation judgments compared to unisensory auditory or visual stimuli. *The Journal of the Acoustical Society of America, 143*(6), EL516-EL522. https://doi.org/10.1121/1.5042759
Fukui, Y. (2001). Human Eyes as an Image Sensor. Retrieved from https://web.archive.org/web/20050522102952/http://faculty-web.at.northwestern.edu/med/fukui/Human%20eye.pdf
Furber, S. B., Galluppi, F., Temple, S., & Plana, L. A. (2014). The SpiNNaker Project. *Proceedings of the IEEE, 102*(5), 652-665. https://doi.org/10.1109/JPROC.2014.2304638
Gagin, G., Bohon, K. S., Butensky, A., Gates, M. A., Hu, J.-Y., Lafer-Sousa, R., . . . Conway, B. R. (2014). Color-detection thresholds in rhesus macaque monkeys and humans. *Journal of Vision, 14*(8), 12. https://doi.org/10.1167/14.8.12
Gales, M. J. F. (1998). Maximum likelihood linear transformations for HMM-based speech recognition. *Computer Speech & Language, 12*(2), 75-98. https://doi.org/10.1006/csla.1998.0043
Gales, M. J. F., & Young, S. (2008). The Application of Hidden Markov Models in Speech Recognition. *Foundations and Trends® in Signal Processing, 1*(3), 195-304. https://doi.org/10.1561/2000000004
Ganapathiraju, A., Hamaker, J. E., & Picone, J. (2004). Applications of support vector machines to speech recognition. *IEEE Transactions on Signal Processing, 52*(8), 2348-2355. https://doi.org/10.1109/TSP.2004.831018
Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., Dahlgren, N. L., & Zue, V. (1993). *TIMIT Acoustic-Phonetic Continuous Speech Corpus*. Philadelphia, PA. Retrieved on 26th Oct 2015 from https://catalog.ldc.upenn.edu/LDC93S1
Ge, L., Liang, H., Yuan, J., & Thalmann, D. (2019). Real-Time 3D Hand Pose Estimation with 3D Convolutional Neural Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(4), 956-970. https://doi.org/10.1109/TPAMI.2018.2827052
Gee, A. H., Garcia-Olano, D., Ghosh, J., & Paydarfar, D. (2019, 10 Aug). *Explaining Deep Classification of Time-Series Data with Learned Prototypes*. Presented at the International

Workshop on Knowledge Discovery in Healthcare Data, Macao, China. Retrieved from http://ceur-ws.org/Vol-2429/paper3.pdf

Gegenfurtner, K. R. (2003). Cortical mechanisms of colour vision. *Nature Reviews Neuroscience, 4*(7), 563-572. https://doi.org/10.1038/nrn1138

George, A. M., Banerjee, D., Dey, S., Mukherjee, A., & Balamurali, P. (2020, 19-24 Jul). *A Reservoir-based Convolutional Spiking Neural Network for Gesture Recognition from DVS Input*. Presented at the International Joint Conference on Neural Networks (IJCNN), Glasgow, United Kingdom. https://doi.org/10.1109/IJCNN48605.2020.9206681

Gerl, E. J., & Morris, M. R. (2008). The Causes and Consequences of Color Vision. *Evolution: Education and Outreach, 1*(4), 476-486. https://doi.org/10.1007/s12052-008-0088-x

Gilbert, C. D. (2013). The Constructive Nature of Visual Processing. In E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, A. J. Hudspeth, & S. Mack (Eds.), *Principles of Neural Science* (5th ed., pp. 556-576). New York, NY: McGraw-Hill.

Glackin, B., Wall, J. A., McGinnity, T. M., Maguire, L. P., & McDaid, L. J. (2010). A Spiking Neural Network Model of the Medial Superior Olive Using Spike Timing Dependent Plasticity for Sound Localization. *Frontiers in Computational Neuroscience, 4*(18). https://doi.org/10.3389/fncom.2010.00018

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., . . . Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage, 80*, 105-124. https://doi.org/10.1016/j.neuroimage.2013.04.127

Glazebrook, C. M., Welsh, T. N., & Tremblay, L. (2016). The processing of visual and auditory information for reaching movements. *Psychological Research, 80*(5), 757-773. https://doi.org/10.1007/s00426-015-0689-2

Goebel, R., Muckli, L., & Kim, D.-S. (2012). Chapter 37 - Visual System. In J. K. Mai & G. Paxinos (Eds.), *The Human Nervous System* (3rd ed., pp. 1301-1327). San Diego, CA: Academic Press. https://doi.org/10.1016/B978-0-12-374236-0.10037-9

Goertzel, B., Lian, R., Arel, I., de Garis, H., & Chen, S. (2010). A world survey of artificial brain projects, Part II: Biologically inspired cognitive architectures. *Neurocomputing, 74*(1), 30-49. https://doi.org/10.1016/j.neucom.2010.08.012

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15*(1), 20-25. https://doi.org/10.1016/0166-2236(92)90344-8

Goodman, D. F. M., & Brette, R. (2008). Brian: a simulator for spiking neural networks in Python. *Frontiers in Neuroinformatics, 2*(5). https://doi.org/10.3389/neuro.11.005.2008

Goodman, D. F. M., & Brette, R. (2010). Spike-Timing-Based Computation in Sound Localization. *PLOS Computational Biology, 6*(11), e1000993. https://doi.org/10.1371/journal.pcbi.1000993

Gorelick, L., Blank, M., Shechtman, E., Irani, M., & Basri, R. (2007). Actions as Space-Time Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(12), 2247-2253. https://doi.org/10.1109/TPAMI.2007.70711

Goutman, J. D., Elgoyhen, A. B., & Gómez-Casati, M. E. (2015). Cochlear hair cells: The sound-sensing machines. *FEBS Letters, 589*(22), 3354-3361. https://doi.org/10.1016/j.febslet.2015.08.030

Graves, A., Eck, D., Beringer, N., & Schmidhuber, J. (2004, 29-30 Jan). *Biologically Plausible Speech Recognition with LSTM Neural Nets*. Presented at the First International Workshop on Biologically Inspired Approaches to Advanced Information Technology, Lausanne, Switzerland. https://doi.org/10.1007/978-3-540-27835-1_10

Graves, A., Mohamed, A.-r., & Hinton, G. E. (2013, 26-31 May). *Speech recognition with deep recurrent neural networks*. Presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, Canada. https://doi.org/10.1109/ICASSP.2013.6638947

Greenwood, D. D. (1961). Critical Bandwidth and the Frequency Coordinates of the Basilar Membrane. *The Journal of the Acoustical Society of America, 33*(10), 1344-1356. https://doi.org/10.1121/1.1908437

Greenwood, D. D. (1990). A cochlear frequency‐position function for several species—29 years later. *The Journal of the Acoustical Society of America, 87*(6), 2592-2605. https://doi.org/10.1121/1.399052

Griffiths, T. D., Micheyl, C., & Overath, T. (2012). Auditory Object Analysis. In D. Poeppel, T. Overath, A. N. Popper, & R. R. Fay (Eds.), *The Human Auditory Cortex* (pp. 199-223). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-2314-0_8

Groh, J. M., Trause, A. S., Underhill, A. M., Clark, K. R., & Inati, S. (2001). Eye Position Influences Auditory Responses in Primate Inferior Colliculus. *Neuron, 29*(2), 509-518. https://doi.org/10.1016/S0896-6273(01)00222-7

Grothe, B., Pecka, M., & McAlpine, D. (2010). Mechanisms of Sound Localization in Mammals. *Physiological Reviews, 90*(3), 983-1012. https://doi.org/10.1152/physrev.00026.2009

Grünert, U., & Martin, P. R. (2020). Cell types and cell circuits in human and non-human primate retina. *Progress in Retinal and Eye Research*, 100844. https://doi.org/10.1016/j.preteyeres.2020.100844

Gruters, K., & Groh, J. M. (2012). Sounds and beyond: multisensory and other non-auditory signals in the inferior colliculus. *Frontiers in Neural Circuits, 6*(96). https://doi.org/10.3389/fncir.2012.00096

Gütig, R., Gollisch, T., Sompolinsky, H., & Meister, M. (2013). Computing Complex Visual Features with Retinal Spike Times. *PLOS ONE, 8*(1), e53063. https://doi.org/10.1371/journal.pone.0053063

Guyon, I., Athitsos, V., Jangyodsuk, P., & Escalante, H. J. (2014). The ChaLearn gesture dataset (CGD 2011). *Machine Vision and Applications, 25*(8), 1929-1951. https://doi.org/10.1007/s00138-014-0596-3

Habekost, M. (2013). Which color differencing equation should be used? *International Circular of Graphic Education and Research*(6), 20-33. Retrieved from https://www.internationalcircle.net/international_circle/international_circle/international_circle/circular/issues/13_01/ICJ_06_2013_02_069.pdf

Hadjiivanov, A. (2016, 24-29 Jul). *Adaptive conversion of real-valued input into spike trains*. Presented at the International Joint Conference on Neural Networks (IJCNN), Vancouver, Canada. https://doi.org/10.1109/IJCNN.2016.7727314

Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1493), 1055-1069. https://doi.org/10.1098/rstb.2007.2159

Hanson, A. R. (2012). 1 - What is colour? In J. Best (Ed.), *Colour Design (Second Edition)* (pp. 3-21). Duxford, United Kingdom: Woodhead Publishing. https://doi.org/10.1016/B978-0-08-101270-3.00001-1

Hanson, S. J., Matsuka, T., & Haxby, J. V. (2004). Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a "face" area? *NeuroImage, 23*(1), 156-166. https://doi.org/10.1016/j.neuroimage.2004.05.020

Harris, C. R., Millman, J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., . . . Oliphant, T. E. (2020). Array programming with NumPy. *Nature, 585*(7825), 357-362. https://doi.org/10.1038/s41586-020-2649-2

Harshita, P., & Adiga, A. R. (2018, 3-5 Apr). *Speech Recognition with Frequency Domain Linear Prediction*. Presented at the IEEE International Conference on Communication and Signal Processing (ICCSP), Chennai, India. https://doi.org/10.1109/ICCSP.2018.8524270

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-Inspired Artificial Intelligence. *Neuron, 95*(2), 245-258. https://doi.org/10.1016/j.neuron.2017.06.011

Hassan, M., Assaleh, K., & Shanableh, T. (2019). Multiple Proposals for Continuous Arabic Sign Language Recognition. *Sensing and Imaging, 20*(1), 4. https://doi.org/10.1007/s11220-019-0225-3

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science, 293*(5539), 2425-2430. https://doi.org/10.1126/science.1063736

He, K., Zhang, X., Ren, S., & Sun, J. (2016, 27-30 Jun). *Deep Residual Learning for Image Recognition*. Presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV. https://doi.org/10.1109/CVPR.2016.90

Heil, P., & Peterson, A. J. (2015). Basic response properties of auditory nerve fibers: a review. *Cell and Tissue Research, 361*(1), 129-158. https://doi.org/10.1007/s00441-015-2177-9

Hendry, S. H., Schwark, H. D., Jones, E. G., & Yan, J. (1987). Numbers and proportions of GABA-immunoreactive neurons in different areas of monkey cerebral cortex. *The Journal of Neuroscience, 7*(5), 1503-1519. https://doi.org/10.1523/jneurosci.07-05-01503.1987

Henriksson, L., Karvonen, J., Salminen-Vaparanta, N., Railo, H., & Vanni, S. (2012). Retinotopic Maps, Spatial Tuning, and Locations of Human Visual Areas in Surface Coordinates Characterized with Multifocal and Blocked fMRI Designs. *PLOS ONE, 7*(5), e36859. https://doi.org/10.1371/journal.pone.0036859

Higgins, I., Stringer, S., & Schnupp, J. W. H. (2017). Unsupervised learning of temporal features for word categorization in a spiking neural network model of the auditory brain. *PLOS ONE, 12*(8), e0180174. https://doi.org/10.1371/journal.pone.0180174

Higgins, I., Stringer, S., & Schnupp, J. W. H. (2018). A Computational Account of the Role of Cochlear Nucleus and Inferior Colliculus in Stabilizing Auditory Nerve Firing for Auditory Category Learning. *Neural Computation, 30*(7), 1801-1829. https://doi.org/10.1162/neco_a_01085

Hill, J. (2020). Do deaf communities actually want sign language gloves? *Nature Electronics, 3*(9), 512-513. https://doi.org/10.1038/s41928-020-0451-7

Hinton, G. E. (2009). Deep belief networks. *Scholarpedia, 4*(5), 5947. https://doi.org/10.4249/scholarpedia.5947

Hinton, G. E., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., . . . Kingsbury, B. (2012). Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Processing Magazine, 29*(6), 82-97. https://doi.org/10.1109/MSP.2012.2205597

Hirsch, H.-G., & Pearce, D. (2000, 18-20 Sep). *The AURORA Experimental Framework For The Performance Evaluation of Speech Recognition Systems Under Noisy Conditions*. Presented at the Automatic Speech Recognition: Challenges for the new Millenium, Paris, France. Retrieved from http://dnt.kr.hsnr.de/aurora/download/asr2000_final_footer.pdf

Holmberg, M., Gelbart, D., & Hemmert, W. (2007). Speech encoding in a model of peripheral auditory processing: Quantitative assessment by means of automatic speech recognition. *Speech Communication, 49*(12), 917-932. https://doi.org/10.1016/j.specom.2007.05.009

Holmes, G. (1918). Disturbances of Vision by Cerebral Lesions. *British Journal of Ophthalmology, 2*(7), 353-384. https://doi.org/10.1136/bjo.2.7.353

Hong, C., Wei, X., Wang, J., Deng, B., Yu, H., & Che, Y. (2020). Training Spiking Neural Networks for Cognitive Tasks: A Versatile Framework Compatible With Various Temporal Codes. *IEEE Transactions on Neural Networks and Learning Systems, 31*(4), 1285-1296. https://doi.org/10.1109/TNNLS.2019.2919662

Hoon, M., Okawa, H., Della Santina, L., & Wong, R. O. L. (2014). Functional architecture of the retina: Development and disease. *Progress in Retinal and Eye Research, 42*, 44-84. https://doi.org/10.1016/j.preteyeres.2014.06.003

Hopkins, M., Pineda-García, G., Bogdan, P. A., & Furber, S. B. (2018). Spiking neural networks for computer vision. *Interface Focus, 8*(4), 20180007. https://doi.org/10.1098/rsfs.2018.0007

Houlihan, M., Campbell, K., & Stelmack, R. M. (1994). Reaction time and movement time as measures of stimulus evaluation and response processes. *Intelligence, 18*(3), 289-307. https://doi.org/10.1016/0160-2896(94)90031-0

Huang, J., & Kingsbury, B. (2013, 26-31 May). *Audio-visual deep learning for noise robust speech recognition*. Presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, Canada. https://doi.org/10.1109/ICASSP.2013.6639140

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature, 532*(7600), 453-458. https://doi.org/10.1038/nature17637

Iakymchuk, T., Rosado-Muñoz, A., Guerrero-Martínez, J. F., Bataller-Mompeán, M., & Francés-Víllora, J. V. (2015). Simplified spiking neural network architecture and STDP learning algorithm applied to image classification. *EURASIP Journal on Image and Video Processing, 2015*(1), 4. https://doi.org/10.1186/s13640-015-0059-4

Inouye, T. (1909). *Die Sehstörungen bei Schußverletzungen der kortikalen Sehsphäre* [The vision impairments from gunshot wounds in the cortical vision sphere]. Leipzig, Germany: Verlag von Wilhelm Engelmann. Retrieved from https://archive.org/details/b21295955

Iranmehr, E., Shouraki, S. B., Faraji, M. M., Bagheri, N., & Linares-Barranco, B. (2019). Bio-Inspired Evolutionary Model of Spiking Neural Networks in Ionic Liquid Space. *Frontiers in Neuroscience, 13*(1085). https://doi.org/10.3389/fnins.2019.01085

Izhikevich, E. M. (2003). Simple model of spiking neurons. *IEEE Transactions on Neural Networks, 14*(6), 1569-1572. https://doi.org/10.1109/TNN.2003.820440

Izhikevich, E. M. (2004). Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks, 15*(5), 1063-1070. https://doi.org/10.1109/TNN.2004.832719

Jackson, Z., Hereman, N., Walker, T., & Weveler, Y.-N. (2016). *Free Spoken Digits Dataset*. https://doi.org/10.5281/zenodo.1000741. Retrieved on 17 Oct 2019 from https://github.com/Jakobovski/free-spoken-digit-dataset/

Jay, M. F., & Sparks, D. L. (1987). Sensorimotor integration in the primate superior colliculus. I. Motor convergence. *Journal of Neurophysiology, 57*(1), 22-34. https://doi.org/10.1152/jn.1987.57.1.22

Ji, S., Xu, W., Yang, M., & Yu, K. (2013). 3D Convolutional Neural Networks for Human Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(1), 221-231. https://doi.org/10.1109/TPAMI.2012.59

Jiang, Z., Lin, Z., & Davis, L. S. (2012). Recognizing Human Actions by Learning and Matching Shape-Motion Prototype Trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(3), 533-547. https://doi.org/10.1109/tpami.2011.147

Jiménez-Fernández, Á., Cerezuela-Escudero, E., Miró-Amarante, L., Domínguez-Morales, M. J., de Asís Gómez-Rodríguez, F., Linares-Barranco, A., & Jiménez-Moreno, G. (2017). A Binaural Neuromorphic Auditory Sensor for FPGA: A Spike Signal Processing Approach. *IEEE Transactions on Neural Networks and Learning Systems, 28*(4), 804-818. https://doi.org/10.1109/TNNLS.2016.2583223

Kaggle Inc. (2018). *TensorFlow Speech Recognition Challenge - Private Leaderboard*. Retrieved on 30 Jul 2018 from https://www.kaggle.com/c/tensorflow-speech-recognition-challenge/leaderboard

Kagirov, I., Ryumin, D., & Axyonov, A. (2019, 20-25 Aug). *Method for Multimodal Recognition of One-Handed Sign Language Gestures Through 3D Convolution and LSTM Neural Networks*.

Presented at the International Conference on Speech and Computer (SPECOM), Istanbul, Turkey. https://doi.org/10.1007/978-3-030-26061-3_20

Kagirov, I., Ryumin, D., Axyonov, A., & Karpov, A. (2020). Мультимедийная база данных жестов русского жестового языка в трехмерном формате. [Multimedia Database of Russian Sign Language Items in 3D]. *Вопросы языкознания. [Topics in the Study of Language.]*(1), 104-123. https://doi.org/10.31857/S0373658X0008302-1

Kamal, S. M., Chen, Y., Li, S., Shi, X., & Zheng, J. (2019). Technical Approaches to Chinese Sign Language Processing: A Review. *IEEE Access, 7*, 96926-96935. https://doi.org/10.1109/ACCESS.2019.2929174

Kasabov, N. K. (2010). To spike or not to spike: A probabilistic spiking neuron model. *Neural Networks, 23*(1), 16-19. https://doi.org/10.1016/j.neunet.2009.08.010

Kasabov, N. K. (2014). NeuCube: A spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data. *Neural Networks, 52*, 62-76. https://doi.org/10.1016/j.neunet.2014.01.006

Kasabov, N. K., & Capecci, E. (2015). Spiking neural network methodology for modelling, classification and understanding of EEG spatio-temporal data measuring cognitive processes. *Information Sciences, 294*, 565-575. https://doi.org/10.1016/j.ins.2014.06.028

Kasabov, N. K., Dhoble, K., Nuntalid, N., & Indiveri, G. (2013). Dynamic evolving spiking neural networks for on-line spatio- and spectro-temporal pattern recognition. *Neural Networks, 41*, 188-201. https://doi.org/10.1016/j.neunet.2012.11.014

Kasabov, N. K., Postma, E., & van den Herik, J. (2000). AVIS: a connectionist-based framework for integrated auditory and visual information processing. *Information Sciences, 123*(1-2), 127-148. https://doi.org/10.1016/S0020-0255(99)00114-0

Kasabov, N. K., Scott, N. M., Tu, E., Marks, S., Sengupta, N., Capecci, E., . . . Yang, J. (2016). Evolving spatio-temporal data machines based on the NeuCube neuromorphic framework: Design methodology and selected applications. *Neural Networks, 78*, 1-14. https://doi.org/10.1016/j.neunet.2015.09.011

Kerr, D., McGinnity, T. M., Coleman, S., & Clogenson, M. (2015). A biologically inspired spiking model of visual processing for image feature detection. *Neurocomputing, 158*, 268-280. https://doi.org/10.1016/j.neucom.2015.01.011

Kheradpisheh, S. R., Ganjtabesh, M., Thorpe, S. J., & Masquelier, T. (2018). STDP-based spiking deep convolutional neural networks for object recognition. *Neural Networks, 99*, 56-67. https://doi.org/10.1016/j.neunet.2017.12.005

Kim, J.-G., Yu, B., & Lee, Y.-K. (2009). Correlations between Color Differences Based on Three Color-Difference Formulas Using Dental Shade Guide Tabs. *Journal of Prosthodontics, 18*(2), 135-140. https://doi.org/10.1111/j.1532-849X.2008.00393.x

Kim, T.-K., Wong, S.-F., & Cipolla, R. (2007, 17-22 Jun). *Tensor Canonical Correlation Analysis for Action Classification*. Presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Minneapolis, MN. https://doi.org/10.1109/cvpr.2007.383137

Kimpe, T., & Tuytschaever, T. (2007). Increasing the Number of Gray Shades in Medical Display Systems—How Much is Enough? *Journal of Digital Imaging, 20*(4), 422-432. https://doi.org/10.1007/s10278-006-1052-3

King, A. J., & Bajo, V. (2013). Cortical modulation of auditory processing in the midbrain. *Frontiers in Neural Circuits, 6*(114). https://doi.org/10.3389/fncir.2012.00114

King, A. J., Hammond-Kenny, A., & Nodal, F. R. (2019). Multisensory Processing in the Auditory Cortex. In A. K. C. Lee, M. T. Wallace, A. B. Coffin, A. N. Popper, & R. R. Fay (Eds.), *Multisensory Processes: The Auditory Perspective* (pp. 105-133). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-030-10461-0_6

Klein, D. J., König, P., & Körding, K. P. (2003). Sparse Spectrotemporal Coding of Sounds. *EURASIP Journal on Advances in Signal Processing, 2003*(7), 902061. https://doi.org/10.1155/S1110865703303051

Klekamp, J., Riedel, A., Harper, C., & Kretschmann, H.-J. (1991). Quantitative changes during the postnatal maturation of the human visual cortex. *Journal of the Neurological Sciences, 103*(2), 136-143. https://doi.org/10.1016/0022-510X(91)90156-2

Knösche, T., & Tittgemeyer, M. (2011). The Role of Long-Range Connectivity for the Characterization of the Functional–Anatomical Organization of the Cortex. *Frontiers in Systems Neuroscience, 5*(58). https://doi.org/10.3389/fnsys.2011.00058

Kolb, H. (2011a). Circuitry for rod signals through the retina. In H. Kolb (Ed.), *Webvision*. Retrieved from https://webvision.med.utah.edu/book/part-iii-retinal-circuits/circuitry-for-rod-cells-through-the-retina/

Kolb, H. (2011b). Cone pathways through the retina. In H. Kolb (Ed.), *Webvision*. Retrieved from https://webvision.med.utah.edu/book/part-iii-retinal-circuits/cone-pathways-through-the-retina/

Kolb, H. (2012). Midget pathways of the primate retina underlie resolution and red green color opponency. In H. Kolb (Ed.), *Webvision*. Retrieved from https://webvision.med.utah.edu/book/part-iii-retinal-circuits/midget-pathways-of-the-primate-retina-underly-resolution/

Koller, O., Forster, J., & Ney, H. (2015). Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. *Computer Vision and Image Understanding, 141*, 108-125. https://doi.org/10.1016/j.cviu.2015.09.013

König, A. (1895). Über die Anzahl der unterscheidbaren Spektralfarben und Helligkeitsstufen [On the number of distinguishable spectral colours and brightness levels]. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane, 8*, 375-380. Retrieved from https://digitalesammlungen.uni-weimar.de/viewer/image/lit29567/1/

Köpüklü, O., Gunduz, A., Kose, N., & Rigoll, G. (2020). Online Dynamic Hand Gesture Recognition Including Efficiency Analysis. *IEEE Transactions on Biometrics, Behavior, and Identity Science, 2*(2), 85-97. https://doi.org/10.1109/TBIOM.2020.2968216

Köpüklü, O., Köse, N., & Rigoll, G. (2018, 18-22 Jun). *Motion Fused Frames: Data Level Fusion Strategy for Hand Gesture Recognition*. Presented at the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT. https://doi.org/10.1109/CVPRW.2018.00284

Kreit, E., Mäthger, L. M., Hanlon, R. T., Dennis, P. B., Naik, R. R., Forsythe, E., & Heikenfeld, J. (2013). Biological versus electronic adaptive coloration: how can one inform the other? *Journal of The Royal Society Interface, 10*(78), 20120601. https://doi.org/10.1098/rsif.2012.0601

Kriegeskorte, N. (2015). Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. *Annual Review of Vision Science, 1*(1), 417-446. https://doi.org/10.1146/annurev-vision-082114-035447

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012, 3-6 Dec). *ImageNet classification with deep convolutional neural networks*. Presented at the International Conference on Neural Information Processing Systems (NIPS), Lake Tahoe, NV. https://doi.org/10.1145/3065386

Kubovy, M., & van Valkenburg, D. (2001). Auditory and visual objects. *Cognition, 80*(1-2), 97-126. https://doi.org/10.1016/s0010-0277(00)00155-4

Kuehni, R. G. (2002). CIEDE2000, milestone or final answer? *Color Research & Application, 27*(2), 126-127. https://doi.org/10.1002/col.10035

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology, 16*(1), 37-68. https://doi.org/10.1152/jn.1953.16.1.37

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology, 62*(1), 621-647. https://doi.org/10.1146/annurev.psych.093008.131123

Laird, A. R., Robinson, J. L., McMillan, K. M., Tordesillas-Gutiérrez, D., Moran, S. T., Gonzales, S. M., . . . Lancaster, J. L. (2010). Comparison of the disparity between Talairach and MNI coordinates in functional neuroimaging data: Validation of the Lancaster transform. *NeuroImage, 51*(2), 677-683. https://doi.org/10.1016/j.neuroimage.2010.02.048

Lambertz, N., Gizewski, E. R., de Greiff, A., & Forsting, M. (2005). Cross-modal plasticity in deaf subjects dependent on the extent of hearing loss. *Cognitive Brain Research, 25*(3), 884-890. https://doi.org/10.1016/j.cogbrainres.2005.09.010

Lancaster, J. L., Rainey, L. H., Summerlin, J. L., Freitas, C. S., Fox, P. T., Evans, A. C., . . . Mazziotta, J. C. (1997). Automated labeling of the human brain: A preliminary report on the development and evaluation of a forward-transform method. *Human Brain Mapping, 5*(4), 238-242. https://doi.org/10.1002/(sici)1097-0193(1997)5:4<238::Aid-hbm6>3.0.Co;2-4

Lancaster, J. L., Tordesillas-Gutiérrez, D., Martinez, M., Salinas, F., Evans, A. C., Zilles, K., . . . Fox, P. T. (2007). Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human Brain Mapping, 28*(11), 1194-1205. https://doi.org/10.1002/hbm.20345

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L. H., . . . Fox, P. T. (2000). Automated Talairach Atlas labels for functional brain mapping. *Human Brain Mapping, 10*(3), 120-131. https://doi.org/10.1002/1097-0193(200007)10:3<120::Aid-hbm30>3.0.Co;2-8

Langers, D. R. M., Sanchez-Panchuelo, R. M., Francis, S. T., Krumbholz, K., & Hall, D. A. (2014). Neuroimaging paradigms for tonotopic mapping (II): The influence of acquisition protocol. *NeuroImage, 100*, 663-675. https://doi.org/10.1016/j.neuroimage.2014.07.042

Langers, D. R. M., van Dijk, P., & Backes, W. H. (2005). Lateralization, connectivity and plasticity in the human central auditory system. *NeuroImage, 28*(2), 490-499. https://doi.org/10.1016/j.neuroimage.2005.06.024

Laptev, I. (2004). *Local spatio-temporal image features for motion interpretation*. Kungliga Tekniska Högskolan [Royal Institute of Technology], Stockholm, Sweden. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-3797

Lech, R. K., & Suchan, B. (2014). Involvement of the human medial temporal lobe in a visual discrimination task. *Behavioural Brain Research, 268*, 22-30. https://doi.org/10.1016/j.bbr.2014.03.030

LeCun, Y., Bengio, Y., & Hinton, G. E. (2015). Deep learning. *Nature, 521*(7553), 436-444. https://doi.org/10.1038/nature14539

Lee, A. K. C., Maddox, R. K., & Bizley, J. K. (2019). An Object-Based Interpretation of Audiovisual Processing. In A. K. C. Lee, M. T. Wallace, A. B. Coffin, A. N. Popper, & R. R. Fay (Eds.), *Multisensory Processes: The Auditory Perspective* (pp. 59-83). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-030-10461-0_4

Lee, A. K. C., & Wallace, M. T. (2019). Visual Influence on Auditory Perception. In A. K. C. Lee, M. T. Wallace, A. B. Coffin, A. N. Popper, & R. R. Fay (Eds.), *Multisensory Processes: The Auditory Perspective* (pp. 1-8). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-030-10461-0_1

Lee, C., Srinivasan, G., Panda, P., & Roy, K. (2019). Deep Spiking Convolutional Neural Network Trained With Unsupervised Spike-Timing-Dependent Plasticity. *IEEE Transactions on Cognitive and Developmental Systems, 11*(3), 384-394. https://doi.org/10.1109/TCDS.2018.2833071

Lee, J., & Groh, J. M. (2014). Different Stimuli, Different Spatial Codes: A Visual Map and an Auditory Rate Code for Oculomotor Space in the Primate Superior Colliculus. *PLOS ONE, 9*(1), e85017. https://doi.org/10.1371/journal.pone.0085017

Lee, J. H., Delbrück, T., & Pfeiffer, M. (2016). Training Deep Spiking Neural Networks Using Backpropagation. *Frontiers in Neuroscience, 10*(508). https://doi.org/10.3389/fnins.2016.00508

Lee, M., Lee, S., Son, S., Park, G., & Kwak, N. (2018, 8-14 Sep). *Motion Feature Network: Fixed Motion Filter for Action Recognition*. Presented at the European Conference on Computer Vision (ECCV), Munich, Germany. https://doi.org/10.1007/978-3-030-01249-6_24

Lee, S. C. S., Martin, P. R., & Grünert, U. (2019). Topography of Neurons in the Rod Pathway of Human Retina. *Investigative Ophthalmology & Visual Science, 60*(8), 2848-2859. https://doi.org/10.1167/iovs.19-27217

Lee, Y.-K. (2005). Comparison of CIELAB ΔE* and CIEDE2000 color-differences after polymerization and thermocycling of resin composites. *Dental Materials, 21*(7), 678-682. https://doi.org/10.1016/j.dental.2004.09.005

Leonard, R. (1984, 19-21 Mar). *A database for speaker-independent digit recognition*. Presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), San Diego, CA. https://doi.org/10.1109/ICASSP.1984.1172716

Leuba, G., & Kraftsik, R. (1994a). Changes in volume, surface estimate, three-dimensional shape and total number of neurons of the human primary visual cortex from midgestation until old age. *Anatomy and Embryology, 190*(4), 351-366. https://doi.org/10.1007/BF00187293

Leuba, G., & Kraftsik, R. (1994b). Visual cortex in Alzheimer's disease: Occurencee of neuronal death and glial proliferation, and correlation with pathological hallmarks. *Neurobiology of Aging, 15*(1), 29-43. https://doi.org/10.1016/0197-4580(94)90142-2

Lewis, J. W., Beauchamp, M. S., & DeYoe, E. A. (2000). A Comparison of Visual and Auditory Motion Processing in Human Cerebral Cortex. *Cerebral Cortex, 10*(9), 873-888. https://doi.org/10.1093/cercor/10.9.873

Li, K., & Príncipe, J. C. (2018). Biologically-Inspired Spike-Based Automatic Speech Recognition of Isolated Digits Over a Reproducing Kernel Hilbert Space. *Frontiers in Neuroscience, 12*(194). https://doi.org/10.3389/fnins.2018.00194

Lian, Y., Grayden, D. B., Kameneva, T., Meffin, H., & Burkitt, A. N. (2019). Toward a Biologically Plausible Model of LGN-V1 Pathways Based on Efficient Coding. *Frontiers in Neural Circuits, 13*(13). https://doi.org/10.3389/fncir.2019.00013

Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In *Phonology and reading disability: Solving the reading puzzle.* (pp. 1-33). Ann Arbor, MI: The University of Michigan Press. Retrieved from https://eric.ed.gov/?id=ED427291

Liberman, M. C. (1978). Auditory-nerve response from cats raised in a low-noise chamber. *The Journal of the Acoustical Society of America, 63*(2), 442-455. https://doi.org/10.1121/1.381736

Lichtsteiner, P., Posch, C., & Delbrück, T. (2008). A 128 × 128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor. *IEEE Journal of Solid-State Circuits, 43*(2), 566-576. https://doi.org/10.1109/JSSC.2007.914337

Lin, T.-H. (2016). Hidden Watermark of 3D Models by Just Noticeable Color Difference. *Electronic Imaging, 2016*(19), 1-4. https://doi.org/10.2352/ISSN.2470-1173.2016.19.COIMG-175

Litovsky, R. Y., Fligor, B. J., & Tramo, M. J. (2002). Functional role of the human inferior colliculus in binaural hearing. *Hearing Research, 165*(1), 177-188. https://doi.org/10.1016/S0378-5955(02)00304-0

Liu, D., Bellotto, N., & Yue, S. (2020). Deep Spiking Neural Network for Video-Based Disguise Face Recognition Based on Dynamic Facial Movements. *IEEE Transactions* *on Neural Networks and Learning Systems, 31*(6), 1843-1855. https://doi.org/10.1109/TNNLS.2019.2927274

Liu, H., Shu, N., Tang, Q., & Zhang, W. (2018). Computational Model Based on Neural Network of Visual Cortex for Human Action Recognition. *IEEE Transactions on Neural Networks and Learning Systems, 29*(5), 1427-1440. https://doi.org/10.1109/TNNLS.2017.2669522

Liu, J., Perez-Gonzalez, D., Rees, A., Erwin, H., & Wermter, S. (2010). A biologically inspired spiking neural network model of the auditory midbrain for sound source localisation. *Neurocomputing, 74*(1), 129-139. https://doi.org/10.1016/j.neucom.2009.10.030

Liu, Q., Pan, G., Ruan, H., Xing, D., Xu, Q., & Tang, H. (2020). Unsupervised AER Object Recognition Based on Multiscale Spatio-Temporal Features and Spiking Neurons. *IEEE Transactions on Neural Networks and Learning Systems, 31*(12), 5300-5311. https://doi.org/10.1109/TNNLS.2020.2966058

Liu, S.-C., van Schaik, A., Minch, B. A., & Delbrück, T. (2014). Asynchronous Binaural Spatial Audition Sensor With 2×64×4 Channel Output. *IEEE Transactions on Biomedical Circuits and Systems, 8*(4), 453-464. https://doi.org/10.1109/TBCAS.2013.2281834

Lu, L., Zhang, H.-J., & Li, S. Z. (2003). Content-based audio classification and segmentation by using support vector machines. *Multimedia Systems, 8*(6), 482-492. https://doi.org/10.1007/s00530-002-0065-0

Luo, M. R., Cui, G., & Rigg, B. (2001). The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research & Application, 26*(5), 340-350. https://doi.org/10.1002/col.1049

Luo, M. R., Cui, G., & Rigg, B. (2002). Further comments on CIEDE2000. *Color Research & Application, 27*(2), 127-128. https://doi.org/10.1002/col.10043

Luo, M. R., Minchew, C., Kenyon, P., & Cui, G. (2004, 3-5 Nov). *Verification of CIEDE2000 using industrial data*. Presented at the AIC 2004 Color and Paints - Interim Meeting of the International Color Association, Porto Alegre, Brazil. Retrieved from https://www.aic-color.org/resources/Documents/aic2004.pdf

Lyon, R. F. (1982, 3-5 May). *A computational model of filtering, detection, and compression in the cochlea*. Presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Paris, France. https://doi.org/10.1109/ICASSP.1982.1171644

Lyon, R. F., & Dyer, L. (1986, 7-11 Apr). *Experiments with a computational model of the cochlea*. Presented at the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Tokyo, Japan. https://doi.org/10.1109/ICASSP.1986.1168650

Lyon, R. F., Ponte, J., & Chechik, G. (2011, 22-27 May). *Sparse coding of auditory features for machine hearing in interference*. Presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic. https://doi.org/10.1109/ICASSP.2011.5947698

Lyon, R. F., Rehn, M., Bengio, S., Walters, T. C., & Chechik, G. (2010). Sound Retrieval and Ranking Using Sparse Auditory Representations. *Neural Computation, 22*(9), 2390-2416. https://doi.org/10.1162/NECO_a_00011

Maass, W. (1997). Networks of spiking neurons: The third generation of neural network models. *Neural Networks, 10*(9), 1659-1671. https://doi.org/10.1016/S0893-6080(97)00011-7

Maass, W., Natschläger, T., & Markram, H. (2002). Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations. *Neural Computation, 14*(11), 2531-2560. https://doi.org/10.1162/089976602760407955

Mansencal, T., Mauderer, M., Parsons, M., Shaw, N., Wheatley, K., Cooper, S., . . . Schambach, M. (2020). Colour Science for Python.

https://doi.org/10.5281/zenodo.605791. Retrieved from https://github.com/colour-science/colour
Marcus, M. P., Marcinkiewicz, M. A., & Santorini, B. (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19*(2), 313-330.
Marks, S. (2017). Immersive visualisation of 3-dimensional spiking neural networks. *Evolving Systems, 8*(3), 193-201. https://doi.org/10.1007/s12530-016-9170-8
Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature, 379*(6566), 649-652. https://doi.org/10.1038/379649a0
Masland, R. H. (2012). The Neuronal Organization of the Retina. *Neuron, 76*(2), 266-280. https://doi.org/10.1016/j.neuron.2012.10.002
Materzynska, J., Berger, G., Bax, I., & Memisevic, R. (2019, 27-28 Oct). *The Jester Dataset: A Large-Scale Video Dataset of Human Gestures*. Presented at the IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, South Korea. https://doi.org/10.1109/ICCVW.2019.00349
McKee, D., McKee, R., Pivac Alexander, S., Pivac, L., & Vale, M. (2011). *Online dictionary of New Zealand Sign Language*. Retrieved on 20 Oct 2020 from http://nzsl.nz
McLoughlin, I., Zhang, H., Xie, Z., Song, Y., & Xiao, W. (2015). Robust Sound Event Classification Using Deep Neural Networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 23*(3), 540-552. https://doi.org/10.1109/TASLP.2015.2389618
Mecklinger, A., & Ullsperger, P. (1993). P3 varies with stimulus categorization rather than probability. *Electroencephalography and Clinical Neurophysiology, 86*(6), 395-407. https://doi.org/10.1016/0013-4694(93)90135-I
Meddis, R., Lecluyse, W., Clark, N. R., Jürgens, T., Tan, C. M., Panda, M. R., & Brown, G. J. (2013). A Computer Model of the Auditory Periphery and Its Application to the Study of Hearing. In B. C. J. Moore, R. D. Patterson, I. M. Winter, R. P. Carlyon, & H. E. Gockel (Eds.), *Basic Aspects of Hearing. Advances in Experimental Medicine and Biology, vol 787* (pp. 11-20). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-1590-9_2
Meijer, D., Veselič, S., Calafiore, C., & Noppeney, U. (2019). Integration of audiovisual spatial signals is not consistent with maximum likelihood estimation. *Cortex, 119*, 74-88. https://doi.org/10.1016/j.cortex.2019.03.026
Merbs, S. L., & Nathans, J. (1992). Absorption spectra of human cone pigments. *Nature, 356*(6368), 433-435. https://doi.org/10.1038/356433a0
Merriam-Webster. (n.d.-a). *Hearing*. Retrieved on 10 Mar 2020 from https://www.merriam-webster.com/dictionary/hearing
Merriam-Webster. (n.d.-b). *Vision*. Retrieved on 19 Apr 2018 from https://www.merriam-webster.com/dictionary/vision
Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature, 485*, 233. https://doi.org/10.1038/nature11020
Miller, G. A., & Fellbaum, C. (2007). WordNet then and now. *Language Resources and Evaluation, 41*(2), 209-214. https://doi.org/10.1007/s10579-007-9044-6
Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia, 46*(3), 774-785. https://doi.org/10.1016/j.neuropsychologia.2007.10.005
Minsky, M. L. (1961). Steps toward Artificial Intelligence. *Proceedings of the IRE, 49*(1), 8-30. https://doi.org/10.1109/JRPROC.1961.287775
Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting Human Brain Activity Associated with the Meanings of Nouns. *Science, 320*(5880), 1191-1195. https://doi.org/10.1126/science.1152876
Molholm, S., Sehatpour, P., Mehta, A. D., Shpaner, M., Gomez-Ramirez, M., Ortigue, S., . . . Foxe, J. J. (2006). Audio-Visual Multisensory Integration in Superior Parietal Lobule

Revealed by Human Intracranial Recordings. *Journal of Neurophysiology, 96*(2), 721-729. https://doi.org/10.1152/jn.00285.2006

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., & Zilles, K. (2001). Human Primary Auditory Cortex: Cytoarchitectonic Subdivisions and Mapping into a Spatial Reference System. *NeuroImage, 13*(4), 684-701. https://doi.org/10.1006/nimg.2000.0715

Murphy, B., & Poesio, M. (2010, 6 Jun). *Detecting semantic category in simultaneous EEG/MEG recordings*. Presented at the First Workshop on Computational Neurolinguistics, Los Angeles, CA. Retrieved from https://dl.acm.org/doi/10.5555/1866686.1866691

Murphy, B., Poesio, M., Bovolo, F., Bruzzone, L., Dalponte, M., & Lakany, H. (2011). EEG decoding of semantic category reveals distributed representations for single concepts. *Brain and Language, 117*(1), 12-22. https://doi.org/10.1016/j.bandl.2010.09.013

Murphy, G. M. (1985). Volumetric asymmetry in the human striate cortex. *Experimental Neurology, 88*(2), 288-302. https://doi.org/10.1016/0014-4886(85)90192-X

Nadol, J. B. (1988). Comparative anatomy of the cochlea and auditory nerve in mammals. *Hearing Research, 34*(3), 253-266. https://doi.org/10.1016/0378-5955(88)90006-8

Näger, C., Storck, J., & Deco, G. (2002). Speech recognition with spiking neurons and dynamic synapses: a model motivated by the human auditory pathway. *Neurocomputing, 44-46*, 937-942. https://doi.org/10.1016/S0925-2312(02)00494-0

Namarvar, H. H., Liaw, J.-S., & Berger, T. W. (2001, 15-19 Jul). *A new dynamic synapse neural network for speech recognition*. Presented at the International Joint Conference on Neural Networks (IJCNN), Washington, DC. https://doi.org/10.1109/IJCNN.2001.938853

Neil, D., & Liu, S.-C. (2016, 22-25 May). *Effective sensor fusion with event-based sensors and deep network architectures*. Presented at the IEEE International Symposium on Circuits and Systems (ISCAS), Montreal, Canada. https://doi.org/10.1109/ISCAS.2016.7539039

Neitz, J., & Neitz, M. (2008). Colour Vision: The Wonder of Hue. *Current Biology, 18*(16), R700-R702. https://doi.org/10.1016/j.cub.2008.06.062

Nelken, I., Chechik, G., Mrsic-Flogel, T. D., King, A. J., & Schnupp, J. W. H. (2005). Encoding Stimulus Information by Spike Numbers and Mean Response Time in Primary Auditory Cortex. *Journal of Computational Neuroscience, 19*(2), 199-221. https://doi.org/10.1007/s10827-005-1739-3

NeuroBehavioralSystems. (n.d.). Presentation. Retrieved from https://www.neurobs.com/index_html

Noda, K., Yamaguchi, Y., Nakadai, K., Okuno, H. G., & Ogata, T. (2015). Audio-visual speech recognition using deep learning. *Applied Intelligence, 42*(4), 722-737. https://doi.org/10.1007/s10489-014-0629-7

Oertel, D., & Doupe, A. J. (2013). The Auditory Central Nervous System. In E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, A. J. Hudspeth, & S. Mack (Eds.), *Principles of Neural Science* (5th ed., pp. 682-711). New York, NY: McGraw-Hill.

Oess, T., Löhr, M. P. R., Schmid, D., Ernst, M. O., & Neumann, H. (2020). From Near-Optimal Bayesian Integration to Neuromorphic Hardware: A Neural Network Model of Multisensory Integration. *Frontiers in Neurorobotics, 14*(29). https://doi.org/10.3389/fnbot.2020.00029

Okkalides, D. (1996). Perception of detail and greyscale range in X-ray fluoroscopy images captured with a personal computer and frame-grabber. *European Journal of Radiology, 23*(2), 149-158. https://doi.org/10.1016/0720-048X(96)01064-9

Olson, C. R., & Colby, C. L. (2013). The Organization of Cognition. In E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, A. J. Hudspeth, & S. Mack (Eds.), *Principles of Neural Science* (5th ed., pp. 392-411). New York, NY: McGraw-Hill.

Olson, I. R., Gatenby, J. C., & Gore, J. C. (2002). A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Cognitive Brain Research, 14*(1), 129-138. https://doi.org/10.1016/S0926-6410(02)00067-8

Ortiz-Jaramillo, B., Kumcu, A., & Philips, W. (2016, 6-8 Jun). *Evaluating color difference measures in images*. Presented at the International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal. https://doi.org/10.1109/QoMEX.2016.7498922

Ortiz-Jaramillo, B., Kumcu, A., Platisa, L., & Philips, W. (2019). Evaluation of color differences in natural scene color images. *Signal Processing: Image Communication, 71*, 128-137. https://doi.org/10.1016/j.image.2018.11.009

Pakhomov, S. (2014). *MNI Space utility*. Retrieved on 7 Oct 2020 from https://web.archive.org/web/20141015201028/http://www.ihb.spb.ru:80/%7Epet_lab/MSU/MSUMain.html

Paulun, L., Wendt, A., & Kasabov, N. K. (2018). A Retinotopic Spiking Neural Network System for Accurate Recognition of Moving Objects Using NeuCube and Dynamic Vision Sensors. *Frontiers in Computational Neuroscience, 12*(42). https://doi.org/10.3389/fncom.2018.00042

Pernet, C., & Belin, P. (2012). The Role of Pitch and Timbre in Voice Gender Categorization. *Frontiers in Psychology, 3*(23). https://doi.org/10.3389/fpsyg.2012.00023

Petró, B., Kasabov, N. K., & Kiss, R. M. (2020). Selection and Optimization of Temporal Spike Encoding Methods for Spiking Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems, 31*(2), 358-370. https://doi.org/10.1109/TNNLS.2019.2906158

Petro, L. S., Paton, A. T., & Muckli, L. (2017). Contextual modulation of primary visual cortex by auditory signals. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1714), 20160104. https://doi.org/10.1098/rstb.2016.0104

Pfeiffer, M., & Pfeil, T. (2018). Deep Learning With Spiking Neurons: Opportunities and Challenges. *Frontiers in Neuroscience, 12*(774). https://doi.org/10.3389/fnins.2018.00774

Piczak, K. J. (2015, 17-20 Sep). *Environmental sound classification with convolutional neural networks*. Presented at the IEEE International Workshop on Machine Learning for Signal Processing (MLSP), Boston, MA. https://doi.org/10.1109/MLSP.2015.7324337

Plakke, B., & Romanski, L. M. (2019). Audiovisual Integration in the Primate Prefrontal Cortex. In A. K. C. Lee, M. T. Wallace, A. B. Coffin, A. N. Popper, & R. R. Fay (Eds.), *Multisensory Processes: The Auditory Perspective* (pp. 135-159). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-030-10461-0_7

Polimeni, J. R., Fischl, B., Greve, D. N., & Wald, L. L. (2010). Laminar analysis of 7T BOLD using an imposed spatial activation pattern in human V1. *NeuroImage, 52*(4), 1334-1346. https://doi.org/10.1016/j.neuroimage.2010.05.005

Poria, S., Cambria, E., Howard, N., Huang, G.-B., & Hussain, A. (2016). Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing, 174*, 50-59. https://doi.org/10.1016/j.neucom.2015.01.095

Porter, K. K., & Groh, J. M. (2006). The "other" transformation required for visual-auditory integration: representational format. In S. Martinez-Conde, S. L. Macknik, L. M. Martinez, J. M. Alonso, & P. U. Tse (Eds.), *Progress in Brain Research* (Vol. 155, pp. 313-323). Amsterdam, The Netherlands: Elsevier. https://doi.org/10.1016/S0079-6123(06)55018-6

Pourtois, G., Schettino, A., & Vuilleumier, P. (2013). Brain mechanisms for emotional influences on perception and attention: What is magic and what is not. *Biological Psychology, 92*(3), 492-512. https://doi.org/10.1016/j.biopsycho.2012.02.007

Princeton University. (n.d.). *wnstats - WordNet 3.0 database statistics*. Retrieved on 28 Feb 2020 from https://wordnet.princeton.edu/documentation/wnstats7wn

Pugh, E. N., & Cobbs, W. H. (1986). Visual transduction in vertebrate rods and cones: A tale of two transmitters, calcium and cyclic GMP. *Vision Research, 26*(10), 1613-1643. https://doi.org/10.1016/0042-6989(86)90051-9

Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A.-S., Mooney, R. D., . . . White, L. E. (2017). *Neuroscience* (6th ed.). Oxford, United Kingdom: Oxford University Press.

Rabin, J., Gooch, J., & Ivan, D. (2011). Rapid Quantification of Color Vision: The Cone Contrast Test. *Investigative Ophthalmology & Visual Science, 52*(2), 816-820. https://doi.org/10.1167/iovs.10-6283

Rabiner, L. R., Wilpon, J. G., & Soong, F. K. (1989). High performance connected digit recognition using hidden Markov models. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 37*(8), 1214-1225. https://doi.org/10.1109/29.31269

Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H. J., & Zilles, K. (2001). Probabilistic Mapping and Volume Measurement of Human Primary Auditory Cortex. *NeuroImage, 13*(4), 669-683. https://doi.org/10.1006/nimg.2000.0714

Rafegas, I., & Vanrell, M. (2018). Color encoding in biologically-inspired convolutional neural networks. *Vision Research, 151*, 7-17. https://doi.org/10.1016/j.visres.2018.03.010

Rao, R. (2016, 24-29 Jul). *A spatio-temporal model of multi-sensory learning that demonstrates improved object recall*. Presented at the International Joint Conference on Neural Networks (IJCNN), Vancouver, Canada. https://doi.org/10.1109/IJCNN.2016.7727864

Rathi, N., & Roy, K. (2021). STDP Based Unsupervised Multimodal Learning With Cross-Modal Processing in Spiking Neural Networks. *IEEE Transactions on Emerging Topics in Computational Intelligence, 5*(1), 143-153. https://doi.org/10.1109/TETCI.2018.2872014

Rauschecker, J. P. (2013). Processing Streams in Auditory Cortex. In Y. E. Cohen, A. N. Popper, & R. R. Fay (Eds.), *Neural Correlates of Auditory Cognition* (pp. 7-43). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-2350-8_2

Rauschecker, J. P. (2015). Auditory and visual cortex of primates: a comparison of two sensory systems. *The European Journal of Neuroscience, 41*(5), 579-585. https://doi.org/10.1111/ejn.12844

Rauschecker, J. P. (2018). Where, When, and How: Are they all sensorimotor? Towards a unified view of the dorsal pathway in vision and audition. *Cortex, 98*, 262-268. https://doi.org/10.1016/j.cortex.2017.10.020

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences, 97*(22), 11800-11806. https://doi.org/10.1073/pnas.97.22.11800

Rawat, W., & Wang, Z. (2017). Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation, 29*(9), 2352-2449. https://doi.org/10.1162/neco_a_00990

Read, H. L., Winer, J. A., & Schreiner, C. E. (2002). Functional architecture of auditory cortex. *Current Opinion in Neurobiology, 12*(4), 433-440. https://doi.org/10.1016/S0959-4388(02)00342-2

Reichenbach, A., Agte, S., Francke, M., & Franze, K. (2014). How light traverses the inverted vertebrate retina. *e-Neuroforum, 5*(4), 93-100. https://doi.org/10.1007/s13295-014-0054-8

Reimann, H. M. (2011). Signal processing in the cochlea: the structure equations. *The Journal of Mathematical Neuroscience, 1*(1), 5. https://doi.org/10.1186/2190-8567-1-5

Ren, Z., Qian, K., Zhang, Z., Pandit, V., Baird, A., & Schuller, B. (2018). Deep Scalogram Representations for Acoustic Scene Classification. *IEEE/CAA Journal of Automatica Sinica, 5*(3), 662-669. https://doi.org/10.1109/JAS.2018.7511066

Rolls, E. T. (1996). Visual Processing in the Temporal Lobe for Invariant Object Recognition. In V. Torre & F. Conti (Eds.), *Neurobiology: Ionic Channels, Neurons and the Brain* (pp. 325-353). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4615-5899-6_27

Romanski, L. M. (2007). Representation and Integration of Auditory and Visual Stimuli in the Primate Ventral Lateral Prefrontal Cortex. *Cerebral Cortex, 17*(suppl_1), i61-i69. https://doi.org/10.1093/cercor/bhm099

Rose, A. (1948). The Sensitivity Performance of the Human Eye on an Absolute Scale. *Journal of the Optical Society of America, 38*(2), 196-208. https://doi.org/10.1364/JOSA.38.000196

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review, 65*(6), 386. https://doi.org/10.1037/h0042519

Roy, D. K., & Pentland, A. P. (2002). Learning words from sights and sounds: a computational model. *Cognitive Science, 26*(1), 113-146. https://doi.org/10.1207/s15516709cog2601_4

Rudnicki, M., & Hemmert, W. (2016). Cochlea: inner ear models in Python. https://doi.org/10.5281/zenodo.595747. Retrieved from https://github.com/mrkrd/cochlea

Rudnicki, M., Schoppe, O., Isik, M., Völk, F., & Hemmert, W. (2015). Modeling auditory coding: from sound to spikes. *Cell and Tissue Research, 361*(1), 159-175. https://doi.org/10.1007/s00441-015-2202-z

Ruffieux, S., Lalanne, D., Mugellini, E., & Abou Khaled, O. (2014, 22-27 Jun). *A Survey of Datasets for Human Gesture Recognition*. Presented at the International Conference on Human-Computer Interaction, Heraklion, Greece. https://doi.org/10.1007/978-3-319-07230-2_33

Rule, J. S., & Riesenhuber, M. (2021). Leveraging Prior Concept Learning Improves Generalization From Few Examples in Computational Models of Human Object Recognition. *Frontiers in Computational Neuroscience, 14*(117). https://doi.org/10.3389/fncom.2020.586671

Ruyter, I. E., Nilner, K., & Möller, B. (1987). Color stability of dental composite resin materials for crown and bridge veneers. *Dental Materials, 3*(5), 246-251. https://doi.org/10.1016/S0109-5641(87)80081-7

Saenz, M., & Langers, D. R. M. (2014). Tonotopic mapping of human auditory cortex. *Hearing Research, 307*, 42-52. https://doi.org/10.1016/j.heares.2013.07.016

Sahidullah, M., & Saha, G. (2012). Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. *Speech Communication, 54*(4), 543-565. https://doi.org/10.1016/j.specom.2011.11.004

Sainath, T. N., Kingsbury, B., Saon, G., Soltau, H., Mohamed, A.-r., Dahl, G. E., & Ramabhadran, B. (2015). Deep Convolutional Neural Networks for Large-scale Speech Tasks. *Neural Networks, 64*, 39-48. https://doi.org/10.1016/j.neunet.2014.08.005

Sainath, T. N., Mohamed, A.-r., Kingsbury, B., & Ramabhadran, B. (2013, 26-31 May). *Deep convolutional neural networks for LVCSR*. Presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, Canada. https://doi.org/10.1109/ICASSP.2013.6639347

Salamon, J., & Bello, J. P. (2017). Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification. *IEEE Signal Processing Letters, 24*(3), 279-283. https://doi.org/10.1109/LSP.2017.2657381

Sander, K., & Scheich, H. (2001). Auditory perception of laughing and crying activates human amygdala regardless of attentional state. *Cognitive Brain Research, 12*(2), 181-198. https://doi.org/10.1016/S0926-6410(01)00045-3

Saraceno, G. (2017). *Deep learning and memorizing of spectro-temporal data (music) in the spatio-temporal brain*. University of Trento, Trento, Italy. Retrieved from https://webapps.unitn.it/Biblioteca/it/Web/RichiestaConsultazioneTesi/362886

Schafer, P. B., & Jin, D. Z. (2014). Noise-Robust Speech Recognition Through Auditory Feature Detection and Spike Sequence Decoding. *Neural Computation, 26*(3), 523-556. https://doi.org/10.1162/NECO_a_00557

Scheibe, P., Lazareva, A., Braumann, U.-D., Reichenbach, A., Wiedemann, P., Francke, M., & Rauscher, F. G. (2014). Parametric model for the 3D reconstruction of individual fovea shape from OCT data. *Experimental Eye Research, 119*, 19-26. https://doi.org/10.1016/j.exer.2013.11.008

Schira, M. M., Wade, A. R., & Tyler, C. W. (2007). Two-Dimensional Mapping of the Central and Parafoveal Visual Field to Human Visual Cortex. *Journal of Neurophysiology, 97*(6), 4284-4295. https://doi.org/10.1152/jn.00972.2006

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85-117. https://doi.org/10.1016/j.neunet.2014.09.003

Schmidt, B. P., Neitz, M., & Neitz, J. (2014). Neurobiological hypothesis of color appearance and hue perception. *Journal of the Optical Society of America, 31*(4), A195-A207. https://doi.org/10.1364/JOSAA.31.00A195

Schnupp, J. W. H., Honey, C., & Willmore, B. D. B. (2013). Neural Correlates of Auditory Object Perception. In Y. E. Cohen, A. N. Popper, & R. R. Fay (Eds.), *Neural Correlates of Auditory Cognition* (pp. 115-149). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-2350-8_5

Schnupp, J. W. H., Nelken, I., & King, A. J. (2011). *Auditory neuroscience: making sense of sound*. Cambridge, MA: MIT Press. Retrieved from https://mitpress.mit.edu/books/auditory-neuroscience

Schrauwen, B., D' Haene, M., Verstraeten, D., & Van Campenhout, J. (2007, 12-17 Aug). *Compact hardware for real-time speech recognition using a Liquid State Machine*. Presented at the International Joint Conference on Neural Networks (IJCNN), Orlando, FL. https://doi.org/10.1109/IJCNN.2007.4371111

Scott, N. M. (2015). *Evolving Spiking Neural Networks for Spatio- and Spectro- Temporal Data Analysis: Models, Implementations, Applications*. Auckland University of Technology, Auckland, New Zealand. Retrieved from http://hdl.handle.net/10292/10601

Semple, M. N., & Scott, B. H. (2003). Cortical mechanisms in hearing. *Current Opinion in Neurobiology, 13*(2), 167-173. https://doi.org/10.1016/s0959-4388(03)00048-5

Sharma, G. (2003). Color fundamentals for digital imaging. In G. Sharma (Ed.), *Digital Color Imaging Handbook* (pp. 1-114). Boca Raton, FL: CRC Press. https://doi.org/10.1201/9781420041484-1

Sharma, G., Wu, W., & Dalal, E. N. (2005). The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application, 30*(1), 21-30. https://doi.org/10.1002/col.20070

Sharmin, R., Rahut, S. K., & Huq, M. R. (2020). Bengali Spoken Digit Classification: A Deep Learning Approach Using Convolutional Neural Network. *Procedia Computer Science, 171*, 1381-1388. https://doi.org/10.1016/j.procs.2020.04.148

Shen, J., Zhao, Y., Liu, J. K., & Wang, Y. (2020, 19-24 Jul). *Recognizing Scoring in Basketball Game from AER Sequence by Spiking Neural Networks*. Presented at the International Joint Conference on Neural Networks (IJCNN), Glasgow, United Kingdom. https://doi.org/10.1109/IJCNN48605.2020.9207568

Shen, S., & Berns, R. S. (2011). Color-difference formula performance for several datasets of small color differences based on visual uncertainty. *Color Research & Application, 36*(1), 15-26. https://doi.org/10.1002/col.20595

Shepard, K. N., Kilgard, M. P., & Liu, R. C. (2013). Experience-Dependent Plasticity and Auditory Cortex. In Y. E. Cohen, A. N. Popper, & R. R. Fay (Eds.), *Neural Correlates of Auditory Cognition* (pp. 293-327). New York, NY: Springer. https://doi.org/10.1007/978-1-4614-2350-8_10

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences, 12*(5), 182-186. https://doi.org/10.1016/j.tics.2008.02.003

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., . . . Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature, 529*(7587), 484-489. https://doi.org/10.1038/nature16961

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., . . . Hassabis, D. (2017). Mastering the game of Go without human knowledge. *Nature, 550*(7676), 354-359. https://doi.org/10.1038/nature24270

Simanova, I., van Gerven, M., Oostenveld, R., & Hagoort, P. (2010). Identifying object categories from event-related EEG: toward decoding of conceptual representations. *PLOS ONE, 5*(12), e14465. https://doi.org/10.1371/journal.pone.0014465

Sinha, H., Awasthi, V., & Ajmera, P. K. (2020). Audio classification using braided convolutional neural networks. *IET Signal Processing, 14*(7), 448-454. https://doi.org/10.1049/iet-spr.2019.0381

Sjöström, J., & Gerstner, W. (2010). Spike-timing dependent plasticity. *Scholarpedia, 5*(2), 1362. https://doi.org/10.4249/scholarpedia.1362

Smith, A. T., Singh, K. D., Williams, A. L., & Greenlee, M. W. (2001). Estimating Receptive Field Size from fMRI Data in Human Striate and Extrastriate Visual Cortex. *Cerebral Cortex, 11*(12), 1182-1190. https://doi.org/10.1093/cercor/11.12.1182

Solomon, S. G., & Lennie, P. (2007). The machinery of colour vision. *Nature Reviews Neuroscience, 8*(4), 276-286. https://doi.org/10.1038/nrn2094

Spille, C., Kollmeier, B., & Meyer, B. T. (2017). Combining Binaural and Cortical Features for Robust Speech Recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25*(4), 756-767. https://doi.org/10.1109/TASLP.2017.2661712

Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. S. (2006). Converging Language Streams in the Human Temporal Lobe. *The Journal of Neuroscience, 26*(28), 7328-7336. https://doi.org/10.1523/jneurosci.0559-06.2006

Spoendlin, H., & Schrott, A. (1989). Analysis of the human auditory nerve. *Hearing Research, 43*(1), 25-38. https://doi.org/10.1016/0378-5955(89)90056-7

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science, 253*(5026), 1380-1386. https://doi.org/10.1126/science.1896849

Stakhovskaya, O., Sridhar, D., Bonham, B. H., & Leake, P. A. (2007). Frequency Map for the Human Cochlear Spiral Ganglion: Implications for Cochlear Implants. *Journal of the Association for Research in Otolaryngology, 8*(2), 220-233. https://doi.org/10.1007/s10162-007-0076-9

Stefanucci, J. K., Gagnon, K. T., & Lessard, D. A. (2011). Follow Your Heart: Emotion Adaptively Influences Perception. *Social and Personality Psychology Compass, 5*(6), 296-308. https://doi.org/10.1111/j.1751-9004.2011.00352.x

Sterling, P., Freed, M. A., & Smith, R. G. (1988). Architecture of rod and cone circuits to the on-beta ganglion cell. *The Journal of Neuroscience, 8*(2), 623-642. https://doi.org/10.1523/jneurosci.08-02-00623.1988

Storck, J., Jäkel, F., & Deco, G. (2001). Temporal clustering with spiking neurons and dynamic synapses: towards technological applications. *Neural Networks, 14*(3), 275-285. https://doi.org/10.1016/S0893-6080(00)00101-5

Stoughton, C. M., & Conway, B. R. (2008). Neural basis for unique hues. *Current Biology, 18*(16), R698-R699. https://doi.org/10.1016/j.cub.2008.06.018

Stowell, D., Benetos, E., & Gill, L. F. (2017). On-Bird Sound Recordings: Automatic Acoustic Recognition of Activities and Contexts. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25*(6), 1193-1206. https://doi.org/10.1109/TASLP.2017.2690565

Strauss, O. (2005). The Retinal Pigment Epithelium in Visual Function. *Physiological Reviews, 85*(3), 845-881. https://doi.org/10.1152/physrev.00021.2004

Sudre, G., Pomerleau, D., Palatucci, M., Wehbe, L., Fyshe, A., Salmelin, R., & Mitchell, T. M. (2012). Tracking neural coding of perceptual and semantic features of concrete nouns. *NeuroImage, 62*(1), 451-463. https://doi.org/10.1016/j.neuroimage.2012.04.048

Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P., & Meddis, R. (2002). A revised model of the inner-hair cell and auditory-nerve complex. *The Journal of the Acoustical Society of America, 111*(5), 2178-2188. https://doi.org/10.1121/1.1453451

Sun, F., Harwath, D., & Glass, J. (2016, 13-16 Dec). *Look, listen, and decode: Multimodal speech recognition with images*. Presented at the IEEE Spoken Language Technology Workshop (SLT), San Diego, CA. https://doi.org/10.1109/SLT.2016.7846320

Swanston, M., & Wade, N. (2013). *Visual Perception*. London, United Kingdom: Taylor and Francis. https://doi.org/10.4324/9780203082263

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging*. Stuttgart, Germany: Georg Thieme Verlag. https://doi.org/10.1017/S0022215100111879

Tan, Y., Qin, J., Xiang, X., Ma, W., Pan, W., & Xiong, N. N. (2019). A Robust Watermarking Scheme in YCbCr Color Space Based on Channel Coding. *IEEE Access, 7*, 25026-25036. https://doi.org/10.1109/ACCESS.2019.2896304

Tavanaei, A., & Maida, A. S. (2017a, 14-18 Nov). *Bio-inspired Multi-layer Spiking Neural Network Extracts Discriminative Features from Speech Signals*. Presented at the International Conference on Neural Information Processing (ICONIP), Guangzhou, China. https://doi.org/10.1007/978-3-319-70136-3_95

Tavanaei, A., & Maida, A. S. (2017b). A spiking network that learns to extract spike signatures from speech signals. *Neurocomputing, 240*, 191-199. https://doi.org/doi.org/10.1016/j.neucom.2017.01.088

Tavanaei, A., Masquelier, T., & Maida, A. S. (2016, 24-29 Jul). *Acquisition of visual features through probabilistic spike-timing-dependent plasticity*. Presented at the International Joint Conference on Neural Networks (IJCNN), Vancouver, Canada. https://doi.org/10.1109/IJCNN.2016.7727213

Taylor, P., Hobbs, J. N., Burroni, J., & Siegelmann, H. T. (2015). The global landscape of cognition: hierarchical aggregation as an organizational principle of human cortical networks and functions. *Scientific Reports, 5*, 18112. https://doi.org/10.1038/srep18112

The MathWorks Inc. (2019a). Curve Fitting Toolbox. Natick, MA. Retrieved from https://au.mathworks.com/products/curvefitting.html

The MathWorks Inc. (2019b). Signal Processing Toolbox. Natick, MA. Retrieved from https://au.mathworks.com/products/signal.html

Thermos, S., & Potamianos, G. (2016, 13-16 Dec). *Audio-visual speech activity detection in a two-speaker scenario incorporating depth information from a profile or frontal view*. Presented at the IEEE Spoken Language Technology Workshop (SLT), San Diego, CA. https://doi.org/10.1109/SLT.2016.7846321

Thongkor, K., Amornraksa, T., & Delp, E. J. (2018). Digital watermarking for camera-captured images based on just noticeable distortion and Wiener filtering. *Journal of Visual Communication and Image Representation, 53*, 146-160. https://doi.org/10.1016/j.jvcir.2018.03.005

Tick, S., Rossant, F., Ghorbel, I., Gaudric, A., Sahel, J.-A., Chaumet-Riffaud, P., & Paques, M. (2011). Foveal Shape and Structure in a Normal Population. *Investigative Ophthalmology & Visual Science, 52*(8), 5105-5110. https://doi.org/10.1167/iovs.10-7005

Tikidji-Hamburyan, R. A., Narayana, V., Bozkus, Z., & El-Ghazawi, T. A. (2017). Software for Brain Network Simulations: A Comparative Study. *Frontiers in Neuroinformatics, 11*(46). https://doi.org/10.3389/fninf.2017.00046

Tjandra, A., Sakti, S., Neubig, G., Toda, T., Adriani, M., & Nakamura, S. (2015, 19-24 Apr). *Combination of two-dimensional cochleogram and spectrogram features for deep learning-based ASR*. Presented at the IEEE International Conference on Acoustics, Speech and

Signal Processing (ICASSP), Brisbane, Australia. https://doi.org/10.1109/ICASSP.2015.7178827

Tokozume, Y., & Harada, T. (2017, 5-9 Mar). *Learning environmental sounds with end-to-end convolutional neural network*. Presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA. https://doi.org/10.1109/ICASSP.2017.7952651

Tonelli, A., Cuturi, L. F., & Gori, M. (2017). The Influence of Auditory Information on Visual Size Adaptation. *Frontiers in Neuroscience, 11*(594). https://doi.org/10.3389/fnins.2017.00594

Tu, E., Kasabov, N. K., & Yang, J. (2017). Mapping Temporal Variables Into the NeuCube for Improved Pattern Recognition, Predictive Modeling, and Understanding of Stream Data. *IEEE Transactions on Neural Networks and Learning Systems, 28*(6), 1305-1317. https://doi.org/10.1109/TNNLS.2016.2536742

Tzanetakis, G., & Cook, P. R. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing, 10*(5), 293-302. https://doi.org/10.1109/TSA.2002.800560

Tzounopoulos, T., & Leão, R. M. (2012). Mechanisms of Memory and Learning in the Auditory System. In L. O. Trussell, A. N. Popper, & R. R. Fay (Eds.), *Synaptic Mechanisms in the Auditory System* (pp. 203-226). New York, NY: Springer. https://doi.org/10.1007/978-1-4419-9517-9_9

Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Current Opinion in Neurobiology, 4*(2), 157-165. https://doi.org/10.1016/0959-4388(94)90066-3

Vanarse, A., Espinosa-Ramos, J. I., Osseiran, A., Rassau, A., & Kasabov, N. K. (2020). Application of a Brain-Inspired Spiking Neural Network Architecture to Odor Data Classification. *Sensors, 20*(10), 2756. https://doi.org/10.3390/s20102756

Vance, P. J., Das, G. P., Kerr, D., Coleman, S., McGinnity, T. M., Gollisch, T., & Liu, J. K. (2018). Bioinspired Approach to Modeling Retinal Ganglion Cells Using System Identification Techniques. *IEEE Transactions on Neural Networks and Learning Systems, 29*(5), 1796-1808. https://doi.org/10.1109/TNNLS.2017.2690139

Vartiainen, J., Parviainen, T., & Salmelin, R. (2009). Spatiotemporal Convergence of Semantic Processing in Reading and Speech Perception. *The Journal of Neuroscience, 29*(29), 9271-9280. https://doi.org/10.1523/jneurosci.5860-08.2009

Västfjäll, D. (2002). Influences of Current Mood and Noise Sensitivity on Judgments of Noise Annoyance. *The Journal of Psychology, 136*(4), 357-370. https://doi.org/10.1080/00223980209604163

Vergin, R., O'Shaughnessy, D., & Farhat, A. (1999). Generalized mel frequency cepstral coefficients for large-vocabulary speaker-independent continuous-speech recognition. *IEEE Transactions on Speech and Audio Processing, 7*(5), 525-532. https://doi.org/10.1109/89.784104

Verhulst, S., Dau, T., & Shera, C. A. (2012). Nonlinear time-domain cochlear model for transient stimulation and human otoacoustic emission. *The Journal of the Acoustical Society of America, 132*(6), 3842-3848. https://doi.org/10.1121/1.4763989

Verstraeten, D., Schrauwen, B., Stroobandt, D., & Van Campenhout, J. (2005). Isolated word recognition with the Liquid State Machine: a case study. *Information Processing Letters, 95*(6), 521-528. https://doi.org/10.1016/j.ipl.2005.05.019

Vetter, P., Smith, Fraser W., & Muckli, L. (2014). Decoding Sound and Imagery Content in Early Visual Cortex. *Current Biology, 24*(11), 1256-1262. https://doi.org/10.1016/j.cub.2014.04.020

von Helmholtz, H. (1852). LXXXI. On the theory of compound colours. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 4*(28), 519-534. https://doi.org/10.1080/14786445208647175 von Helmholtz, H. (1867). Handbuch der Physiologischen Optik [Handbook of Physiological Optics]. *Allgemeine Encyclopädie der Physik, 9*. Retrieved from https://archive.org/details/handbuchderphysi00helm

von Kriegstein, K., Dogan, Ö., Grüter, M., Giraud, A.-L., Kell, C. A., Grüter, T., . . . Kiebel, S. J. (2008). Simulation of talking faces in the human brain improves auditory speech recognition. *Proceedings of the National Academy of Sciences, 105*(18), 6747-6752. https://doi.org/10.1073/pnas.0710826105

von Kriegstein, K., Patterson, R. D., & Griffiths, T. D. (2008). Task-Dependent Modulation of Medial Geniculate Body Is Behaviorally Relevant for Speech Recognition. *Current Biology, 18*(23), 1855-1859. https://doi.org/10.1016/j.cub.2008.10.052

Voutsas, K., & Adamy, J. (2007). A Biologically Inspired Spiking Neural Network for Sound Source Lateralization. *IEEE Transactions on Neural Networks, 18*(6), 1785-1799. https://doi.org/10.1109/TNN.2007.899623

Voutsas, K., Langner, G., Adamy, J., & Ochse, M. (2005). A brain-like neural network for periodicity analysis. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 35*(1), 12-22. https://doi.org/10.1109/TSMCB.2004.837751

Vroomen, J., & Gelder, B. d. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance, 26*(5), 1583-1590. https://doi.org/10.1037/0096-1523.26.5.1583

Wade, J. J., McDaid, L. J., Santos, J. A., & Sayers, H. M. (2010). SWAT: A Spiking Neural Network Training Algorithm for Classification Problems. *IEEE Transactions on Neural Networks, 21*(11), 1817-1830. https://doi.org/10.1109/TNN.2010.2074212

Wagner-Schuman, M., Dubis, A. M., Nordgren, R. N., Lei, Y., Odell, D., Chiao, H., . . . Carroll, J. (2011). Race- and Sex-Related Differences in Retinal Thickness and Foveal Pit Morphology. *Investigative Ophthalmology & Visual Science, 52*(1), 625-634. https://doi.org/10.1167/iovs.10-5886

Wall, J. A., McDaid, L. J., Maguire, L. P., & McGinnity, T. M. (2012). Spiking Neural Network Model of Sound Localization Using the Interaural Intensity Difference. *IEEE Transactions on Neural Networks and Learning Systems, 23*(4), 574-586. https://doi.org/10.1109/TNNLS.2011.2178317

Wan, V., & Campbell, W. M. (2000, 11-13 Dec). *Support vector machines for speaker verification and identification*. Presented at the IEEE Signal Processing Society Workshop Neural Networks for Signal Processing, Sydney, Australia. https://doi.org/10.1109/NNSP.2000.890157

Wandell, B. A. (1999). Computational Neuroimaging of Human Visual Cortex. *Annual Review of Neuroscience, 22*(1), 145-173. https://doi.org/10.1146/annurev.neuro.22.1.145

Wandell, B. A., & Winawer, J. (2011). Imaging retinotopic maps in the human brain. *Vision Research, 51*(7), 718-737. https://doi.org/10.1016/j.visres.2010.08.004

Wandell, B. A., & Winawer, J. (2015). Computational neuroimaging and population receptive fields. *Trends in Cognitive Sciences, 19*(6), 349-357. https://doi.org/10.1016/j.tics.2015.03.009

Wang, C.-Y., Tai, T.-C., Wang, J.-C., Santoso, A., Mathulaprangsan, S., Chiang, C.-C., & Wu, C.-H. (2020). Sound Events Recognition and Retrieval Using Multi-Convolutional-Channel Sparse Coding Convolutional Neural Networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28*, 1875-1887. https://doi.org/10.1109/TASLP.2020.2964959

Wang, J., Wan, W. B., Li, X. X., Sun, J. D., & Zhang, H. X. (2020). Color image watermarking based on orientation diversity and color complexity. *Expert Systems with Applications, 140*, 112868. https://doi.org/10.1016/j.eswa.2019.112868

Wang, L., Mruczek, R. E. B., Arcaro, M. J., & Kastner, S. (2015). Probabilistic Maps of Visual Topography in Human Cortex. *Cerebral Cortex, 25*(10), 3911-3931. https://doi.org/10.1093/cercor/bhu277

Warden, P. (2018). Speech commands: A dataset for limited-vocabulary speech recognition. *arXiv*. Retrieved from https://arxiv.org/abs/1804.03209
Wässle, H., Grünert, U., Röhrenbeck, J., & Boycott, B. B. (1989). Cortical magnification factor and the ganglion cell density of the primate retina. *Nature, 341*(6243), 643-646. https://doi.org/10.1038/341643a0
Wegel, R. L., & Lane, C. E. (1924). The Auditory Masking of One Pure Tone by Another and its Probable Relation to the Dynamics of the Inner Ear. *Physical Review, 23*(2), 266-285. https://doi.org/10.1103/PhysRev.23.266
Weisz, C., Glowatzki, E., & Fuchs, P. (2009). The postsynaptic function of type II cochlear afferents. *Nature, 461*, 1126. https://doi.org/10.1038/nature08487
Weyand, T. G. (2016). The multifunctional lateral geniculate nucleus. *Reviews in the Neurosciences, 27*(2), 135-157. https://doi.org/10.1515/revneuro-2015-0018
Wilkinson-Berka, J. L. (2004). Diabetes and retinal vascular disorders: role of the renin-angiotensin system. *Expert Reviews in Molecular Medicine, 6*(15), 1-18. https://doi.org/10.1017/S1462399404008129
Willett, S. M., Groh, J. M., & Maddox, R. K. (2019). Hearing in a "Moving" Visual World: Coordinate Transformations Along the Auditory Pathway. In A. K. C. Lee, M. T. Wallace, A. B. Coffin, A. N. Popper, & R. R. Fay (Eds.), *Multisensory Processes: The Auditory Perspective* (pp. 85-104). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-030-10461-0_5
Wilpon, J. G., Rabiner, L. R., Lee, C.-H., & Goldman, E. R. (1990). Automatic recognition of keywords in unconstrained speech using hidden Markov models. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 38*(11), 1870-1878. https://doi.org/10.1109/29.103088
Wilson, S. M., Bautista, A., & McCarron, A. (2018). Convergence of spoken and written language processing in the superior temporal sulcus. *NeuroImage, 171*, 62-74. https://doi.org/10.1016/j.neuroimage.2017.12.068
Winawer, J., & Horiguchi, H. (2015). *Population Receptive Field Size Estimates in 3 Human Retinotopic Maps*. Retrieved on 24 Oct 2020 from https://archive.nyu.edu/handle/2451/33887
Witten, I. B., & Knudsen, E. I. (2005). Why Seeing Is Believing: Merging Auditory and Visual Worlds. *Neuron, 48*(3), 489-496. https://doi.org/10.1016/j.neuron.2005.10.020
Wohrer, A., & Kornprobst, P. (2009). Virtual Retina: A biological retina model and simulator, with contrast gain control. *Journal of Computational Neuroscience, 26*(2), 219-249. https://doi.org/10.1007/s10827-008-0108-4
Wolken, J. J. (1975). *Photoprocesses, photoreceptors, and evolution*. New York, NY: Academic Press. https://doi.org/10.1016/C2013-0-07649-9
Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation, 1*(1), 67-82. https://doi.org/10.1109/4235.585893
Wright, A., Davis, A., Bredberg, G., Ulehlova, L., & Spencer, H. (1987). Hair cell distributions in the normal human cochlea. *Acta oto-laryngologica. Supplementum, 444*, 1-48. https://doi.org/10.3109/00016488709098971
Wu, J., Chua, Y., & Li, H. (2018, 8-13 Jul). *A Biologically Plausible Speech Recognition Framework Based on Spiking Neural Networks*. Presented at the International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil. https://doi.org/10.1109/IJCNN.2018.8489535
Wu, J., Chua, Y., Zhang, M., Li, H., & Tan, K. C. (2018). A Spiking Neural Network Framework for Robust Sound Classification. *Frontiers in Neuroscience, 12*(836). https://doi.org/10.3389/fnins.2018.00836
Wu, J., Yan, T., Zhang, Z., Jin, F., & Guo, Q. (2012). Retinotopic mapping of the peripheral visual field to human visual cortex by functional magnetic resonance imaging. *Human Brain Mapping, 33*(7), 1727-1740. https://doi.org/10.1002/hbm.21324

Wu, J., Yılmaz, E., Zhang, M., Li, H., & Tan, K. C. (2020). Deep Spiking Neural Networks for Large Vocabulary Automatic Speech Recognition. *Frontiers in Neuroscience, 14*(199). https://doi.org/10.3389/fnins.2020.00199

Wysoski, S. G., Benuskova, L., & Kasabov, N. K. (2010). Evolving spiking neural networks for audiovisual information processing. *Neural Networks, 23*(7), 819-835. https://doi.org/10.1016/j.neunet.2010.04.009

Xiao, F., & Weibei, D. (2016, 24-29 Jul). *A biologically plausible spiking model for interaural level difference processing auditory pathway in human brain*. Presented at the International Joint Conference on Neural Networks (IJCNN), Vancouver, Canada. https://doi.org/10.1109/IJCNN.2016.7727862

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017, 21-26 Jul). *Aggregated Residual Transformations for Deep Neural Networks*. Presented at the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI. https://doi.org/10.1109/CVPR.2017.634

Yang, K., Li, R., Qiao, P., Wang, Q., Li, D., & Dou, Y. (2018, 7-10 Oct). *Temporal Pyramid Relation Network for Video-Based Gesture Recognition*. Presented at the IEEE International Conference on Image Processing (ICIP), Athens, Greece. https://doi.org/10.1109/ICIP.2018.8451700

Yang, M., Chien, C.-H., Delbrück, T., & Liu, S.-C. (2016, 31 Jan-4 Feb). *A 0.5V 55 μW 64 × 2-channel binaural silicon cochlea for event-driven stereo-audio sensing*. Presented at the IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA. https://doi.org/10.1109/ISSCC.2016.7418070

Yang, Y., Wang, J., Bailer, C., Cherkassky, V., & Just, M. A. (2016). Commonality of neural representations of sentences across languages: Predicting brain activation during Portuguese sentence comprehension using an English-based model of brain function. *NeuroImage, 146*, 658-666. https://doi.org/10.1016/j.neuroimage.2016.10.029

Yin, P., Qiao, H., Wu, W., Qi, L., Li, Y., Zhong, S., & Zhang, B. (2018). A Novel Biologically Inspired Visual Cognition Model: Automatic Extraction of Semantics, Formation of Integrated Concepts, and Reselection Features for Ambiguity. *IEEE Transactions on Cognitive and Developmental Systems, 10*(2), 420-431. https://doi.org/10.1109/TCDS.2017.2749978

Yu, B., Luo, Z., Wu, H., & Li, S. (2020). Hand gesture recognition based on attentive feature fusion. *Concurrency and Computation: Practice and Experience, 32*(22), e5910. https://doi.org/10.1002/cpe.5910

Yu, Q., Tang, H., Tan, K. C., & Li, H. (2013). Rapid Feedforward Computation by Temporal Encoding and Learning With Spiking Neurons. *IEEE Transactions on Neural Networks and Learning Systems, 24*(10), 1539-1552. https://doi.org/10.1109/TNNLS.2013.2245677

Yu, X., Dou, Y., & Li, J. (2018, 19-21 May). *Grayscale difference sensitivity of human eyes on LCD screen*. Presented at the Third International Workshop on Pattern Recognition, Jinan, China. Retrieved from https://doi.org/10.1117/12.2501921

Zaidi, Q., & Conway, B. R. (2019). Steps towards neural decoding of colors. *Current Opinion in Behavioral Sciences, 30*, 169-177. https://doi.org/10.1016/j.cobeha.2019.10.011

Zeddies, D. G., & Siegel, J. H. (2004). A biophysical model of an inner hair cell. *The Journal of the Acoustical Society of America, 116*(1), 426-441. https://doi.org/10.1121/1.1755237

Zeng, H., & Sanes, J. R. (2017). Neuronal cell-type classification: challenges, opportunities and the path forward. *Nature Reviews Neuroscience, 18*(9), 530-546. https://doi.org/10.1038/nrn.2017.85

Zeng, Y., Zhang, T., & Xu, B. (2017). Improving multi-layer spiking neural networks by incorporating brain-inspired rules. *Science China Information Sciences, 60*(5), 052201. https://doi.org/10.1007/s11432-016-0439-4

Zhang, H., McLoughlin, I., & Song, Y. (2015, 19-24 Apr). *Robust sound event recognition using convolutional neural networks*. Presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, Australia. https://doi.org/10.1109/ICASSP.2015.7178031

Zhang, J., Shen, F., Xu, X., & Shen, H. T. (2020). Temporal Reasoning Graph for Activity Recognition. *IEEE Transactions on Image Processing, 29*, 5491-5506. https://doi.org/10.1109/TIP.2020.2985219

Zhang, J., Zhou, W., Xie, C., Pu, J., & Li, H. (2016, 11-15 Jul). *Chinese sign language recognition with adaptive HMM*. Presented at the IEEE International Conference on Multimedia and Expo (ICME), Seattle, WA. https://doi.org/10.1109/ICME.2016.7552950

Zhang, J. X., Brooks, S., & Whalley, J. L. (2009, 28 Jun-3 Jul). *Audio classification based on adaptive partitioning*. Presented at the IEEE International Conference on Multimedia and Expo (ICME), Cancun, Mexiko. https://doi.org/10.1109/ICME.2009.5202541

Zhang, S., Zhang, S., Huang, T., Gao, W., & Tian, Q. (2018). Learning Affective Features With a Hybrid Deep Model for Audio-Visual Emotion Recognition. *IEEE Transactions on Circuits and Systems for Video Technology, 28*(10), 3030-3043. https://doi.org/10.1109/TCSVT.2017.2719043

Zhang, Y., Cao, C., Cheng, J., & Lu, H. (2018). EgoGesture: A new dataset and benchmark for egocentric hand gesture recognition. *IEEE Transactions on Multimedia, 20*(5), 1038-1050. https://doi.org/10.1109/TMM.2018.2808769

Zhang, Y., Li, P., Jin, Y., & Choe, Y. (2015). A Digital Liquid State Machine With Biologically Inspired Learning and Its Application to Speech Recognition. *IEEE Transactions on Neural Networks and Learning Systems, 26*(11), 2635-2649. https://doi.org/10.1109/TNNLS.2015.2388544

Zhang, Y., Shi, L., Wu, Y., Cheng, K., Cheng, J., & Lu, H. (2020). Gesture recognition based on deep deformable 3D convolutional neural networks. *Pattern Recognition, 107*, 107416. https://doi.org/10.1016/j.patcog.2020.107416

Zheng, F., Zhang, G., & Song, Z. (2001). Comparison of different implementations of MFCC. *Journal of Computer Science and Technology, 16*(6), 582-589. https://doi.org/10.1007/BF02943243

Zhou, B., Andonian, A., Oliva, A., & Torralba, A. (2018, 8-14 Sep). *Temporal Relational Reasoning in Videos*. Presented at the European Conference on Computer Vision (ECCV), Munich, Germany. https://doi.org/10.1007/978-3-030-01246-5_49

Zilany, M. S. A., Bruce, I. C., & Carney, L. H. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. *The Journal of the Acoustical Society of America, 135*(1), 283-286. https://doi.org/10.1121/1.4837815

Zilany, M. S. A., Bruce, I. C., Nelson, P. C., & Carney, L. H. (2009). A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics. *The Journal of the Acoustical Society of America, 126*(5), 2390-2412. https://doi.org/10.1121/1.3238250

GLOSSARY

This glossary contains, in alphabetical order, all key terms and abbreviations that were introduced throughout the thesis document. The table also includes a short description and the page number where the term was first used in the document.

Term	Description	Page
AER	Address Event Representation. A communication protocol commonly used in neuromorphic systems. Processing units with unique addresses notify each other of events they observed.	86
ANN	Artificial Neural Network. A computational method for solving high-dimensional problems through unsupervised learning of connection weights between multiple spatially distributed neurons.	101
Auditory nerve fibres	Form the cochlear nerve and transmit electrical signals created from sound to the brain.	59
Basilar membrane	A bone-like structure in the cochlea that vibrates with sound waves.	55
Bipolar cell	A neuron in the retina transmitting signals from photoreceptors to ganglion cells.	70
cGMP	Cyclic guanosine monophosphate. A molecule that causes photoreceptors to stay in a depolarised resting state in the absence of light.	69
CIE	International Commission on Illumination.	143
CNN	Convolutional Neural Network. A type of ANN that consists of several layers that perform data convolution and pooling.	84
Cochlea	The spiral-shaped hearing organ.	55
Cochlear nucleus	A processing station at the beginning of the primary auditory pathway.	60

Corti	An organ that is responsible for transforming vibrations from sound waves into electrochemical signals.	58
deSNN	Dynamic evolving SNN. An algorithm that learns the connectivity of an SNN by connecting its connection weights to single output neurons and then classifying these.	102
DVS	Dynamic vision sensor. An event-based, biologically inspired neuromorphic camera system that detects changes in light intensity in dynamic visual scenes and transforms them into spikes.	89
EEG	Electroencephalography. A technique that captures electrical activity in the brain from electrodes placed on the scalp.	33
Extrastriate cortex	Higher visual processing areas specialised in specific aspects of vision.	72
fMRI	Functional magnetic resonance imaging. A technique that captures and visualises brain activity by measuring blood flow.	49
Fovea	The area in the centre of the retina with the highest visual resolution.	67
FSDD	Free spoken digits dataset. The dataset that was used for evaluating the sound processing model.	175
Ganglion cell	A neuron in the retina transmitting visual information into further processing areas of the primary visual pathway.	70
Hair cell	Sensory receptors in the organ of Corti.	58
HCP	Human Connectome Project.	154
HMM	Hidden Markov Model. A stochastic model based on Markov chains in which the states are hidden but their emissions can be observed.	84

Horizontal cell	A neuron in the retina supporting visual feature detection by bipolar cells.	70
Inferior colliculi	A processing station of the primary auditory pathway that is involved in multi-modal stimulus integration.	61
Lateral geniculate nucleus	A processing station of the primary visual pathway forwarding signals from the ganglion cells to the primary visual cortex.	72
LIF	Leaky Integrate-and-Fire. A type of neuron that collects incoming signals until it reaches a threshold and fires a signal as well.	106
JND	Just noticeable difference. The minimum required difference to distinguish between two colours	146
Medial geniculate body	A processing station of the primary auditory pathway involved in integrating visual cues, emotional responses, and somatosensory input.	62
MFCC	Mel-frequency cepstral coefficients. Method to extract perceptual features from sound data.	84
NeuCube	The spiking neural network architecture used in this research.	101
NZSL	New Zealand Sign Language. The primary language developed and used by the New Zealand Deaf community.	224
Optic chiasm	The cross-over point of visual signals in the brain.	71
Photoreceptor	A cell in the retina that is capable of phototransduction.	68
Phototransduction	The process of converting light rays into electrical signals.	69
Plasticity	The ability of the brain to adapt to change.	64
Primary auditory cortex	The central processing unit for sound signals in the brain.	62

Primary auditory pathway	The pathway of the brain signals from the cochleae to the auditory cortices.	59
Primary visual cortex	The central processing unit for light signals in the brain.	72
Primary visual pathway	The pathway of the brain signals from the retinae to the visual cortex.	71
Retina	A multi-cell layer at the back of the eye that enables vision.	67
Retinotopy	The mapping of locations of stimuli in the visual field to neural processing areas in the primary visual cortex.	152
RGB	Red, green, blue. A colour format commonly used in computational systems.	143
STDP	Spike-timing-dependent plasticity. An algorithm that learns patterns in spike data by adjusting connection weights of pre- and post-synaptic neurons based on their firing times.	102
SNN	Spiking Neural Network An artificial neural network in which the neurons communicate by sending electrical impulses (spikes).	101
Striate cortex	An alternative name for the primary visual cortex, based on its striped (striate) appearance.	72
Superior olivary complex	A processing unit of the primary auditory pathway that is mainly responsible for sound localisation.	60
SVM	Support Vector Machine. A supervised machine learning model that can separate data points by finding an optimum hyperplane in higher-dimensional space.	84
SWC	Small-world connectivity. The principle that during the initialisation of the network, neurons are only connected to neighbouring neurons within a defined small radius.	107

Tonotopy	The mapping of frequencies of sound stimuli to neural processing areas in the primary auditory cortex.	125
V1	Primary visual cortex.	72
VRF	Visual Receptive Field. A group of retinal neurons that process stimuli from an area of the visual field in union.	157

APPENDICES

APPENDIX A SOURCE CODE

This appendix contains all code that was referenced throughout the thesis document, however, it does not contain all code that was written for this research. All source code and data that were created as an original piece of work for this PhD project can be found on GitHub at https://github.com/AnneWendt/PhD-thesis.

LISTING I RESAMPLE_SOUND_DATA.M

```
% resample the spoken words corpus from 8kHz to 100kHz for
  cochlea.py
source_folder = free-spoken-digit-dataset-master\recordings\';
target_folder = free-spoken-digit-dataset-master\upsampled\';

source_fs = 8000;
target_fs = 100000;

% calculate the two values that are needed for MATLAB's resample
  function
[p, q] = rat(target_fs/source_fs);

% get all the filenames
info = dir(source_folder);
info = info(3:end); % the first two entries are . and .. pointing
                      to itself and the parent directory

for file_id = 1:length(info)
    filename = [source_folder info(file_id).name];
    [sound, fs] = audioread(filename);

    assert(fs == source_fs);
    new_sound = resample(sound, p, q);

    filename = [target_folder info(file_id).name];
    audiowrite(filename, new_sound, target_fs);
end
```

LISTING II CONVERT_JACKSON_SAMPLES_TO_CSV.PY

```python
### This script takes the speech wav files and creates NeuCube-compatible spike matrices

import cochlea
import matplotlib.pyplot as plt
import numpy as np
import thorns as th
import thorns.waves as wv
from scipy.io import wavfile
import sys

def scale_down(spike_matrix, left_cf, right_cf, left_target_cf, right_target_cf, scaling_factor):

    # we scale down vertically (time) and horizontally (frequencies) using the target frequencies as cut-off values

    number_of_short_samples = int(len(spike_matrix)/scaling_factor)
    short_spike_matrix = np.zeros((number_of_short_samples, len(left_target_cf) + len(right_target_cf) - 2), dtype=int)

    left_threshold = 2 * np.mean(spike_matrix[:, :num_freq])
    right_threshold = 2 * np.mean(spike_matrix[:, num_freq:])

    # find cut-off indices in the spike matrix based on the frequency values of the target cf
    left_freq_indices = np.zeros(len(left_target_cf), dtype=int)
    right_freq_indices = np.zeros(len(right_target_cf), dtype=int)

    for idx, freq in enumerate(left_target_cf):
        left_freq_indices[idx] = np.where(left_cf >= freq)[0][0]

    for idx, freq in enumerate(right_target_cf):
        right_freq_indices[idx] = np.where(right_cf >= freq)[0][0]

    # now use these indices to compress the data
    for short_sam_num in xrange(number_of_short_samples):

        start_time = short_sam_num * scaling_factor
        stop_time = start_time + scaling_factor

        for freq in xrange(len(left_target_cf) - 1):

            start_freq = left_freq_indices[freq]
            stop_freq = left_freq_indices[freq + 1]

            if np.mean(spike_matrix[start_time:stop_time, start_freq:stop_freq]) > left_threshold:
                short_spike_matrix[short_sam_num, freq] = 1

        for freq in xrange(len(right_target_cf) - 1):

            start_freq = right_freq_indices[freq]
            stop_freq = right_freq_indices[freq + 1]

            if np.mean(spike_matrix[start_time:stop_time, start_freq:stop_freq]) > right_threshold:
                short_spike_matrix[short_sam_num, freq + len(left_target_cf) - 1] = 1

    return short_spike_matrix
```

```
### set model parameters ###

source_folder = free-spoken-digit-dataset-master\upsampled\'
target_folder = free-spoken-digit-dataset-master\samples\'

speakers = ['jackson', 'nicolas', 'theo', 'yweveler']
sound_frequency = 100e3
scaling_factor = 100  # the spike rate is quite low and the files are quite large, so we will sum every 100 time points together

min_freq = 125  # lowest in cochlea.py
max_freq = 8001  # highest in human hearing would be 20k but I only work with 8k because that is what I have Langers data for
num_freq = 3500  # based on Wright et al. 1987
cf = (min_freq, max_freq, num_freq)

anf_num = (5, 2, 1)  # number of auditory nerve fibres with high/medium/low spontaneous spike rate - this is per CF! based on Liberman 1978 and 30,000 Type I fibres for 3,500 CFs (Spoendlin & Schrott 1989)
species = 'human'  # could also be cat
seed = 0  # this is a random seed parameter

# this is to keep track of the NeuCube sample id
sample_id = 0

### this is the big loop that does everything - per speaker, per digit, per recording
print "Starting program..."

for speaker in speakers:
    print "Reading", speaker, "..."

    # we have ten digits from 0 to 9
    for digit in xrange(10):
        sys.stdout.write(str(digit) + '...')

        # every speaker recorded every digit 50 times
        for speaker_sample in xrange(50):

            ### load data and make sure it is in the right format
            fs, sound = wavfile.read(source_folder + str(digit) + '_' + speaker + '_' + str(speaker_sample) + '.wav')

            assert fs == sound_frequency, "Frequency should be 100 kHz for cochlea module"
            assert sound.dtype == 'int16', "Data type should be int16"
            assert sound.ndim == 1 or sound.ndim == 2, "Sound file must have one or two channels"

            #normalise data (required for cochlea.py)
            sound = (sound / 2.**15)

            #if we only have one channel, duplicate to make it stereo, otherwise extract channels
            if sound.ndim == 1:
                left = sound
                right = left
            else:
                left = sound[:,0]
                right = sound[:,1]
```

```
            ### create spike trains ###

            # output is a Pandas Data Frame with headings ['cf', 'duration', 'spikes', 'type']
            # 'cf' is the exact frequency
            # 'duration' is the length of the sound sample
            # 'spikes' is an array of the exact spike times for that particular frequency
            # 'type' is hsr, msr, or lsr depending on the settings in anf_num

            left_trains = cochlea.run_zilany2014(
                sound=left,
                fs=sound_frequency,
                anf_num=anf_num,
                cf=cf,
                species=species,
                seed=seed
            )

            right_trains = cochlea.run_zilany2014(
                sound=right,
                fs=sound_frequency,
                anf_num=anf_num,
                cf=cf,
                species=species,
                seed=seed
            )

            # put the signals from the different ANFs together so that we do not have so much data
            left_trains = th.accumulate(left_trains, ignore=['type'])
            right_trains = th.accumulate(right_trains,ignore=['type'])

        # the spikes are saved as time points, e.g., [0.123, 0.345, 0.567]
        # the values are between 0 and the sample length
        left_spikes = left_trains['spikes'] * sound_frequency
        right_spikes = right_trains['spikes'] * sound_frequency

        # create a matrix of zeros and replace all spike times with 1
        # dimension is the length of the sample and 2x columns (channels for left plus channels for right)
        # since all sound files have different length, we need to set the array size dynamically
        sample_length = len(left)
        spike_matrix = np.zeros((sample_length, num_freq * 2), dtype=int)

        # unpack the spike trains by frequencies
        for col_id in xrange(num_freq):
            # convert the spike times of this channel into integers that we can use as index for the spike matrix
            row_ids = (np.fix(left_spikes[col_id])).astype(int)
            spike_matrix[row_ids, col_id] = 1

        for col_id in xrange(num_freq):
            # and now for the right side as well - remember the offset to change the second half of the channels!
            row_ids = (np.fix(right_spikes[col_id])).astype(int)
            spike_matrix[row_ids, col_id + num_freq] = 1

        sample_id +=1
```

```
    # get actual values for cf
    left_cf = np.array(left_trains['cf'])
    right_cf = np.array(right_trains['cf'])

    #scale down the data to different sizes and save it

    #MNI_times_2
    left_target_cf = np.array([166, 240, 346, 355, 368, 374,
        390, 397, 400, 402, 409, 417, 421, 423, 425, 428, 431,
        434, 439, 442, 444, 448, 451, 455, 457, 459, 463, 464,
        468, 470, 471, 475, 479, 483, 486, 489, 491, 493, 495,
        497, 498, 499, 501, 501, 503, 504, 505, 505, 507, 508,
        509, 510, 511, 512, 513, 514, 516, 516, 517, 518, 518,
        519, 520, 520, 521, 521, 522, 523, 523, 524, 525, 526,
        526, 527, 528, 529, 530, 531, 531, 532, 532, 533, 534,
        534, 535, 536, 537, 538, 538, 539, 540, 540, 541, 541,
        542, 543, 543, 544, 544, 545, 546, 546, 547, 547, 548,
        549, 550, 550, 551, 551, 552, 552, 553, 553, 554, 554,
        555, 557, 557, 558, 559, 560, 560, 561, 561, 562, 562,
        563, 563, 564, 564, 565, 566, 566, 566, 567, 567, 568,
        569, 569, 570, 570, 570, 571, 572, 572, 573, 573, 573,
        574, 574, 575, 575, 576, 576, 577, 577, 577, 578, 578,
        578, 579, 579, 580, 580, 581, 581, 582, 582, 582, 582,
        582, 583, 583, 583, 583, 584, 584, 585, 585, 585, 586,
        586, 586, 587, 587, 587, 588, 588, 588, 589, 589, 589,
        590, 590, 591, 591, 591, 591, 592, 592, 593, 593, 593,
        593, 594, 594, 594, 595, 595, 596, 596, 597, 597, 597,
        597, 598, 598, 598, 598, 599, 599, 599, 600, 600, 600,
        600, 601, 601, 601, 602, 602, 602, 602, 603, 603, 603,
        603, 604, 604, 604, 605, 605, 605, 605, 605, 606, 606,
        606, 607, 607, 607, 608, 608, 608, 608, 609, 609, 609,
        609, 610, 610, 610, 610, 611, 611, 611, 612, 612, 612,
        612, 612, 613, 613, 613, 613, 613, 614, 614, 614, 615,
        615, 615, 615, 616, 616, 616, 616, 617, 617, 617, 618,
        618, 618, 618, 619, 619, 619, 619, 620, 620, 620, 620,
        621, 621, 621, 621, 621, 622, 622, 622, 622, 622, 623,
        623, 623, 624, 624, 624, 624, 625, 625, 625, 625, 625,
        626, 626, 626, 626, 626, 627, 627, 627, 627, 627, 627,
        627, 628, 628, 628, 628, 628, 629, 629, 629, 629,
        629, 629, 630, 630, 630, 630, 631, 631, 631, 631,
        631, 632, 632, 632, 632, 633, 633, 633, 633, 633, 634,
        634, 634, 634, 634, 634, 634, 634, 634, 634, 635, 635,
        635, 635, 635, 635, 635, 636, 636, 636, 636, 636, 636,
        637, 637, 637, 637, 637, 638, 638, 638, 638, 638, 638,
        638, 638, 638, 639, 639, 639, 639, 639, 639, 639, 640,
        640, 640, 640, 640, 641, 641, 641, 641, 641, 641, 641,
        642, 642, 642, 642, 643, 643, 643, 643, 643, 643, 643,
        644, 644, 644, 644, 644, 645, 645, 645, 645, 645, 645,
        646, 646, 646, 646, 646, 646, 647, 647, 647, 647, 647,
        647, 647, 648, 648, 648, 648, 648, 648, 648, 648, 648,
        649, 649, 649, 649, 649, 649, 649, 649, 649, 649, 650,
        650, 650, 650, 650, 651, 651, 651, 651, 651, 651, 652,
        652, 652, 652, 652, 652, 652, 652, 653, 653, 653, 653,
        653, 653, 653, 653, 653, 654, 654, 654, 654, 654, 654,
        654, 655, 655, 655, 655, 655, 655, 655, 655, 656, 656,
        656, 656, 656, 656, 656, 657, 657, 657, 657, 657, 657,
        657, 657, 658, 658, 658, 658, 658, 658, 658, 658, 659,
        659, 659, 659, 659, 659, 659, 659, 659, 660, 660, 660,
        660, 660, 660, 660, 661, 661, 661, 661, 661, 661, 661,
        662, 662, 662, 662, 662, 662, 662, 662, 662, 662, 663,
        663, 663, 663, 663, 663, 663, 663, 663, 664, 664, 664,
        664, 664, 664, 664, 664, 665, 665, 665, 665, 665, 665,
        665, 665, 666, 666, 666, 666, 666, 666, 666, 667, 667,
```

667, 667, 667, 667, 667, 667, 668, 668, 668, 668, 668,
668, 668, 669, 669, 669, 669, 669, 669, 669, 669, 670,
670, 670, 670, 670, 670, 670, 670, 670, 670, 671, 671,
671, 671, 671, 671, 671, 671, 672, 672, 672, 672, 672,
672, 672, 672, 672, 673, 673, 673, 673, 673, 673, 674,
674, 674, 674, 674, 674, 674, 674, 674, 674, 675, 675,
675, 675, 675, 675, 675, 675, 676, 676, 676, 676, 676,
676, 676, 676, 677, 677, 677, 677, 677, 677, 678, 678,
678, 678, 678, 679, 679, 679, 679, 679, 679, 679, 679,
679, 680, 680, 680, 680, 680, 680, 680, 680, 681, 681,
681, 681, 681, 681, 682, 682, 682, 682, 682, 682, 682,
683, 683, 683, 683, 683, 684, 684, 684, 684, 684, 684,
684, 685, 685, 685, 685, 685, 685, 685, 686, 686, 686,
686, 686, 686, 686, 687, 687, 687, 687, 687, 687, 687,
688, 688, 688, 688, 688, 688, 688, 688, 688, 689, 689,
689, 689, 689, 689, 689, 689, 689, 690, 690, 690, 690,
690, 691, 691, 691, 691, 691, 692, 692, 692, 692, 692,
692, 692, 693, 693, 693, 693, 693, 693, 693, 693, 693,
694, 694, 694, 694, 694, 694, 694, 694, 694, 695, 695,
695, 695, 695, 695, 695, 696, 696, 696, 696, 696, 696,
696, 696, 697, 697, 697, 697, 697, 697, 697, 697, 697,
698, 698, 698, 699, 699, 699, 699, 699, 699, 699, 699,
700, 700, 700, 700, 700, 700, 700, 701, 701, 701, 701,
701, 701, 701, 701, 702, 702, 702, 702, 702, 702, 703,
703, 703, 703, 703, 703, 703, 703, 703, 704, 704, 704,
704, 704, 704, 704, 705, 705, 705, 705, 705, 705, 705,
706, 706, 706, 706, 706, 706, 706, 706, 707, 707, 707,
707, 707, 707, 708, 708, 708, 708, 708, 708, 708, 708,
709, 709, 709, 709, 709, 709, 710, 710, 710, 710, 710,
710, 710, 710, 711, 711, 711, 711, 711, 711, 711, 712,
712, 712, 712, 712, 712, 712, 712, 712, 713, 713, 713,
713, 713, 713, 713, 713, 714, 714, 714, 714, 714, 714,
714, 715, 715, 715, 715, 716, 716, 716, 716, 716, 716,
716, 716, 717, 717, 717, 717, 717, 717, 717, 718, 718,
718, 718, 718, 718, 718, 719, 719, 719, 719, 719, 719,
719, 720, 720, 720, 720, 720, 720, 720, 721, 721, 721,
721, 721, 722, 722, 722, 722, 722, 723, 723, 723, 723,
723, 723, 724, 724, 724, 724, 724, 724, 724, 724, 725,
725, 725, 725, 725, 726, 726, 726, 726, 726, 727, 727,
727, 727, 727, 727, 727, 727, 727, 728, 728, 728, 728,
728, 728, 728, 728, 729, 729, 729, 729, 729, 729, 730,
730, 730, 730, 730, 731, 731, 731, 731, 731, 731, 731,
732, 732, 732, 732, 732, 732, 732, 733, 733, 733, 733,
733, 733, 734, 734, 734, 734, 734, 734, 734, 734, 734,
734, 735, 735, 735, 735, 735, 735, 735, 736, 736, 736,
736, 736, 736, 736, 737, 737, 737, 737, 737, 737, 737,
738, 738, 738, 738, 738, 739, 739, 739, 739, 739, 739,
739, 739, 740, 740, 740, 740, 740, 740, 740, 741, 741,
741, 741, 741, 741, 741, 742, 742, 742, 742, 743, 743,
743, 743, 743, 743, 744, 744, 744, 744, 744, 744, 744,
744, 744, 745, 745, 745, 745, 745, 745, 745, 745, 746,
746, 746, 746, 746, 746, 747, 747, 747, 747, 747, 748,
748, 748, 748, 748, 748, 748, 749, 749, 749, 749, 749,
749, 750, 750, 750, 750, 750, 750, 750, 751, 751, 751,
751, 752, 752, 752, 752, 752, 752, 753, 753, 753, 753,
753, 754, 754, 754, 754, 754, 754, 754, 755, 755, 755,
755, 755, 755, 755, 756, 756, 756, 756, 756, 756, 756,
757, 757, 757, 757, 758, 758, 758, 758, 758, 759, 759,
759, 759, 759, 759, 759, 760, 760, 760, 760, 760, 760,
761, 761, 761, 761, 761, 761, 761, 762, 762, 762, 762,
762, 762, 763, 763, 763, 763, 763, 763, 763, 764, 764,
764, 764, 764, 764, 764, 765, 765, 765, 765, 765, 765,
765, 765, 766, 766, 766, 766, 766, 766, 766, 766, 767,
767, 767, 767, 767, 768, 768, 768, 768, 768, 768, 769,

769, 769, 769, 769, 769, 770, 770, 770, 770, 770, 771,
771, 771, 771, 771, 771, 772, 772, 772, 772, 772, 772,
773, 773, 773, 773, 773, 773, 774, 774, 774, 774, 774,
775, 775, 775, 775, 775, 775, 776, 776, 776, 776, 776,
776, 777, 777, 777, 777, 777, 777, 778, 778, 778, 778,
778, 779, 779, 779, 779, 779, 779, 779, 779, 780, 780,
780, 780, 780, 781, 781, 781, 781, 781, 781, 782, 782,
782, 782, 783, 783, 783, 783, 783, 783, 783, 783, 784,
784, 784, 784, 784, 784, 785, 785, 785, 785, 785, 785,
786, 786, 786, 786, 786, 786, 787, 787, 787, 787, 787,
787, 788, 788, 788, 788, 788, 788, 788, 789, 789, 789,
789, 789, 789, 789, 790, 790, 790, 790, 790, 791, 791,
791, 791, 791, 791, 791, 791, 792, 792, 792, 792, 792,
792, 792, 793, 793, 793, 793, 793, 794, 794, 794, 794,
794, 794, 795, 795, 795, 795, 795, 795, 795, 795, 795,
796, 796, 796, 796, 796, 796, 796, 796, 797, 797, 797,
797, 797, 797, 797, 798, 798, 798, 798, 798, 798, 798,
799, 799, 799, 799, 799, 799, 800, 800, 800, 800, 800,
800, 801, 801, 801, 801, 801, 801, 802, 802, 802, 802,
802, 802, 803, 803, 803, 803, 803, 803, 803, 804, 804,
804, 804, 804, 804, 805, 805, 805, 805, 805, 805, 805,
806, 806, 806, 806, 807, 807, 807, 807, 807, 807, 808,
808, 808, 808, 808, 808, 808, 808, 809, 809, 809, 809,
809, 809, 809, 809, 810, 810, 810, 810, 810, 810, 810,
811, 811, 811, 811, 811, 811, 811, 812, 812, 812, 812,
812, 812, 813, 813, 813, 813, 813, 813, 813, 814, 814,
814, 814, 814, 814, 815, 815, 815, 815, 815, 815, 816,
816, 816, 816, 816, 816, 817, 817, 817, 817, 817, 817,
818, 818, 818, 818, 818, 818, 819, 819, 819, 819, 820,
820, 820, 820, 820, 820, 820, 821, 821, 821, 821, 822,
822, 822, 822, 822, 823, 823, 823, 823, 823, 824, 824,
824, 824, 824, 825, 825, 825, 825, 825, 826, 826, 826,
826, 827, 827, 828, 828, 828, 829, 829, 829, 829, 830,
830, 830, 831, 831, 831, 831, 831, 832, 832, 832, 833,
833, 833, 834, 834, 834, 834, 834, 835, 835, 835, 835,
835, 835, 836, 836, 836, 836, 836, 837, 837, 837, 838,
838, 838, 838, 839, 839, 839, 840, 840, 840, 840, 841,
841, 841, 842, 842, 842, 842, 842, 843, 843, 843, 844,
844, 844, 844, 844, 845, 845, 845, 846, 846, 847, 847,
847, 847, 848, 848, 848, 848, 849, 849, 849, 849, 850,
850, 851, 851, 851, 851, 852, 852, 852, 853, 853, 853,
854, 854, 854, 854, 855, 855, 856, 856, 856, 857, 857,
857, 858, 858, 858, 858, 859, 859, 860, 860, 860, 861,
861, 862, 862, 862, 863, 863, 864, 864, 864, 864, 865,
865, 866, 866, 866, 866, 867, 867, 868, 868, 868, 869,
869, 869, 870, 870, 871, 871, 871, 872, 872, 872, 873,
873, 874, 874, 875, 875, 876, 876, 877, 877, 878, 878,
878, 879, 879, 879, 880, 880, 880, 881, 881, 881, 882,
882, 882, 883, 884, 884, 884, 885, 885, 886, 887, 887,
888, 889, 889, 890, 890, 891, 891, 892, 892, 893, 893,
894, 894, 894, 894, 895, 895, 896, 897, 897, 897, 897,
898, 898, 898, 899, 899, 900, 900, 901, 901, 902, 902,
902, 903, 903, 904, 905, 905, 906, 906, 907, 907, 908,
908, 909, 910, 910, 911, 911, 912, 913, 914, 914, 915,
915, 916, 916, 917, 918, 918, 919, 919, 920, 920, 920,
921, 922, 922, 923, 925, 925, 926, 927, 928, 928, 930,
930, 931, 931, 931, 933, 933, 934, 934, 934, 935, 936,
937, 938, 939, 940, 940, 941, 942, 943, 944, 944, 945,
945, 946, 947, 947, 948, 949, 950, 950, 951, 952, 953,
954, 954, 955, 956, 957, 957, 958, 959, 959, 960, 961,
963, 964, 965, 966, 966, 966, 967, 967, 968, 969, 970,
971, 972, 973, 974, 975, 976, 977, 977, 978, 979, 980,
981, 983, 984, 985, 987, 987, 989, 991, 991, 992, 993,
995, 997, 999, 1001, 1001, 1002, 1003, 1004, 1006, 1007,

1008, 1010, 1010, 1011, 1011, 1012, 1014, 1015, 1017,
1019, 1020, 1021, 1022, 1023, 1024, 1025, 1026, 1028,
1029, 1030, 1031, 1032, 1034, 1035, 1037, 1039, 1039,
1040, 1042, 1043, 1043, 1045, 1047, 1049, 1049, 1051,
1052, 1052, 1054, 1058, 1060, 1063, 1064, 1065, 1066,
1067, 1071, 1073, 1074, 1075, 1076, 1077, 1078, 1079,
1080, 1082, 1083, 1086, 1086, 1090, 1091, 1092, 1093,
1094, 1095, 1097, 1098, 1099, 1101, 1102, 1102, 1105,
1107, 1108, 1109, 1111, 1112, 1113, 1115, 1117, 1119,
1119, 1120, 1122, 1124, 1125, 1127, 1128, 1130, 1131,
1132, 1133, 1136, 1137, 1138, 1140, 1142, 1143, 1145,
1147, 1150, 1152, 1154, 1156, 1156, 1158, 1160, 1162,
1163, 1165, 1167, 1168, 1169, 1170, 1172, 1174, 1178,
1180, 1182, 1184, 1184, 1186, 1187, 1191, 1193, 1195,
1196, 1199, 1200, 1200, 1203, 1204, 1206, 1209, 1210,
1211, 1213, 1213, 1215, 1219, 1219, 1221, 1222, 1224,
1224, 1227, 1229, 1234, 1237, 1239, 1241, 1242, 1245,
1246, 1249, 1250, 1251, 1252, 1256, 1257, 1258, 1259,
1261, 1263, 1264, 1265, 1266, 1271, 1273, 1275, 1277,
1280, 1281, 1283, 1284, 1286, 1288, 1290, 1293, 1295,
1297, 1297, 1302, 1304, 1306, 1308, 1311, 1313, 1315,
1317, 1318, 1320, 1323, 1325, 1327, 1332, 1333, 1334,
1337, 1340, 1342, 1345, 1347, 1349, 1351, 1353, 1354,
1355, 1358, 1359, 1361, 1363, 1365, 1367, 1369, 1372,
1375, 1379, 1381, 1383, 1387, 1389, 1393, 1395, 1396,
1398, 1402, 1403, 1406, 1408, 1410, 1413, 1414, 1416,
1419, 1423, 1428, 1428, 1430, 1432, 1434, 1437, 1438,
1441, 1444, 1447, 1449, 1452, 1453, 1455, 1457, 1460,
1461, 1467, 1470, 1472, 1474, 1476, 1477, 1479, 1480,
1485, 1487, 1489, 1491, 1492, 1495, 1497, 1500, 1502,
1503, 1506, 1508, 1511, 1515, 1516, 1521, 1524, 1526,
1529, 1532, 1534, 1535, 1536, 1537, 1538, 1542, 1545,
1546, 1548, 1552, 1553, 1555, 1557, 1559, 1563, 1565,
1569, 1573, 1576, 1578, 1581, 1584, 1587, 1592, 1593,
1600, 1602, 1604, 1606, 1609, 1613, 1615, 1616, 1625,
1627, 1629, 1631, 1633, 1634, 1636, 1639, 1641, 1644,
1649, 1652, 1654, 1656, 1658, 1661, 1664, 1667, 1673,
1676, 1681, 1682, 1686, 1688, 1690, 1691, 1693, 1696,
1700, 1704, 1708, 1712, 1714, 1716, 1718, 1719, 1722,
1726, 1729, 1731, 1735, 1739, 1740, 1743, 1744, 1748,
1750, 1752, 1757, 1761, 1763, 1769, 1773, 1776, 1780,
1782, 1785, 1788, 1789, 1795, 1798, 1802, 1803, 1806,
1811, 1814, 1818, 1821, 1829, 1836, 1840, 1844, 1845,
1847, 1852, 1854, 1857, 1865, 1874, 1877, 1881, 1883,
1887, 1891, 1894, 1899, 1901, 1905, 1911, 1914, 1917,
1921, 1925, 1927, 1931, 1938, 1940, 1948, 1951, 1955,
1957, 1962, 1965, 1970, 1973, 1974, 1977, 1980, 1987,
1991, 1993, 1997, 2001, 2005, 2008, 2010, 2014, 2020,
2023, 2024, 2027, 2032, 2033, 2036, 2044, 2046, 2049,
2053, 2055, 2057, 2059, 2064, 2065, 2073, 2076, 2079,
2081, 2085, 2093, 2099, 2104, 2107, 2112, 2113, 2119,
2120, 2126, 2126, 2137, 2139, 2142, 2146, 2149, 2153,
2156, 2158, 2161, 2163, 2167, 2170, 2176, 2180, 2182,
2185, 2189, 2192, 2196, 2200, 2206, 2210, 2213, 2214,
2218, 2219, 2223, 2226, 2228, 2231, 2234, 2237, 2246,
2246, 2251, 2254, 2255, 2259, 2260, 2267, 2268, 2270,
2273, 2275, 2280, 2286, 2289, 2292, 2295, 2307, 2308,
2312, 2316, 2319, 2322, 2324, 2328, 2331, 2334, 2340,
2345, 2346, 2351, 2355, 2361, 2365, 2368, 2376, 2379,
2382, 2385, 2390, 2395, 2397, 2401, 2408, 2416, 2422,
2427, 2435, 2437, 2443, 2447, 2452, 2455, 2458, 2462,
2468, 2475, 2479, 2480, 2482, 2486, 2488, 2489, 2496,
2498, 2504, 2510, 2516, 2526, 2532, 2536, 2541, 2546,
2550, 2552, 2559, 2561, 2565, 2570, 2574, 2582, 2587,

```
            2591, 2596, 2599, 2602, 2613, 2620, 2623, 2628, 2633,
            2639, 2642, 2650, 2656, 2660, 2665, 2674, 2676, 2679,
            2689, 2693, 2696, 2701, 2707, 2715, 2717, 2720, 2731,
            2740, 2746, 2753, 2758, 2766, 2774, 2778, 2788, 2796,
            2803, 2820, 2826, 2837, 2847, 2859, 2870, 2875, 2884,
            2895, 2905, 2916, 2938, 2955, 2968, 2976, 2995, 3005,
            3018, 3022, 3026, 3040, 3045, 3055, 3073, 3093, 3116,
            3134, 3175, 3182, 3202, 3212, 3238, 3251, 3276, 3304,
            3326, 3338, 3364, 3399, 3418, 3443, 3464, 3535, 3552,
            3563, 3594, 3627, 3664, 3714, 3754, 3775, 3806, 3863,
            3895, 4014, 4055, 4115, 4135, 4326, 4379, 4493, 4655,
            4747, 4868, 4989, 5196, 5294, 5401, 5747, 6067, 6337,
            6688, 6942, 7181, 7281, 8000])

        right_target_cf = np.array([161, 236, 288, 337, 354, 364,
            370, 380, 384, 390, 391, 396, 398, 399, 402, 404, 406,
            409, 413, 419, 421, 429, 432, 438, 448, 451, 459, 462,
            465, 472, 473, 477, 480, 485, 487, 492, 494, 496, 500,
            502, 507, 511, 511, 514, 517, 519, 521, 523, 525, 527,
            528, 530, 531, 532, 534, 535, 535, 537, 538, 541, 541,
            543, 545, 547, 548, 549, 550, 550, 551, 552, 553, 554,
            555, 555, 556, 557, 558, 559, 559, 560, 560, 560, 561,
            561, 562, 562, 563, 564, 564, 564, 565, 565, 566, 566,
            566, 567, 567, 568, 569, 569, 570, 570, 570, 571, 572,
            572, 572, 573, 574, 574, 575, 575, 575, 575, 576, 576,
            576, 577, 577, 577, 578, 578, 578, 578, 579, 579, 580,
            580, 580, 580, 581, 581, 581, 581, 582, 582, 582, 582,
            583, 583, 584, 584, 584, 585, 585, 586, 586, 587, 587,
            587, 587, 588, 588, 588, 589, 589, 589, 589, 590, 590,
            590, 590, 591, 591, 592, 592, 592, 592, 593, 593, 594,
            594, 594, 594, 595, 595, 595, 596, 596, 596, 597, 597,
            597, 597, 598, 598, 598, 598, 598, 599, 599, 599, 600,
            600, 600, 600, 601, 601, 601, 602, 602, 602, 602, 602,
            603, 603, 603, 603, 603, 604, 604, 604, 604, 604, 604,
            604, 605, 605, 605, 605, 606, 606, 606, 606, 606, 606,
            607, 607, 607, 607, 607, 608, 608, 608, 608, 608, 609,
            609, 609, 609, 609, 609, 610, 610, 610, 611, 611, 611,
            611, 611, 612, 612, 612, 612, 612, 612, 613, 613, 613,
            613, 613, 614, 614, 614, 614, 614, 615, 615, 615, 615,
            616, 616, 616, 616, 616, 617, 617, 617, 617, 617, 618,
            618, 618, 618, 618, 618, 619, 619, 619, 619, 620, 620,
            620, 620, 620, 620, 621, 621, 621, 621, 622, 622, 622,
            622, 623, 623, 623, 623, 623, 624, 624, 624, 624, 624,
            625, 625, 625, 625, 625, 626, 626, 626, 626, 627, 627,
            627, 627, 627, 627, 628, 628, 628, 628, 628, 629, 629,
            629, 629, 629, 630, 630, 630, 630, 630, 631, 631, 631,
            631, 631, 632, 632, 632, 632, 632, 632, 633, 633, 633,
            633, 633, 633, 633, 634, 634, 634, 634, 634, 635, 635,
            635, 635, 636, 636, 636, 636, 636, 636, 636, 637, 637,
            637, 637, 637, 638, 638, 638, 638, 638, 639, 639, 639,
            639, 639, 639, 640, 640, 640, 641, 641, 641, 641, 642,
            642, 642, 642, 642, 643, 643, 643, 643, 644, 644, 644,
            644, 644, 644, 645, 645, 645, 645, 646, 646, 646, 646,
            646, 646, 646, 647, 647, 647, 647, 647, 648, 648, 648,
            649, 649, 649, 649, 649, 649, 650, 650, 650, 650, 650,
            650, 651, 651, 651, 651, 652, 652, 652, 652, 652, 653,
            653, 653, 653, 653, 654, 654, 654, 654, 654, 655, 655,
            655, 655, 655, 656, 656, 656, 656, 657, 657, 657, 657,
            657, 657, 658, 658, 658, 658, 659, 659, 659, 659, 659,
            659, 660, 660, 660, 660, 660, 661, 661, 661, 661, 661,
            662, 662, 662, 662, 662, 662, 663, 663, 663, 663, 663,
            664, 664, 664, 664, 664, 664, 665, 665, 665, 665, 665,
            665, 666, 666, 666, 666, 666, 666, 667, 667, 667, 667,
            667, 667, 667, 668, 668, 668, 668, 668, 668, 669, 669,
```

669, 669, 669, 669, 670, 670, 670, 670, 671, 671, 671,
671, 671, 671, 672, 672, 672, 672, 672, 673, 673, 673,
673, 673, 673, 673, 674, 674, 674, 674, 675, 675, 675,
675, 675, 675, 676, 676, 676, 676, 676, 677, 677, 677,
677, 677, 678, 678, 678, 678, 678, 678, 679, 679, 679,
679, 679, 680, 680, 680, 680, 680, 680, 681, 681, 681,
681, 681, 682, 682, 682, 682, 682, 682, 683, 683, 683,
683, 683, 684, 684, 684, 684, 684, 684, 685, 685, 685,
685, 685, 686, 686, 686, 686, 686, 686, 687, 687, 687,
687, 687, 687, 688, 688, 688, 688, 688, 689, 689, 689,
689, 689, 690, 690, 690, 690, 690, 691, 691, 691, 691,
691, 691, 692, 692, 692, 692, 692, 692, 692, 693, 693,
693, 693, 693, 693, 694, 694, 694, 694, 694, 694, 695,
695, 695, 695, 695, 695, 696, 696, 696, 696, 696, 696,
697, 697, 697, 697, 697, 698, 698, 698, 698, 698, 698,
699, 699, 699, 699, 699, 700, 700, 700, 700, 700, 700,
700, 701, 701, 701, 701, 701, 701, 702, 702, 702, 702,
702, 702, 702, 702, 703, 703, 703, 703, 703, 703, 703,
703, 704, 704, 704, 704, 704, 704, 704, 705, 705, 705,
705, 705, 705, 705, 706, 706, 706, 706, 707, 707, 707,
707, 707, 707, 707, 707, 708, 708, 708, 708, 708, 708,
708, 708, 709, 709, 709, 709, 709, 709, 710, 710, 710,
710, 710, 710, 710, 710, 711, 711, 711, 711, 711, 711,
711, 712, 712, 712, 712, 712, 712, 712, 713, 713, 713,
713, 713, 714, 714, 714, 714, 714, 714, 715, 715, 715,
715, 715, 715, 716, 716, 716, 716, 716, 716, 716, 717,
717, 717, 717, 717, 717, 717, 718, 718, 718, 718, 718,
718, 718, 719, 719, 719, 719, 719, 719, 719, 720, 720,
720, 720, 720, 721, 721, 721, 721, 721, 721, 721, 721,
721, 721, 721, 722, 722, 722, 722, 722, 723, 723, 723,
723, 723, 723, 724, 724, 724, 724, 724, 724, 724, 724,
725, 725, 725, 725, 725, 725, 726, 726, 726, 726, 726,
726, 727, 727, 727, 727, 727, 727, 727, 728, 728, 728,
728, 728, 729, 729, 729, 729, 729, 729, 729, 730, 730,
730, 730, 730, 730, 730, 731, 731, 731, 731, 731, 731,
731, 731, 731, 732, 732, 732, 732, 732, 732, 733, 733,
733, 733, 733, 734, 734, 734, 734, 735, 735, 735, 735,
735, 735, 736, 736, 736, 736, 736, 737, 737, 737, 737,
737, 737, 738, 738, 738, 738, 738, 738, 738, 738, 739,
739, 739, 739, 739, 739, 740, 740, 740, 740, 740, 740,
740, 740, 741, 741, 741, 741, 741, 742, 742, 742, 742,
742, 742, 743, 743, 743, 743, 743, 744, 744, 744, 744,
744, 744, 744, 745, 745, 745, 745, 745, 746, 746, 746,
746, 746, 747, 747, 747, 747, 747, 747, 747, 748, 748,
748, 748, 749, 749, 749, 749, 749, 749, 750, 750, 750,
750, 750, 750, 751, 751, 751, 751, 751, 751, 751, 752,
752, 752, 752, 753, 753, 753, 753, 754, 754, 754, 754,
754, 755, 755, 755, 755, 755, 756, 756, 756, 756, 756,
756, 756, 757, 757, 757, 757, 757, 757, 758, 758, 758,
758, 758, 759, 759, 759, 759, 759, 759, 759, 760, 760,
760, 760, 760, 760, 761, 761, 761, 761, 761, 761, 762,
762, 762, 762, 762, 762, 762, 763, 763, 763, 763, 764,
764, 764, 764, 764, 764, 764, 765, 765, 765, 765, 765,
765, 766, 766, 766, 766, 766, 766, 767, 767, 767, 767,
767, 767, 768, 768, 768, 769, 769, 769, 769, 769, 769,
770, 770, 770, 770, 770, 770, 771, 771, 771, 771, 771,
771, 772, 772, 772, 772, 772, 773, 773, 773, 773, 773,
773, 774, 774, 774, 774, 774, 774, 775, 775, 775, 775,
775, 775, 775, 776, 776, 776, 777, 777, 777, 777, 777,
777, 778, 778, 778, 778, 778, 779, 779, 779, 779, 779,
779, 779, 779, 780, 780, 780, 780, 780, 781, 781, 781,
781, 781, 781, 781, 782, 782, 782, 782, 782, 783, 783,
783, 783, 783, 784, 784, 784, 784, 785, 785, 785, 785,
785, 785, 786, 786, 786, 786, 786, 786, 787, 787, 787,

787, 787, 788, 788, 788, 788, 788, 788, 789, 789, 789,
790, 790, 790, 790, 790, 790, 791, 791, 791, 791, 791,
792, 792, 792, 792, 792, 793, 793, 793, 793, 793, 793,
794, 794, 794, 794, 795, 795, 795, 795, 795, 795, 795,
796, 796, 796, 796, 796, 796, 797, 797, 797, 797, 797,
798, 798, 798, 798, 798, 799, 799, 799, 799, 799, 799,
799, 800, 800, 800, 800, 800, 800, 800, 801, 801, 801,
801, 802, 802, 802, 802, 802, 802, 802, 802, 803, 803,
803, 803, 803, 804, 804, 804, 804, 804, 804, 805, 805,
805, 806, 806, 806, 806, 806, 807, 807, 807, 807, 808,
808, 808, 808, 808, 809, 809, 809, 809, 809, 810, 810,
810, 810, 810, 811, 811, 811, 811, 811, 812, 812, 812,
812, 812, 813, 813, 813, 813, 813, 814, 814, 814, 814,
815, 815, 815, 815, 815, 815, 816, 816, 816, 816, 816,
817, 817, 817, 817, 817, 817, 818, 818, 818, 818, 818,
818, 819, 819, 819, 819, 819, 819, 820, 820, 820, 820,
821, 821, 821, 821, 821, 821, 822, 822, 822, 822, 822,
823, 823, 823, 823, 823, 824, 824, 824, 824, 824, 825,
825, 825, 825, 825, 825, 826, 826, 826, 826, 826, 827,
827, 827, 827, 827, 827, 828, 828, 828, 828, 828, 829,
829, 829, 829, 830, 830, 830, 830, 830, 831, 831, 831,
831, 831, 832, 832, 832, 832, 832, 833, 833, 833, 833,
834, 834, 834, 835, 835, 835, 835, 835, 835, 836, 836,
836, 836, 837, 837, 837, 837, 837, 837, 838, 838, 838,
838, 839, 839, 839, 839, 840, 840, 840, 841, 841, 841,
841, 841, 842, 842, 842, 842, 843, 843, 843, 843, 844,
844, 844, 844, 844, 845, 845, 845, 846, 846, 846, 846,
846, 846, 846, 847, 847, 847, 847, 847, 848, 848, 848,
848, 849, 849, 849, 849, 850, 850, 850, 850, 851, 851,
851, 851, 851, 851, 852, 852, 852, 853, 853, 853, 853,
854, 854, 855, 855, 855, 855, 855, 856, 856, 856, 856,
857, 857, 857, 857, 858, 858, 858, 858, 859, 859, 859,
860, 860, 860, 860, 860, 861, 861, 861, 861, 862, 862,
862, 863, 863, 863, 863, 863, 864, 864, 864, 864, 865,
865, 865, 866, 866, 866, 866, 867, 867, 867, 868, 868,
868, 868, 869, 869, 869, 869, 870, 870, 871, 871, 871,
871, 872, 873, 873, 873, 873, 874, 874, 874, 874, 875,
875, 876, 876, 876, 877, 877, 877, 878, 878, 878, 878,
879, 879, 879, 880, 880, 881, 881, 881, 882, 882, 882,
882, 883, 883, 883, 883, 884, 884, 884, 885, 885, 885,
886, 886, 887, 887, 887, 888, 888, 888, 888, 889, 889,
889, 890, 890, 890, 890, 891, 891, 892, 892, 893, 893,
893, 894, 894, 894, 895, 895, 895, 896, 896, 897, 897,
898, 898, 899, 899, 899, 900, 900, 900, 901, 901, 901,
902, 902, 902, 903, 904, 904, 905, 905, 905, 906, 906,
906, 907, 907, 907, 908, 908, 908, 909, 909, 909, 909,
910, 910, 911, 912, 912, 913, 913, 914, 914, 914, 915,
915, 916, 916, 916, 917, 918, 919, 919, 919, 920, 921,
921, 922, 922, 923, 923, 924, 924, 924, 925, 925, 926,
926, 926, 928, 928, 928, 929, 929, 930, 930, 930, 931,
931, 932, 932, 932, 932, 933, 933, 934, 934, 935, 935,
936, 937, 938, 940, 940, 941, 941, 942, 943, 943, 944,
944, 945, 945, 946, 946, 947, 948, 948, 949, 950, 950,
951, 951, 952, 953, 953, 953, 954, 954, 955, 956, 956,
957, 958, 958, 959, 960, 960, 961, 962, 962, 963, 964,
965, 966, 966, 968, 968, 969, 970, 970, 972, 973, 974,
975, 976, 976, 976, 977, 978, 979, 980, 980, 981, 982,
983, 984, 985, 985, 986, 987, 988, 990, 991, 992, 993,
994, 995, 995, 996, 997, 997, 998, 999, 1001, 1002, 1004,
1005, 1007, 1008, 1009, 1009, 1010, 1010, 1011, 1012,
1012, 1014, 1015, 1016, 1017, 1018, 1019, 1021, 1022,
1023, 1024, 1025, 1026, 1028, 1028, 1030, 1031, 1033,
1035, 1036, 1037, 1039, 1040, 1041, 1045, 1046, 1047,
1049, 1050, 1052, 1053, 1054, 1055, 1056, 1058, 1059,

1061, 1062, 1062, 1063, 1064, 1066, 1067, 1069, 1071,
1073, 1074, 1075, 1076, 1077, 1078, 1080, 1081, 1082,
1082, 1084, 1084, 1086, 1087, 1089, 1091, 1092, 1093,
1094, 1096, 1099, 1100, 1102, 1105, 1106, 1107, 1109,
1111, 1113, 1113, 1115, 1117, 1118, 1119, 1121, 1123,
1123, 1124, 1125, 1126, 1128, 1129, 1131, 1132, 1134,
1135, 1136, 1139, 1141, 1145, 1147, 1151, 1152, 1152,
1155, 1156, 1160, 1161, 1164, 1165, 1167, 1168, 1170,
1171, 1172, 1173, 1175, 1176, 1177, 1179, 1182, 1183,
1186, 1188, 1190, 1191, 1193, 1195, 1198, 1201, 1204,
1205, 1207, 1208, 1210, 1213, 1214, 1217, 1218, 1220,
1222, 1224, 1225, 1227, 1228, 1231, 1232, 1233, 1236,
1238, 1241, 1243, 1245, 1248, 1250, 1252, 1256, 1258,
1261, 1262, 1263, 1266, 1270, 1272, 1275, 1277, 1279,
1281, 1282, 1285, 1288, 1292, 1295, 1296, 1298, 1300,
1303, 1304, 1306, 1309, 1311, 1312, 1314, 1318, 1320,
1323, 1326, 1328, 1332, 1336, 1340, 1343, 1349, 1354,
1357, 1360, 1364, 1367, 1371, 1375, 1378, 1380, 1383,
1385, 1389, 1391, 1393, 1399, 1401, 1405, 1409, 1414,
1416, 1419, 1421, 1423, 1430, 1432, 1434, 1438, 1441,
1444, 1450, 1451, 1454, 1459, 1462, 1465, 1468, 1473,
1475, 1478, 1482, 1485, 1487, 1491, 1493, 1500, 1502,
1506, 1507, 1510, 1514, 1517, 1520, 1523, 1525, 1530,
1532, 1535, 1537, 1542, 1546, 1549, 1551, 1553, 1558,
1561, 1568, 1570, 1574, 1577, 1582, 1585, 1589, 1597,
1602, 1605, 1607, 1611, 1615, 1618, 1621, 1627, 1632,
1636, 1641, 1645, 1650, 1655, 1657, 1661, 1666, 1674,
1679, 1681, 1685, 1689, 1691, 1697, 1699, 1703, 1705,
1711, 1716, 1717, 1722, 1727, 1730, 1734, 1735, 1740,
1747, 1749, 1755, 1757, 1759, 1768, 1771, 1775, 1780,
1786, 1788, 1792, 1794, 1798, 1802, 1804, 1806, 1807,
1809, 1812, 1814, 1817, 1818, 1821, 1827, 1830, 1833,
1837, 1841, 1843, 1847, 1852, 1855, 1859, 1862, 1866,
1869, 1873, 1876, 1879, 1883, 1890, 1895, 1901, 1907,
1909, 1913, 1920, 1924, 1930, 1931, 1937, 1940, 1942,
1945, 1949, 1953, 1957, 1961, 1965, 1967, 1971, 1973,
1976, 1978, 1980, 1984, 1989, 1992, 1996, 2003, 2007,
2010, 2016, 2019, 2023, 2027, 2032, 2035, 2041, 2045,
2049, 2058, 2060, 2067, 2069, 2071, 2074, 2077, 2082,
2085, 2089, 2097, 2100, 2103, 2108, 2115, 2117, 2120,
2123, 2128, 2130, 2134, 2137, 2141, 2150, 2152, 2156,
2161, 2163, 2167, 2172, 2178, 2184, 2187, 2192, 2195,
2199, 2201, 2205, 2208, 2213, 2220, 2222, 2225, 2228,
2231, 2235, 2240, 2248, 2254, 2256, 2258, 2261, 2264,
2272, 2274, 2278, 2283, 2286, 2290, 2293, 2295, 2299,
2302, 2308, 2311, 2318, 2324, 2328, 2330, 2334, 2339,
2343, 2349, 2351, 2357, 2365, 2368, 2380, 2384, 2389,
2393, 2403, 2408, 2415, 2421, 2426, 2435, 2436, 2444,
2448, 2455, 2460, 2466, 2475, 2481, 2485, 2493, 2498,
2506, 2514, 2518, 2522, 2527, 2531, 2536, 2547, 2555,
2561, 2568, 2574, 2581, 2584, 2593, 2596, 2604, 2607,
2612, 2618, 2623, 2628, 2633, 2639, 2643, 2649, 2653,
2662, 2668, 2681, 2687, 2691, 2705, 2712, 2726, 2732,
2740, 2746, 2749, 2755, 2761, 2766, 2774, 2779, 2794,
2805, 2813, 2820, 2825, 2830, 2836, 2849, 2852, 2859,
2866, 2872, 2878, 2888, 2893, 2898, 2907, 2912, 2918,
2925, 2934, 2938, 2943, 2952, 2958, 2964, 2969, 2973,
2983, 2991, 3001, 3006, 3017, 3020, 3029, 3031, 3039,
3046, 3050, 3054, 3057, 3067, 3072, 3081, 3084, 3088,
3091, 3100, 3105, 3110, 3113, 3117, 3126, 3131, 3135,
3141, 3148, 3158, 3161, 3169, 3178, 3185, 3191, 3194,
3199, 3207, 3210, 3215, 3226, 3231, 3232, 3242, 3255,
3268, 3276, 3286, 3299, 3306, 3315, 3323, 3331, 3338,
3352, 3367, 3379, 3391, 3419, 3433, 3451, 3463, 3479,

```
    3503, 3522, 3532, 3570, 3598, 3621, 3642, 3657, 3699,
    3718, 3738, 3759, 3828, 3904, 3985, 4046, 4104, 4169,
    4239, 4374, 4659, 5105, 5734, 7805])

short_spike_matrix = scale_down(spike_matrix=spike_matrix,
    left_cf=left_cf, right_cf=right_cf, left_target_cf=
    left_target_cf, right_target_cf=right_target_cf,
    scaling_factor=scaling_factor)

filename = target_folder + 'bio_112m5h_mni0/sam' +
    str(sample_id) + '_' + speaker + '_' + str(digit) + '.csv'

np.savetxt(filename, short_spike_matrix, fmt='%i',
    delimiter=",")

#MNI_orig
left_target_cf = np.array([166, 400, 431, 457, 479, 498,
    507, 516, 521, 526, 532, 539, 543, 548, 553, 559, 563,
    567, 572, 575, 578, 582, 584, 587, 589, 592, 595, 598,
    600, 602, 604, 606, 609, 611, 613, 615, 617, 619, 621,
    622, 625, 626, 627, 629, 630, 632, 634, 634, 636, 637,
    638, 639, 640, 642, 643, 645, 646, 647, 648, 649, 650,
    651, 653, 653, 654, 655, 657, 658, 659, 660, 661, 662,
    662, 663, 664, 665, 667, 668, 669, 670, 670, 671, 672,
    674, 674, 675, 676, 678, 679, 680, 681, 682, 683, 685,
    686, 687, 688, 689, 690, 692, 693, 694, 694, 695, 697,
    697, 699, 700, 701, 702, 703, 704, 706, 707, 708, 709,
    710, 711, 712, 713, 714, 716, 717, 718, 719, 720, 722,
    723, 724, 726, 727, 728, 729, 730, 732, 733, 734, 735,
    736, 737, 738, 739, 741, 742, 743, 744, 745, 746, 748,
    749, 750, 752, 754, 755, 756, 757, 759, 760, 761, 763,
    764, 765, 766, 767, 768, 770, 771, 773, 774, 775, 777,
    778, 779, 781, 782, 783, 785, 786, 787, 789, 790, 791,
    792, 794, 795, 796, 797, 798, 799, 800, 802, 803, 804,
    805, 807, 808, 809, 810, 811, 813, 814, 815, 817, 818,
    820, 821, 823, 824, 826, 829, 831, 833, 835, 836, 839,
    841, 843, 845, 848, 850, 852, 855, 858, 860, 863, 866,
    869, 872, 876, 879, 881, 885, 890, 894, 897, 900, 903,
    908, 913, 918, 922, 929, 934, 940, 946, 952, 957, 965,
    971, 978, 987, 998, 1008, 1016, 1026, 1036, 1046, 1060,
    1073, 1082, 1095, 1106, 1118, 1129, 1141, 1156, 1169,
    1184, 1199, 1212, 1224, 1245, 1257, 1272, 1287, 1306,
    1322, 1341, 1356, 1373, 1396, 1414, 1433, 1453, 1474,
    1491, 1508, 1531, 1546, 1564, 1591, 1615, 1635, 1657,
    1686, 1707, 1727, 1750, 1778, 1803, 1840, 1873, 1901,
    1930, 1964, 1992, 2021, 2048, 2075, 2111, 2141, 2167,
    2196, 2223, 2250, 2272, 2308, 2334, 2368, 2401, 2447,
    2480, 2510, 2552, 2591, 2633, 2676, 2717, 2774, 2847,
    2938, 3026, 3175, 3326, 3552, 3806, 4379, 5401, 8000])

right_target_cf = np.array([161, 384, 406, 448, 480, 506,
    525, 535, 547, 555, 560, 564, 567, 572, 575, 578, 580,
    582, 585, 588, 590, 593, 595, 598, 599, 602, 603, 605,
    606, 608, 609, 611, 613, 614, 616, 618, 619, 621, 623,
    624, 626, 627, 629, 630, 632, 633, 635, 636, 638, 639,
    641, 643, 645, 646, 648, 649, 651, 653, 654, 656, 657,
    659, 660, 662, 664, 665, 666, 667, 669, 670, 672, 673,
    675, 676, 678, 679, 681, 682, 683, 685, 686, 688, 689,
    691, 692, 693, 695, 696, 697, 699, 700, 702, 703, 703,
    705, 706, 707, 708, 710, 710, 712, 713, 714, 716, 717,
    718, 719, 720, 721, 722, 724, 725, 726, 727, 729, 730,
    731, 732, 733, 735, 737, 738, 739, 740, 741, 743, 744,
    746, 747, 749, 750, 751, 753, 755, 756, 757, 759, 760,
    761, 762, 764, 765, 767, 768, 770, 771, 773, 774, 775,
```

```
    777, 778, 779, 781, 782, 784, 785, 787, 788, 790, 792,
    793, 795, 796, 797, 799, 800, 802, 803, 804, 806, 808,
    809, 811, 813, 814, 816, 817, 819, 820, 822, 823, 825,
    826, 828, 829, 831, 833, 835, 836, 838, 840, 842, 844,
    846, 847, 849, 851, 853, 855, 857, 859, 861, 863, 865,
    867, 869, 872, 875, 878, 880, 883, 885, 888, 890, 893,
    896, 900, 902, 906, 908, 912, 916, 920, 924, 928, 931,
    934, 941, 945, 950, 954, 959, 966, 972, 979, 985, 993,
    1000, 1010, 1016, 1025, 1037, 1049, 1060, 1071, 1080,
    1090, 1103, 1116, 1125, 1137, 1155, 1170, 1182, 1199,
    1216, 1229, 1247, 1265, 1283, 1304, 1320, 1350, 1379,
    1403, 1430, 1455, 1482, 1508, 1533, 1558, 1590, 1626,
    1658, 1693, 1725, 1755, 1789, 1810, 1837, 1862, 1896,
    1935, 1962, 1987, 2020, 2058, 2086, 2121, 2154, 2187,
    2220, 2254, 2284, 2312, 2349, 2393, 2444, 2493, 2538,
    2593, 2633, 2687, 2751, 2813, 2866, 2919, 2970, 3029,
    3072, 3114, 3161, 3210, 3281, 3352, 3479, 3657, 4046,
    7805])

short_spike_matrix = scale_down(spike_matrix=spike_matrix,
    left_cf=left_cf, right_cf=right_cf, left_target_cf=
    left_target_cf, right_target_cf=right_target_cf,
    scaling_factor=scaling_factor)

filename = target_folder + 'bio_112m5h_mni1/sam' +
    str(sample_id) + '_' + speaker + '_' + str(digit) + '.csv'

np.savetxt(filename, short_spike_matrix, fmt='%i',
    delimiter=",")

#MNI_by_2
left_target_cf = np.array([166, 521, 564, 590, 609, 625,
    636, 646, 655, 663, 671, 679, 688, 697, 706, 715, 725,
    735, 744, 755, 765, 775, 786, 796, 806, 816, 829, 846,
    867, 895, 931, 982, 1065, 1163, 1283, 1428, 1583, 1773,
    2017, 2246, 2505, 2928, 8000])
right_target_cf = np.array([161, 547, 580, 599, 612, 625,
    637, 650, 663, 674, 686, 697, 707, 716, 725, 735, 746,
    758, 769, 780, 792, 803, 816, 827, 842, 856, 874, 895,
    923, 957, 1012, 1094, 1206, 1360, 1570, 1815, 2065, 2320,
    2691, 3115, 7805])

short_spike_matrix = scale_down(spike_matrix=spike_matrix,
    left_cf=left_cf, right_cf=right_cf, left_target_cf=
    left_target_cf, right_target_cf=right_target_cf,
    scaling_factor=scaling_factor)

filename = target_folder + 'bio_112m5h_mni2/sam' +
    str(sample_id) + '_' + speaker + '_' + str(digit) + '.csv'

np.savetxt(filename, short_spike_matrix, fmt='%i',
    delimiter=",")

#MNI_by_3
left_target_cf = np.array([166, 600, 646, 675, 706, 739,
    775, 810, 867, 1022, 1428, 2138, 8000])

right_target_cf = np.array([161, 604, 646, 686, 719, 754,
    792, 832, 888, 1012, 1430, 2233,7805])

short_spike_matrix = scale_down(spike_matrix=spike_matrix,
    left_cf=left_cf, right_cf=right_cf, left_target_cf=
    left_target_cf, right_target_cf=right_target_cf,
    scaling_factor=scaling_factor)
```

```
filename = target_folder + 'bio_112m5h_mni3/sam' +
    str(sample_id) + '_' + speaker + '_' + str(digit) + '.csv'

np.savetxt(filename, short_spike_matrix, fmt='%i',
    delimiter=",")

#MNI_by_4
left_target_cf = np.array([166, 658, 733, 818, 1232,
    8000])

right_target_cf = np.array([161, 663, 746, 842, 1206,
    7805])

short_spike_matrix = scale_down(spike_matrix=spike_matrix,
    left_cf=left_cf, right_cf=right_cf, left_target_cf=
    left_target_cf, right_target_cf=right_target_cf,
    scaling_factor=scaling_factor)

filename = target_folder + 'bio_112m5h_mni4/sam' +
    str(sample_id) + '_' + speaker + '_' + str(digit) + '.csv'

np.savetxt(filename, short_spike_matrix, fmt='%i',
    delimiter=",")

#MNI_by_5
left_target_cf = np.array([166, 706, 867, 8000])
right_target_cf = np.array([161, 719, 888, 7805])

short_spike_matrix = scale_down(spike_matrix=spike_matrix,
    left_cf=left_cf, right_cf=right_cf, left_target_cf=
    left_target_cf, right_target_cf=right_target_cf,
    scaling_factor=scaling_factor)

filename = target_folder + 'bio_112m5h_mni5/sam' +
    str(sample_id) + '_' + speaker + '_' + str(digit) + '.csv'

np.savetxt(filename, short_spike_matrix, fmt='%i',
    delimiter=",")

#TAL_orig
left_target_cf = np.array([166, 270, 353, 368, 374, 394,
    398, 402, 409, 417, 421, 425, 430, 434, 439, 442, 444,
    450, 454, 457, 460, 463, 467, 470, 473, 477, 483, 486,
    490, 492, 495, 497, 498, 500, 501, 503, 504, 505, 507,
    508, 510, 511, 512, 513, 515, 516, 517, 518, 519, 520,
    520, 521, 522, 523, 524, 525, 526, 526, 528, 529, 530,
    531, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540,
    540, 541, 542, 543, 543, 544, 545, 546, 546, 547, 548,
    549, 550, 551, 551, 552, 553, 553, 554, 554, 556, 557,
    558, 559, 560, 561, 561, 562, 562, 563, 564, 564, 565,
    566, 566, 567, 568, 569, 569, 570, 570, 571, 572, 573,
    573, 573, 574, 574, 575, 576, 576, 577, 577, 578, 578,
    579, 579, 580, 580, 581, 582, 582, 582, 582, 583, 583,
    583, 584, 584, 585, 585, 586, 586, 587, 587, 587, 588,
    588, 589, 589, 590, 590, 590, 591, 591, 592, 592, 593,
    593, 593, 594, 594, 595, 595, 596, 596, 597, 597, 597,
    598, 598, 598, 599, 599, 600, 600, 600, 601, 601, 602,
    602, 602, 602, 603, 603, 603, 604, 604, 604, 605, 605,
    605, 606, 606, 606, 607, 607, 608, 608, 608, 609, 609,
    609, 610, 610, 610, 611, 611, 611, 612, 612, 612, 612,
    613, 613, 613, 613, 614, 614, 615, 615, 615, 615, 616,
    616, 616, 617, 617, 618, 618, 618, 619, 619, 619, 620,
    620, 620, 620, 621, 621, 621, 621, 622, 622, 622, 623,
```

623, 623, 624, 624, 624, 625, 625, 625, 625, 626, 626,
626, 626, 627, 627, 627, 627, 627, 628, 628, 628, 628,
628, 629, 629, 629, 629, 629, 630, 630, 630, 630, 631,
631, 631, 631, 632, 632, 632, 633, 633, 633, 633, 634,
634, 634, 634, 634, 634, 634, 635, 635, 635, 635, 635,
635, 636, 636, 636, 636, 636, 637, 637, 637, 637, 638,
638, 638, 638, 638, 638, 638, 639, 639, 639, 639, 639,
639, 640, 640, 640, 640, 641, 641, 641, 641, 641, 642,
642, 642, 643, 643, 643, 643, 643, 643, 644, 644, 644,
644, 645, 645, 645, 645, 645, 646, 646, 646, 646, 647,
647, 647, 647, 647, 647, 648, 648, 648, 648, 648, 648,
648, 649, 649, 649, 649, 649, 649, 649, 650, 650, 650,
650, 650, 651, 651, 651, 651, 651, 652, 652, 652, 652,
652, 652, 653, 653, 653, 653, 653, 653, 653, 654, 654,
654, 654, 654, 654, 655, 655, 655, 655, 655, 655, 656,
656, 656, 656, 656, 657, 657, 657, 657, 657, 657, 657,
658, 658, 658, 658, 658, 658, 659, 659, 659, 659, 659,
659, 659, 660, 660, 660, 660, 660, 660, 661, 661, 661,
661, 661, 662, 662, 662, 662, 662, 662, 662, 662, 663,
663, 663, 663, 663, 663, 663, 664, 664, 664, 664, 664,
664, 665, 665, 665, 665, 665, 665, 665, 666, 666, 666,
666, 666, 667, 667, 667, 667, 667, 667, 668, 668, 668,
668, 668, 668, 669, 669, 669, 669, 669, 669, 669, 670,
670, 670, 670, 670, 670, 670, 671, 671, 671, 671, 671,
671, 671, 672, 672, 672, 672, 672, 672, 672, 673, 673,
673, 673, 674, 674, 674, 674, 674, 674, 674, 674, 675,
675, 675, 675, 675, 675, 675, 676, 676, 676, 676, 676,
676, 677, 677, 677, 677, 677, 678, 678, 678, 678, 679,
679, 679, 679, 679, 679, 679, 680, 680, 680, 680, 680,
680, 681, 681, 681, 681, 681, 682, 682, 682, 682, 682,
683, 683, 683, 683, 683, 684, 684, 684, 684, 684, 685,
685, 685, 685, 685, 686, 686, 686, 686, 686, 686, 687,
687, 687, 687, 687, 687, 688, 688, 688, 688, 688, 688,
689, 689, 689, 689, 689, 689, 689, 690, 690, 690, 690,
690, 691, 691, 691, 691, 692, 692, 692, 692, 692, 692,
693, 693, 693, 693, 693, 693, 694, 694, 694, 694, 694,
694, 694, 694, 695, 695, 695, 695, 695, 696, 696, 696,
696, 696, 696, 697, 697, 697, 697, 697, 697, 697, 698,
698, 698, 699, 699, 699, 699, 699, 699, 700, 700, 700,
700, 700, 701, 701, 701, 701, 701, 701, 701, 702, 702,
702, 702, 702, 703, 703, 703, 703, 703, 703, 703, 704,
704, 704, 704, 704, 705, 705, 705, 705, 705, 705, 706,
706, 706, 706, 706, 706, 707, 707, 707, 707, 707, 708,
708, 708, 708, 708, 708, 709, 709, 709, 709, 709, 710,
710, 710, 710, 710, 710, 711, 711, 711, 711, 711, 711,
712, 712, 712, 712, 712, 712, 712, 713, 713, 713, 713,
713, 713, 714, 714, 714, 714, 714, 714, 715, 715, 715,
716, 716, 716, 716, 716, 716, 717, 717, 717, 717, 717,
717, 718, 718, 718, 718, 718, 718, 719, 719, 719, 719,
719, 720, 720, 720, 720, 720, 721, 721, 721, 721, 722,
722, 722, 722, 723, 723, 723, 723, 723, 724, 724, 724,
724, 724, 724, 725, 725, 725, 725, 726, 726, 726, 726,
726, 727, 727, 727, 727, 727, 727, 728, 728, 728, 728,
728, 728, 728, 729, 729, 729, 729, 730, 730, 730, 730,
731, 731, 731, 731, 731, 731, 732, 732, 732, 732, 732,
733, 733, 733, 733, 733, 734, 734, 734, 734, 734, 734,
734, 734, 735, 735, 735, 735, 735, 735, 736, 736, 736,
736, 736, 737, 737, 737, 737, 737, 737, 738, 738, 738,
738, 739, 739, 739, 739, 739, 739, 740, 740, 740, 740,
740, 740, 741, 741, 741, 741, 741, 742, 742, 742, 743,
743, 743, 743, 743, 744, 744, 744, 744, 744, 744, 744,
745, 745, 745, 745, 745, 745, 746, 746, 746, 746, 746,
747, 747, 747, 747, 748, 748, 748, 748, 748, 748, 749,
749, 749, 749, 750, 750, 750, 750, 750, 751, 751, 751,

751, 752, 752, 752, 752, 753, 753, 753, 753, 754, 754,
754, 754, 754, 754, 755, 755, 755, 755, 755, 755, 756,
756, 756, 756, 756, 757, 757, 757, 758, 758, 758, 758,
759, 759, 759, 759, 759, 759, 760, 760, 760, 760, 761,
761, 761, 761, 761, 761, 762, 762, 762, 762, 762, 763,
763, 763, 763, 763, 764, 764, 764, 764, 764, 764, 765,
765, 765, 765, 765, 765, 766, 766, 766, 766, 766, 766,
766, 767, 767, 767, 767, 768, 768, 768, 768, 769, 769,
769, 769, 769, 770, 770, 770, 770, 771, 771, 771, 771,
771, 772, 772, 772, 772, 772, 773, 773, 773, 773, 774,
774, 774, 774, 775, 775, 775, 775, 775, 776, 776, 776,
776, 777, 777, 777, 777, 777, 777, 778, 778, 778, 778,
779, 779, 779, 779, 779, 780, 780, 780, 780, 781, 781,
781, 781, 781, 782, 782, 782, 782, 783, 783, 783, 783,
783, 783, 784, 784, 784, 784, 784, 785, 785, 785, 785,
785, 786, 786, 786, 786, 787, 787, 787, 787, 787, 788,
788, 788, 788, 788, 789, 789, 789, 789, 789, 789, 790,
790, 790, 790, 791, 791, 791, 791, 791, 791, 792, 792,
792, 792, 792, 793, 793, 793, 793, 793, 794, 794, 794,
794, 795, 795, 795, 795, 795, 795, 795, 795, 796, 796,
796, 796, 796, 796, 797, 797, 797, 797, 797, 798, 798,
798, 798, 798, 798, 799, 799, 799, 799, 800, 800, 800,
800, 800, 801, 801, 801, 801, 801, 802, 802, 802, 802,
802, 803, 803, 803, 803, 803, 804, 804, 804, 804, 804,
805, 805, 805, 805, 805, 806, 806, 806, 806, 807, 807,
807, 807, 807, 808, 808, 808, 808, 808, 808, 809, 809,
809, 809, 809, 809, 810, 810, 810, 810, 810, 810, 811,
811, 811, 811, 811, 812, 812, 812, 812, 812, 813, 813,
813, 813, 813, 814, 814, 814, 814, 814, 815, 815, 815,
815, 815, 816, 816, 816, 816, 816, 817, 817, 817, 817,
818, 818, 818, 818, 818, 819, 819, 819, 820, 820, 820,
820, 820, 820, 821, 821, 821, 822, 822, 822, 822, 823,
823, 823, 823, 824, 824, 824, 825, 825, 825, 825, 825,
826, 826, 826, 827, 827, 828, 828, 829, 829, 829, 830,
830, 830, 831, 831, 831, 832, 832, 832, 833, 833, 834,
834, 834, 834, 835, 835, 835, 835, 835, 836, 836, 836,
836, 837, 837, 838, 838, 838, 839, 839, 839, 840, 840,
840, 841, 841, 842, 842, 842, 842, 843, 843, 844, 844,
844, 844, 845, 845, 845, 846, 846, 847, 847, 848, 848,
848, 848, 849, 849, 849, 850, 850, 851, 851, 851, 852,
852, 853, 853, 853, 854, 854, 855, 855, 856, 856, 857,
857, 857, 858, 858, 858, 859, 859, 860, 860, 861, 861,
862, 862, 863, 863, 864, 864, 865, 865, 865, 866, 866,
867, 867, 868, 868, 869, 869, 869, 870, 871, 871, 871,
872, 872, 873, 874, 874, 874, 875, 876, 877, 877, 878,
878, 879, 879, 879, 880, 880, 881, 881, 882, 882, 882,
883, 884, 884, 885, 886, 886, 887, 888, 889, 889, 890,
891, 891, 892, 893, 893, 894, 894, 895, 895, 896, 897,
897, 897, 898, 898, 899, 899, 900, 900, 901, 902, 902,
902, 903, 903, 905, 905, 906, 907, 907, 908, 909, 909,
910, 911, 912, 913, 914, 914, 915, 916, 917, 917, 918,
919, 919, 920, 920, 921, 922, 923, 924, 925, 926, 927,
928, 930, 930, 931, 931, 933, 933, 934, 934, 935, 936,
938, 939, 940, 941, 942, 943, 944, 945, 945, 946, 947,
948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 957,
958, 959, 960, 962, 964, 965, 966, 966, 967, 967, 969,
969, 971, 972, 973, 975, 976, 977, 978, 979, 980, 982,
984, 985, 987, 987, 990, 991, 992, 993, 997, 999, 1001,
1002, 1002, 1004, 1006, 1008, 1009, 1010, 1011, 1012,
1014, 1015, 1017, 1019, 1021, 1023, 1024, 1025, 1026,
1028, 1029, 1031, 1032, 1034, 1036, 1038, 1039, 1041,
1042, 1043, 1045, 1048, 1049, 1051, 1052, 1054, 1057,
1060, 1063, 1064, 1066, 1067, 1072, 1073, 1074, 1076,
1078, 1078, 1080, 1082, 1084, 1086, 1090, 1091, 1092,

```
        1094, 1095, 1097, 1099, 1101, 1102, 1103, 1107, 1108,
        1109, 1111, 1113, 1115, 1117, 1119, 1120, 1122, 1124,
        1126, 1128, 1131, 1131, 1133, 1135, 1138, 1139, 1142,
        1143, 1146, 1149, 1151, 1154, 1156, 1158, 1160, 1162,
        1163, 1166, 1168, 1169, 1171, 1174, 1178, 1181, 1182,
        1184, 1186, 1188, 1192, 1194, 1196, 1199, 1200, 1202,
        1204, 1208, 1209, 1211, 1213, 1214, 1218, 1219, 1221,
        1223, 1224, 1227, 1230, 1236, 1238, 1241, 1244, 1246,
        1249, 1250, 1251, 1255, 1257, 1259, 1261, 1262, 1264,
        1266, 1270, 1272, 1276, 1278, 1280, 1283, 1284, 1287,
        1289, 1292, 1295, 1297, 1301, 1304, 1308, 1309, 1313,
        1315, 1317, 1320, 1322, 1325, 1329, 1333, 1334, 1338,
        1341, 1343, 1347, 1350, 1352, 1353, 1355, 1358, 1361,
        1362, 1365, 1368, 1371, 1375, 1379, 1381, 1386, 1389,
        1393, 1395, 1397, 1402, 1403, 1407, 1410, 1413, 1415,
        1418, 1423, 1428, 1429, 1431, 1434, 1437, 1441, 1443,
        1447, 1449, 1453, 1455, 1457, 1460, 1466, 1470, 1473,
        1474, 1477, 1478, 1481, 1485, 1489, 1491, 1493, 1496,
        1500, 1502, 1506, 1508, 1510, 1515, 1519, 1523, 1526,
        1531, 1533, 1535, 1536, 1537, 1541, 1543, 1546, 1550,
        1553, 1555, 1558, 1560, 1564, 1566, 1574, 1577, 1581,
        1584, 1588, 1592, 1598, 1602, 1605, 1607, 1612, 1615,
        1622, 1626, 1629, 1631, 1634, 1635, 1639, 1642, 1645,
        1651, 1654, 1657, 1660, 1664, 1669, 1673, 1679, 1682,
        1686, 1689, 1691, 1693, 1697, 1702, 1707, 1712, 1714,
        1717, 1719, 1722, 1727, 1730, 1734, 1739, 1741, 1744,
        1748, 1751, 1756, 1761, 1763, 1772, 1776, 1779, 1782,
        1786, 1789, 1793, 1800, 1802, 1804, 1810, 1814, 1819,
        1829, 1835, 1841, 1845, 1847, 1852, 1856, 1859, 1873,
        1877, 1882, 1886, 1891, 1894, 1900, 1904, 1911, 1914,
        1920, 1923, 1927, 1933, 1939, 1948, 1951, 1955, 1961,
        1964, 1971, 1973, 1977, 1979, 1987, 1992, 1997, 2001,
        2005, 2009, 2013, 2020, 2023, 2027, 2029, 2033, 2038,
        2046, 2049, 2053, 2056, 2059, 2064, 2066, 2075, 2078,
        2081, 2085, 2097, 2103, 2107, 2112, 2116, 2120, 2126,
        2131, 2138, 2141, 2146, 2152, 2156, 2158, 2163, 2165,
        2170, 2176, 2182, 2183, 2189, 2192, 2198, 2206, 2210,
        2214, 2216, 2219, 2223, 2227, 2231, 2234, 2237, 2246,
        2250, 2254, 2255, 2259, 2264, 2268, 2270, 2274, 2280,
        2286, 2290, 2293, 2303, 2308, 2314, 2319, 2322, 2324,
        2329, 2333, 2340, 2345, 2347, 2355, 2361, 2367, 2374,
        2379, 2382, 2386, 2393, 2397, 2403, 2409, 2418, 2426,
        2435, 2440, 2447, 2451, 2456, 2459, 2468, 2475, 2480,
        2482, 2486, 2488, 2492, 2497, 2504, 2510, 2522, 2529,
        2536, 2542, 2548, 2552, 2559, 2562, 2568, 2574, 2582,
        2590, 2593, 2599, 2610, 2617, 2622, 2628, 2634, 2641,
        2650, 2656, 2662, 2669, 2676, 2679, 2691, 2695, 2701,
        2707, 2716, 2719, 2731, 2743, 2751, 2758, 2766, 2776,
        2784, 2795, 2803, 2822, 2832, 2847, 2860, 2871, 2882,
        2895, 2906, 2928, 2951, 2967, 2983, 3002, 3017, 3022,
        3026, 3043, 3052, 3073, 3101, 3123, 3175, 3182, 3204,
        3236, 3251, 3280, 3317, 3336, 3364, 3403, 3435, 3464,
        3535, 3553, 3578, 3627, 3680, 3740, 3766, 3806, 3875,
        3976, 4041, 4115, 4138, 4363, 4443, 4655, 4802, 4953,
        5192, 5294, 5526, 6046, 6337, 6699, 7079, 7271, 8000])

    right_target_cf = np.array([161, 258, 323, 348, 364, 374,
        382, 390, 392, 396, 399, 402, 405, 408, 411, 416, 423,
        430, 436, 448, 454, 461, 464, 472, 475, 480, 484, 487,
        493, 496, 500, 503, 509, 511, 514, 517, 520, 523, 525,
        527, 530, 531, 532, 535, 535, 537, 539, 541, 543, 545,
        547, 548, 550, 550, 551, 553, 554, 555, 556, 557, 558,
        559, 559, 560, 560, 561, 562, 562, 563, 564, 564, 565,
        565, 566, 566, 567, 568, 568, 569, 570, 570, 571, 572,
```

572, 572, 574, 574, 575, 575, 575, 576, 576, 576, 577,
577, 578, 578, 578, 579, 579, 580, 580, 581, 581, 581,
581, 582, 582, 582, 583, 584, 584, 584, 585, 585, 586,
586, 587, 587, 587, 588, 588, 589, 589, 589, 590, 590,
590, 591, 591, 592, 592, 592, 593, 593, 594, 594, 595,
595, 595, 596, 596, 597, 597, 597, 598, 598, 598, 598,
599, 599, 599, 600, 600, 600, 601, 601, 601, 602, 602,
602, 603, 603, 603, 603, 603, 604, 604, 604, 604, 604,
605, 605, 605, 606, 606, 606, 606, 606, 607, 607, 607,
607, 608, 608, 608, 608, 609, 609, 609, 609, 610, 610,
610, 611, 611, 611, 611, 612, 612, 612, 612, 613, 613,
613, 613, 614, 614, 614, 614, 614, 615, 615, 615, 616,
616, 616, 617, 617, 617, 617, 618, 618, 618, 618, 618,
619, 619, 619, 620, 620, 620, 620, 620, 621, 621, 621,
622, 622, 622, 623, 623, 623, 623, 623, 624, 624, 624,
625, 625, 625, 625, 626, 626, 626, 627, 627, 627, 627,
627, 628, 628, 628, 628, 629, 629, 629, 629, 630, 630,
630, 630, 631, 631, 631, 631, 632, 632, 632, 632, 632,
633, 633, 633, 633, 633, 634, 634, 634, 634, 635, 635,
635, 636, 636, 636, 636, 636, 636, 637, 637, 637, 637,
638, 638, 638, 638, 639, 639, 639, 639, 640, 640, 640,
641, 641, 641, 642, 642, 642, 642, 643, 643, 643, 644,
644, 644, 644, 645, 645, 645, 645, 646, 646, 646, 646,
646, 646, 647, 647, 647, 648, 648, 648, 649, 649, 649,
649, 650, 650, 650, 650, 650, 651, 651, 651, 651, 652,
652, 652, 652, 653, 653, 653, 654, 654, 654, 654, 655,
655, 655, 655, 656, 656, 656, 656, 657, 657, 657, 657,
658, 658, 658, 659, 659, 659, 659, 659, 660, 660, 660,
660, 661, 661, 661, 661, 662, 662, 662, 662, 663, 663,
663, 663, 663, 664, 664, 664, 664, 665, 665, 665, 665,
665, 666, 666, 666, 666, 666, 667, 667, 667, 667, 667,
668, 668, 668, 668, 668, 669, 669, 669, 669, 669, 670,
670, 670, 671, 671, 671, 671, 671, 672, 672, 672, 672,
673, 673, 673, 673, 673, 673, 674, 674, 674, 675, 675,
675, 675, 676, 676, 676, 676, 677, 677, 677, 677, 678,
678, 678, 678, 678, 679, 679, 679, 679, 680, 680, 680,
680, 680, 681, 681, 681, 682, 682, 682, 682, 682, 683,
683, 683, 683, 684, 684, 684, 684, 684, 685, 685, 685,
685, 686, 686, 686, 686, 687, 687, 687, 687, 687, 688,
688, 688, 688, 689, 689, 689, 689, 689, 690, 690, 690,
691, 691, 691, 691, 691, 692, 692, 692, 692, 692, 692,
693, 693, 693, 693, 694, 694, 694, 694, 694, 695, 695,
695, 695, 695, 696, 696, 696, 696, 697, 697, 697, 697,
697, 698, 698, 698, 698, 699, 699, 699, 699, 700, 700,
700, 700, 700, 700, 701, 701, 701, 701, 702, 702, 702,
702, 702, 702, 703, 703, 703, 703, 703, 703, 703, 704,
704, 704, 704, 704, 704, 705, 705, 705, 705, 705, 706,
706, 706, 706, 707, 707, 707, 707, 707, 707, 708, 708,
708, 708, 708, 708, 709, 709, 709, 709, 709, 710, 710,
710, 710, 710, 710, 711, 711, 711, 711, 711, 711, 712,
712, 712, 712, 712, 713, 713, 713, 713, 714, 714, 714,
714, 714, 715, 715, 715, 715, 715, 716, 716, 716, 716,
716, 717, 717, 717, 717, 717, 717, 718, 718, 718, 718,
718, 719, 719, 719, 719, 719, 720, 720, 720, 720, 720,
721, 721, 721, 721, 721, 721, 721, 721, 722, 722, 722,
722, 723, 723, 723, 723, 723, 724, 724, 724, 724, 724,
724, 725, 725, 725, 725, 725, 726, 726, 726, 726, 726,
727, 727, 727, 727, 727, 728, 728, 728, 728, 728, 729,
729, 729, 729, 729, 730, 730, 730, 730, 730, 730, 731,
731, 731, 731, 731, 731, 731, 732, 732, 732, 732, 733,
733, 733, 733, 734, 734, 734, 734, 735, 735, 735, 735,
736, 736, 736, 736, 737, 737, 737, 737, 737, 738, 738,
738, 738, 738, 738, 739, 739, 739, 739, 739, 740, 740,
740, 740, 740, 740, 741, 741, 741, 741, 742, 742, 742,

742, 742, 743, 743, 743, 743, 744, 744, 744, 744, 744,
744, 745, 745, 745, 746, 746, 746, 746, 747, 747, 747,
747, 747, 748, 748, 748, 748, 749, 749, 749, 749, 750,
750, 750, 750, 750, 750, 751, 751, 751, 751, 751, 752,
752, 752, 753, 753, 753, 754, 754, 754, 754, 755, 755,
755, 755, 756, 756, 756, 756, 756, 756, 757, 757, 757,
757, 758, 758, 758, 758, 759, 759, 759, 759, 759, 759,
760, 760, 760, 760, 761, 761, 761, 761, 761, 762, 762,
762, 762, 762, 762, 763, 763, 763, 764, 764, 764, 764,
764, 764, 765, 765, 765, 765, 766, 766, 766, 766, 766,
767, 767, 767, 767, 768, 768, 768, 769, 769, 769, 769,
769, 770, 770, 770, 770, 770, 771, 771, 771, 771, 772,
772, 772, 772, 772, 773, 773, 773, 773, 774, 774, 774,
774, 774, 775, 775, 775, 775, 775, 776, 776, 776, 777,
777, 777, 777, 778, 778, 778, 778, 779, 779, 779, 779,
779, 779, 779, 780, 780, 780, 780, 781, 781, 781, 781,
781, 782, 782, 782, 782, 783, 783, 783, 783, 784, 784,
784, 784, 785, 785, 785, 785, 786, 786, 786, 786, 786,
787, 787, 787, 787, 788, 788, 788, 788, 789, 789, 789,
790, 790, 790, 790, 790, 791, 791, 791, 792, 792, 792,
792, 792, 793, 793, 793, 793, 793, 794, 794, 794, 795,
795, 795, 795, 795, 796, 796, 796, 796, 796, 797, 797,
797, 797, 798, 798, 798, 798, 799, 799, 799, 799, 799,
800, 800, 800, 800, 800, 800, 801, 801, 801, 802, 802,
802, 802, 802, 802, 803, 803, 803, 803, 804, 804, 804,
804, 804, 805, 805, 806, 806, 806, 806, 807, 807, 807,
807, 808, 808, 808, 808, 809, 809, 809, 810, 810, 810,
810, 811, 811, 811, 811, 812, 812, 812, 812, 813, 813,
813, 813, 814, 814, 814, 815, 815, 815, 815, 815, 816,
816, 816, 816, 817, 817, 817, 817, 818, 818, 818, 818,
818, 819, 819, 819, 819, 819, 820, 820, 820, 821, 821,
821, 821, 821, 822, 822, 822, 822, 823, 823, 823, 823,
824, 824, 824, 824, 825, 825, 825, 825, 825, 826, 826,
826, 826, 827, 827, 827, 827, 828, 828, 828, 828, 829,
829, 829, 829, 830, 830, 830, 830, 831, 831, 831, 832,
832, 832, 832, 833, 833, 833, 833, 834, 834, 835, 835,
835, 835, 836, 836, 836, 836, 837, 837, 837, 837, 838,
838, 838, 838, 839, 839, 839, 840, 840, 840, 841, 841,
841, 842, 842, 842, 843, 843, 843, 844, 844, 844, 844,
845, 845, 845, 846, 846, 846, 846, 847, 847, 847, 847,
847, 848, 848, 848, 849, 849, 849, 850, 850, 851, 851,
851, 851, 851, 851, 852, 852, 852, 853, 853, 853, 854,
854, 855, 855, 855, 855, 856, 856, 856, 857, 857, 857,
858, 858, 858, 859, 859, 859, 860, 860, 860, 861, 861,
861, 862, 862, 862, 863, 863, 863, 863, 864, 864, 864,
865, 865, 865, 866, 866, 866, 867, 867, 868, 868, 868,
869, 869, 869, 870, 870, 871, 871, 871, 872, 873, 873,
873, 874, 874, 874, 875, 875, 876, 876, 877, 877, 878,
878, 878, 879, 879, 879, 880, 880, 881, 881, 882, 882,
882, 883, 883, 883, 884, 884, 884, 885, 885, 886, 886,
887, 887, 888, 888, 888, 889, 889, 889, 890, 890, 890,
891, 892, 892, 893, 893, 894, 894, 894, 895, 895, 896,
897, 897, 898, 898, 899, 899, 900, 900, 901, 901, 901,
902, 902, 903, 904, 904, 905, 905, 906, 906, 906, 907,
907, 908, 908, 909, 909, 909, 910, 910, 911, 912, 913,
913, 914, 914, 915, 915, 916, 916, 917, 918, 919, 919,
920, 921, 922, 922, 923, 923, 924, 924, 925, 926, 926,
927, 928, 928, 929, 929, 930, 930, 931, 931, 932, 932,
932, 933, 934, 934, 935, 936, 937, 938, 940, 940, 941,
942, 943, 944, 944, 945, 945, 946, 947, 948, 949, 950,
950, 951, 952, 953, 953, 954, 954, 955, 956, 957, 957,
958, 959, 960, 961, 962, 963, 963, 964, 966, 966, 968,
969, 970, 971, 972, 973, 975, 976, 976, 977, 978, 979,
980, 981, 982, 983, 984, 985, 986, 988, 989, 991, 992,

```
    994, 995, 995, 996, 997, 998, 999, 1001, 1003, 1005, 1007,
    1008, 1009, 1010, 1010, 1011, 1012, 1014, 1015, 1016,
    1018, 1019, 1021, 1022, 1024, 1025, 1027, 1028, 1030,
    1031, 1034, 1036, 1037, 1039, 1041, 1044, 1046, 1048,
    1050, 1052, 1053, 1055, 1056, 1058, 1060, 1061, 1062,
    1063, 1065, 1067, 1069, 1071, 1074, 1075, 1076, 1077,
    1079, 1080, 1082, 1082, 1084, 1086, 1087, 1090, 1091,
    1093, 1094, 1096, 1100, 1102, 1105, 1106, 1108, 1111,
    1113, 1115, 1117, 1118, 1120, 1122, 1123, 1124, 1125,
    1127, 1129, 1131, 1133, 1134, 1136, 1139, 1141, 1146,
    1151, 1152, 1154, 1155, 1160, 1162, 1165, 1166, 1169,
    1170, 1172, 1173, 1175, 1176, 1178, 1181, 1184, 1187,
    1189, 1191, 1193, 1196, 1201, 1204, 1205, 1208, 1209,
    1213, 1216, 1218, 1220, 1222, 1225, 1226, 1228, 1231,
    1232, 1234, 1238, 1241, 1244, 1248, 1250, 1254, 1256,
    1260, 1262, 1265, 1268, 1272, 1275, 1278, 1280, 1282,
    1285, 1290, 1295, 1296, 1299, 1301, 1304, 1306, 1310,
    1312, 1314, 1319, 1321, 1325, 1328, 1333, 1337, 1342,
    1349, 1355, 1360, 1363, 1367, 1374, 1377, 1380, 1384,
    1386, 1390, 1393, 1400, 1404, 1408, 1414, 1417, 1420,
    1423, 1430, 1432, 1437, 1441, 1444, 1450, 1452, 1459,
    1463, 1465, 1473, 1475, 1480, 1483, 1487, 1491, 1496,
    1502, 1505, 1508, 1511, 1517, 1520, 1524, 1526, 1532,
    1535, 1539, 1544, 1549, 1552, 1556, 1560, 1565, 1572,
    1576, 1580, 1585, 1590, 1599, 1604, 1607, 1612, 1616,
    1620, 1627, 1633, 1640, 1644, 1651, 1656, 1661, 1666,
    1675, 1680, 1684, 1689, 1693, 1698, 1701, 1706, 1712,
    1717, 1720, 1727, 1733, 1735, 1740, 1747, 1753, 1757,
    1759, 1768, 1774, 1780, 1786, 1789, 1793, 1798, 1802,
    1805, 1807, 1809, 1812, 1815, 1818, 1821, 1827, 1832,
    1837, 1841, 1844, 1849, 1854, 1859, 1863, 1868, 1873,
    1876, 1880, 1889, 1894, 1904, 1907, 1912, 1920, 1925,
    1930, 1937, 1940, 1943, 1946, 1952, 1958, 1962, 1966,
    1971, 1974, 1977, 1979, 1984, 1989, 1996, 1999, 2007,
    2012, 2018, 2021, 2027, 2033, 2040, 2045, 2055, 2059,
    2065, 2069, 2071, 2075, 2080, 2085, 2089, 2098, 2102,
    2109, 2115, 2119, 2122, 2128, 2131, 2137, 2141, 2150,
    2154, 2161, 2163, 2169, 2176, 2182, 2187, 2192, 2197,
    2200, 2205, 2209, 2218, 2222, 2226, 2229, 2234, 2240,
    2249, 2255, 2256, 2261, 2265, 2273, 2276, 2284, 2288,
    2290, 2295, 2299, 2307, 2310, 2317, 2325, 2329, 2334,
    2339, 2346, 2351, 2356, 2365, 2377, 2382, 2388, 2393,
    2404, 2413, 2421, 2427, 2435, 2441, 2448, 2457, 2461,
    2471, 2483, 2487, 2494, 2506, 2514, 2520, 2525, 2530,
    2538, 2549, 2560, 2568, 2574, 2582, 2590, 2596, 2605,
    2608, 2614, 2625, 2628, 2636, 2643, 2650, 2655, 2666,
    2680, 2688, 2695, 2709, 2726, 2736, 2744, 2749, 2755,
    2764, 2771, 2779, 2794, 2808, 2817, 2825, 2831, 2841,
    2852, 2859, 2867, 2876, 2887, 2893, 2902, 2908, 2915,
    2926, 2935, 2941, 2951, 2960, 2964, 2971, 2983, 2997,
    3004, 3016, 3020, 3030, 3035, 3044, 3050, 3054, 3065,
    3070, 3081, 3085, 3091, 3100, 3107, 3111, 3116, 3126,
    3132, 3136, 3147, 3158, 3161, 3174, 3183, 3192, 3197,
    3203, 3210, 3218, 3230, 3231, 3242, 3256, 3274, 3286,
    3299, 3307, 3321, 3330, 3339, 3354, 3370, 3391, 3419,
    3440, 3460, 3479, 3509, 3530, 3568, 3601, 3629, 3652,
    3697, 3718, 3747, 3819, 3899, 3985, 4058, 4134, 4239,
    4402, 4768, 5631, 7805])

short_spike_matrix = scale_down(spike_matrix=spike_matrix,
    left_cf=left_cf, right_cf=right_cf, left_target_cf=
    left_target_cf, right_target_cf=right_target_cf,
    scaling_factor=scaling_factor)
```

```
filename = target_folder + 'bio_1l2m5h_tal1/sam' +
   str(sample_id) + '_' + speaker + '_' + str(digit) + '.csv'

np.savetxt(filename, short_spike_matrix, fmt='%i',
   delimiter=",")

# TAL_by_2
left_target_cf = np.array([166, 409, 444, 471, 498, 509,
   518, 525, 532, 540, 546, 552, 559, 564, 570, 574, 579,
   582, 586, 589, 592, 596, 599, 602, 604, 607, 610, 612,
   615, 617, 620, 622, 625, 627, 628, 630, 632, 634, 635,
   637, 638, 639, 641, 643, 645, 646, 648, 649, 650, 652,
   653, 654, 656, 657, 658, 659, 661, 662, 663, 664, 665,
   667, 668, 669, 670, 672, 673, 674, 675, 677, 678, 680,
   681, 683, 684, 686, 687, 688, 689, 691, 693, 694, 695,
   696, 698, 699, 701, 702, 703, 705, 706, 708, 709, 710,
   712, 713, 714, 716, 718, 719, 720, 722, 724, 725, 727,
   728, 730, 731, 733, 734, 735, 737, 739, 740, 741, 743,
   745, 746, 748, 749, 751, 753, 754, 756, 758, 759, 761,
   763, 764, 765, 767, 768, 770, 772, 774, 775, 777, 779,
   780, 782, 784, 785, 787, 789, 790, 792, 793, 795, 796,
   797, 799, 800, 802, 804, 805, 807, 808, 810, 811, 813,
   814, 816, 818, 820, 822, 823, 826, 829, 831, 834, 836,
   839, 842, 844, 847, 850, 853, 856, 860, 864, 867, 871,
   875, 879, 882, 888, 893, 897, 901, 906, 912, 918, 925,
   932, 939, 946, 954, 962, 969, 978, 991, 1004, 1015, 1027,
   1040, 1052, 1073, 1085, 1099, 1113, 1128, 1143, 1161,
   1180, 1199, 1213, 1232, 1252, 1270, 1289, 1313, 1334,
   1355, 1379, 1403, 1428, 1452, 1476, 1497, 1524, 1545,
   1569, 1602, 1631, 1657, 1688, 1716, 1743, 1776, 1807,
   1850, 1892, 1929, 1971, 2007, 2038, 2075, 2116, 2154,
   2186, 2222, 2254, 2288, 2325, 2368, 2416, 2460, 2497,
   2552, 2598, 2654, 2707, 2774, 2868, 2991, 3111, 3323,
   3594, 4055, 5196, 8000])

right_target_cf = np.array([161, 391, 421, 474, 508, 529,
   542, 553, 560, 565, 570, 575, 578, 581, 584, 587, 590,
   594, 597, 599, 602, 604, 606, 608, 609, 612, 614, 616,
   618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 637,
   639, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660,
   662, 664, 665, 667, 668, 670, 672, 674, 676, 678, 680,
   682, 683, 685, 687, 689, 691, 692, 694, 696, 697, 699,
   701, 702, 703, 705, 707, 708, 710, 711, 712, 714, 716,
   717, 719, 720, 721, 723, 724, 726, 728, 729, 731, 732,
   734, 736, 737, 739, 740, 742, 744, 746, 747, 749, 751,
   753, 755, 757, 758, 760, 762, 763, 765, 767, 769, 771,
   772, 774, 776, 778, 779, 781, 783, 785, 787, 788, 791,
   793, 795, 796, 798, 800, 801, 803, 805, 807, 809, 811,
   813, 815, 817, 819, 821, 823, 825, 827, 828, 831, 833,
   835, 837, 839, 842, 844, 846, 848, 851, 853, 856, 859,
   861, 863, 866, 869, 873, 876, 879, 882, 885, 889, 892,
   896, 900, 904, 908, 912, 916, 922, 926, 931, 935, 943,
   949, 954, 961, 969, 977, 985, 995, 1007, 1015, 1026, 1040,
   1055, 1068, 1081, 1094, 1112, 1124, 1140, 1163, 1176,
   1196, 1217, 1234, 1260, 1281, 1306, 1332, 1371, 1401,
   1434, 1468, 1502, 1533, 1568, 1608, 1653, 1693, 1733,
   1773, 1807, 1837, 1871, 1917, 1956, 1988, 2029, 2073,
   2116, 2155, 2199, 2235, 2279, 2318, 2366, 2427, 2486,
   2548, 2608, 2665, 2747, 2821, 2891, 2955, 3024, 3083,
   3134, 3199, 3274, 3379, 3568, 3904, 7805])

short_spike_matrix = scale_down(spike_matrix=spike_matrix,
   left_cf=left_cf, right_cf=right_cf, left_target_cf=
```

```
        left_target_cf, right_target_cf=right_target_cf,
        scaling_factor=scaling_factor)

    filename = target_folder + 'bio_112m5h_tal2/sam' +
        str(sample_id) + '_' + speaker + '_' + str(digit) + '.csv'

    np.savetxt(filename, short_spike_matrix, fmt='%i',
        delimiter=",")

    # TAL_by_3
    left_target_cf = np.array([166, 487, 524, 547, 568, 583,
        594, 604, 613, 622, 629, 634, 639, 645, 650, 655, 659,
        663, 668, 672, 676, 681, 686, 691, 695, 700, 705, 710,
        714, 719, 725, 730, 735, 740, 745, 750, 756, 762, 766,
        772, 778, 784, 789, 795, 799, 805, 810, 815, 822, 830,
        838, 848, 859, 871, 885, 901, 920, 945, 972, 1011, 1053,
        1106, 1160, 1219, 1284, 1358, 1441, 1523, 1614, 1714,
        1817, 1957, 2077, 2209, 2323, 2479, 2639, 2882, 3453,
        8000])

    right_target_cf = np.array([161, 487, 551, 570, 582, 593,
        603, 609, 616, 623, 630, 636, 643, 650, 656, 663, 668,
        675, 681, 687, 693, 699, 704, 709, 714, 719, 724, 729,
        735, 740, 746, 752, 758, 764, 770, 776, 782, 788, 795,
        800, 807, 814, 820, 827, 834, 842, 849, 857, 866, 877,
        888, 900, 914, 930, 950, 974, 1005, 1046, 1089, 1140,
        1206, 1277, 1376, 1486, 1607, 1740, 1859, 1989, 2130,
        2268, 2436, 2633, 2884, 3092, 3331, 7805])

    short_spike_matrix = scale_down(spike_matrix=spike_matrix,
        left_cf=left_cf, right_cf=right_cf, left_target_cf=
        left_target_cf, right_target_cf=right_target_cf,
        scaling_factor=scaling_factor)

    filename = target_folder + 'bio_112m5h_tal3/sam' +
        str(sample_id) + '_' + speaker + '_' + str(digit) + '.csv'

    np.savetxt(filename, short_spike_matrix, fmt='%i',
        delimiter=",")

    # TAL_by_4
    left_target_cf = np.array([166, 534, 580, 606, 626, 639,
        652, 662, 672, 684, 695, 706, 718, 731, 743, 755, 768,
        783, 795, 808, 821, 841, 867, 902, 957, 1047, 1172, 1333,
        1526, 1750, 2055, 2351, 2758, 8000])

    right_target_cf = np.array([161, 561, 592, 611, 627, 644,
        660, 674, 689, 703, 715, 727, 740, 755, 769, 784, 799,
        815, 830, 849, 870, 898, 933, 993, 1092, 1236, 1475, 1788,
        2097, 2466, 3017, 7805])

    short_spike_matrix = scale_down(spike_matrix=spike_matrix,
        left_cf=left_cf, right_cf=right_cf, left_target_cf=
        left_target_cf, right_target_cf=right_target_cf,
        scaling_factor=scaling_factor)

    filename = target_folder + 'bio_112m5h_tal4/sam' +
        str(sample_id) + '_' + speaker + '_' + str(digit) + '.csv'

    np.savetxt(filename, short_spike_matrix, fmt='%i',
        delimiter=",")

    # TAL_by_5
```

```
    left_target_cf = np.array([166, 578, 624, 650, 670, 692,
        713, 737, 763, 789, 813, 849, 916, 1078, 1362, 1784, 2375,
        8000])
    right_target_cf = np.array([161, 590, 625, 657, 686, 711,
        735, 763, 792, 821, 856, 907, 1012, 1273, 1815, 2498,
        7805])
    short_spike_matrix = scale_down(spike_matrix=spike_matrix,
        left_cf=left_cf, right_cf=right_cf, left_target_cf=
        left_target_cf, right_target_cf=right_target_cf,
        scaling_factor=scaling_factor)
    filename = target_folder + 'bio_112m5h_tal5/sam' +
        str(sample_id) + '_' + speaker + '_' + str(digit) + '.csv'
    np.savetxt(filename, short_spike_matrix, fmt='%i',
        delimiter=",")
    # TAL_by_6
    left_target_cf = np.array([166, 612, 658, 694, 733, 775,
        818, 906, 1232, 1971, 8000])
    right_target_cf = np.array([161, 618, 673, 719, 766, 818,
        888, 1108, 1951, 7805])
    short_spike_matrix = scale_down(spike_matrix=spike_matrix,
        left_cf=left_cf, right_cf=right_cf, left_target_cf=
        left_target_cf, right_target_cf=right_target_cf,
        scaling_factor=scaling_factor)
    filename = target_folder + 'bio_112m5h_tal6/sam' +
        str(sample_id) + '_' + speaker + '_' + str(digit) + '.csv'
    np.savetxt(filename, short_spike_matrix, fmt='%i',
        delimiter=",")
    # TAL_by_7
    left_target_cf = np.array([166, 646, 706, 775, 867, 1428,
        8000])
    right_target_cf = np.array([161, 646, 719, 792, 888, 1430,
        7805])
    short_spike_matrix = scale_down(spike_matrix=spike_matrix,
        left_cf=left_cf, right_cf=right_cf, left_target_cf=
        left_target_cf, right_target_cf=right_target_cf,
        scaling_factor=scaling_factor)
    filename = target_folder + 'bio_112m5h_tal7/sam' +
        str(sample_id) + '_' + speaker + '_' + str(digit) + '.csv'
    np.savetxt(filename, short_spike_matrix, fmt='%i',
        delimiter=",")
    # TAL_by_8
    left_target_cf = np.array([166, 675, 775, 1022, 8000])
    right_target_cf = np.array([161, 686, 792, 1012, 7805])
    short_spike_matrix = scale_down(spike_matrix=spike_matrix,
        left_cf=left_cf, right_cf=right_cf, left_target_cf=
        left_target_cf, right_target_cf=right_target_cf,
        scaling_factor=scaling_factor)
```

```
        filename = target_folder + 'bio_112m5h_tal8/sam' +
           str(sample_id) + '_' + speaker + '_' + str(digit) + '.csv'

        np.savetxt(filename, short_spike_matrix, fmt='%i',
           delimiter=",")

        # TAL_by_9
        left_target_cf = np.array([166, 706, 867, 8000])
        right_target_cf = np.array([161, 719, 888, 7805])

        short_spike_matrix = scale_down(spike_matrix=spike_matrix,
           left_cf=left_cf, right_cf=right_cf, left_target_cf=
           left_target_cf, right_target_cf=right_target_cf,
           scaling_factor=scaling_factor)

        filename = target_folder + 'bio_112m5h_tal9/sam' +
           str(sample_id) + '_' + speaker + '_' + str(digit) + '.csv'

        np.savetxt(filename, short_spike_matrix, fmt='%i',
                   delimiter=",")

        # TAL_by_10
        left_target_cf = np.array([166, 775, 8000])
        right_target_cf = np.array([161, 792, 7805])

        short_spike_matrix = scale_down(spike_matrix=spike_matrix,
           left_cf=left_cf, right_cf=right_cf, left_target_cf=
           left_target_cf, right_target_cf=right_target_cf,
           scaling_factor=scaling_factor)

        filename = target_folder + 'bio_112m5h_tal10/sam' +
                   str(sample_id) + '_' + speaker + '_' +
                   str(digit) + '.csv'

        np.savetxt(filename, short_spike_matrix, fmt='%i',
                   delimiter=",")

    print "Done."
print "Successfully created", str(sample_id), "samples."
```

LISTING III SELECT_AUDITORY_INPUT_COORDINATES.M

```matlab
function [ input_coordinates, indices ] = select_auditory_input_coordinates ( coordinates, number_of_inputs, by_value, plot_result )
% selects a number of evenly spaced points from a list
% if by_value is true, selects them uniformly by value, assuming they are sorted
% otherwise just selects uniformly from list

if (number_of_inputs < 1)
    error('Must select at least one input neuron!');
end

number_of_coordinates = size(coordinates, 1);

if (number_of_inputs == 1)
    % we just return the middle point
    indices = int16(number_of_coordinates / 2);
end

if (number_of_inputs == 2)
    % we just return the first and last point
    indices = [1, number_of_coordinates];
end

if (number_of_inputs > 2)
    % stepwise selection

    if (by_value)
        % select uniformly based on values
        mini = min(coordinates(:, 4));
        maxi = max(coordinates(:, 4));
        values = mini:(maxi-mini) / (number_of_inputs - 1):maxi;

        indices = zeros(1, size(values, 2), 'int16');
        my_copy = coordinates(:, 4);

        for i = 1:size(values, 2)
            [~, idx] = min(abs(my_copy - values(i)));
            indices(i) = idx;
            my_copy(idx) = 10000;
        end
    else
        % select uniformly based on (sorted) index
        indices = int16(1:(number_of_coordinates - 1) / (number_of_inputs - 1):number_of_coordinates);
    end

    % just making sure we have the correct number
    if (length(indices) < number_of_inputs)
        indices = [indices, number_of_coordinates];
    end
end

input_coordinates = coordinates(indices, :);

if (plot_result)
    figure('Color','w', 'NumberTitle', 'off', 'Name', 'Selected input coordinates');
    plot(coordinates(:, 4), 1:number_of_coordinates);
    hold on;
    scatter(input_coordinates(:, 4), indices, 'filled')
```

```
    if (by_value)
        title('Selected indices by value');
    else
        title('Selected indices by index');
    end

    xlabel('Values')
    ylabel('Indices')
end

end
```

LISTING IV CONVERT_JESTER_JPG_TO_MP4.PY

```python
# run this code in the directory where the folders with the jpgs are located

import cv2 as cv  # OpenCV library
import os

codec = cv.VideoWriter_fourcc('m','p','4','v')
fps = 12.0
folders = os.listdir()  # or put in here the absolute path, preferably using os.path.join()

for folder in folders:
    files = sorted(os.listdir(folder))
    images = []

    for file in files:
        if file.endswith("jpg"):
            images.append(file)

    if not images:
        print("Folder " + folder + " does not contain jpg images")
        continue

    frame = cv.imread(os.path.join(folder, images[0]))
    height, width, channels = frame.shape
    video = cv.VideoWriter(os.path.join("0_videos", folder)
                           + ".mp4", codec, fps, (width, height))

    for image in images:
        frame = cv.imread(os.path.join(folder, image))
        video.write(frame)

    video.release()
```

LISTING V ENCODE_VISION_SAMPLES.PY

```python
# -*- coding: utf-8 -*-

"""
Converts video into spikes using retina-inspired encoding.
Subsequent frames are compared with each other pixel by pixel.
A frame is split into blocks, imitating peripheral vision.
These blocks are converted into greyscale to compare brightness.
The foveal area of the image is defined as the block with the most activity.
This focus block is assessed using colour vision.

How spikes are created in peripheral blocks:
When a pixel's brightness changes by more than a defined threshold between frames, a spike is created for this pixel’s location.
If there are more than a defined percentage of spikes in a block, the block creates a spike.

How spikes are created in the focus/foveal block:
For each pixel, the BGR values are transformed into CIELAB colour space. Then the CIEDE2000 formula is used to compute the delta E of the colours.
If delta E is larger than a defined threshold, this pixel creates a spike.
If the fovea pixels are to be summarised in blocks, a block creates a spike if there are more than a defined number of spikes in a block.

@author: Anne Wendt
"""


import os
import sys
import time

import colour.difference as cd
import cv2  # OpenCV library
import matplotlib.pyplot as plt
import numpy as np


#################################################################
###                      GLOBAL PARAMETERS                    ###
#################################################################


DIRECTORY = 'randomly_selected_subset'
DIRECTORY2 = 'encoded_subset'
FILETYPE = '.mp4'  # only these files in the directory will be encoded

TEST_RUN = False  # play videos more slowly and do not close windows at the end

# these thresholds are used when comparing subsequent frames on a pixel base
# if they differ enough, a spike is created
PIXEL_THRESHOLD = 3  # applies to the outer blocks in greyscale = intensity difference
FOVEA_THRESHOLD = 5  # applies to the central block in colour = DELTA E
```

```
# how many blocks do we want in each column and row for each block level
#              c  r
# BLOCKS = [[3, 3],  # block 0 (outermost periphery)
#            [6, 5],  # block 1
#            [5, 4],  # block 2
#            [7, 5],  # block 3
#            [8, 8]]  # block 4 (fovea)


# BLOCKS = [[10, 6], [15, 10], [23, 15], [35, 23], [55, 36], [84, 56], [94, 85]]  # tal_orig
# BLOCKS = [[6, 4], [8, 6], [12, 9], [17, 13], [30, 20], [34, 30]]  # tal_by_2
# BLOCKS = [[4, 4], [6, 5], [7, 10], [16, 12], [18, 16]]  # tal_by_3
# BLOCKS = [[3, 2], [4, 4], [6, 5], [9, 7], [12, 11]]  # tal_by_4
# BLOCKS = [[3, 3], [5, 4], [7, 5], [8, 8]]  # tal_by_5
# BLOCKS = [[2, 2], [4, 3], [6, 4], [6, 6]]  # tal_by_6
# BLOCKS = [[2, 2], [3, 2], [4, 3], [5, 5]]  # tal_by_7
# BLOCKS = [[3, 2], [3, 3], [4, 4]]  # tal_by_8
# BLOCKS = [[2, 2], [3, 2], [4, 3]]  # tal_by_9
# BLOCKS = [[3, 1], [2, 2], [3, 3]]  # tal_by_10
# BLOCKS = [[9, 6], [15, 10], [23, 16], [36, 25], [64, 37], [95, 66], [106, 96]]  # mni_times_2
# BLOCKS = [[5, 4], [8, 6], [15, 9], [21, 15], [33, 22], [38, 34]]  # mni_orig
# BLOCKS = [[3, 2], [6, 4], [7, 5], [12, 8], [13, 12]]  # mni_by_2
# BLOCKS = [[3, 2], [5, 3], [6, 5], [7, 6]]  # mni_by_3
# BLOCKS = [[2, 2], [3, 2], [3, 3], [5, 4]]  # mni_by_4
BLOCKS = [[2, 2], [2, 2], [4, 3]]  # mni_by_5


# threshold for a periphery block to be counted as a spike
BLOCK_THRESHOLD = 0.3

# threshold for a fovea block to be counted as a spike
BLOCK_FOVEA_THRESHOLD = 0.5


# how many times smaller or larger than its direct neighbours is each block
BLOCK_SCALING_FACTOR = 4 ** (1 / (len(BLOCKS) - 1))
if TEST_RUN:
    print("Block scaling factor:", BLOCK_SCALING_FACTOR)


# parameters that will be used to draw the blocks
# OpenCV uses BGR colour format!
COLOURS = [(255, 0, 0),  # level 0 - blue
           (255, 255, 0),  # level 1 - cyan
           (0, 255, 0),  # level 2 - green
           (0, 255, 255),  # level 3 - yellow
           (0, 128, 255),  # level 4 - orange
           (0, 0, 255)]  # level 5 - red

LINE_WIDTH = 1  # in pixels


####################################################################
###                          FUNCTIONS                           ###
####################################################################
```

```python
def clean_up_and_exit(capture, exit_message):
    """
    Releases the video capture, closes all open windows, and exits
    the program with a message.
    """

    capture.release()
    if not TEST_RUN:
        cv2.destroyAllWindows()
    sys.exit(exit_message)


def draw_boundaries(frame, block_info, boundaries):
    """
    Draws the block boundaries onto the frame.
    """

    # draw boundaries for each block
    for i in range(len(BLOCKS)):
        for col in range(BLOCKS[i][0] + 1):
            x = (col * block_info['widths'][i]) + boundaries[i][0]
            frame = cv2.line(frame,
                             (x, boundaries[i][2]),
                             (x, boundaries[i][3]),
                             COLOURS[i],
                             LINE_WIDTH)

        for row in range(BLOCKS[i][1] + 1):
            y = (row * block_info['heights'][i]) + boundaries[i][2]
            frame = cv2.line(frame,
                             (boundaries[i][0], y),
                             (boundaries[i][1], y),
                             COLOURS[i],
                             LINE_WIDTH)

    return frame


def encode(filename):
    """
    Transforms a video file into a sample spike file and saves the
    sample.
    """

    print("Encoding " + filename, end='\t')

    # keep in mind that OpenCV uses BGR!
    capture = cv2.VideoCapture(os.path.join(DIRECTORY, filename))

    # print debug information
    if TEST_RUN:
        show_debug_information(capture)

    # check data
    if capture.get(cv2.CAP_PROP_FRAME_COUNT) < 3:
        clean_up_and_exit(capture,
                          "File must have more than two frames")

    # create output windows
    cv2.namedWindow("Original", cv2.WINDOW_AUTOSIZE)
    cv2.namedWindow("Pixel Spikes", cv2.WINDOW_AUTOSIZE)
    cv2.namedWindow("Fovea Spikes", cv2.WINDOW_AUTOSIZE)
```

```
    # read first frame
    success, frame_t0 = capture.read()
    if not success:
        clean_up_and_exit(capture, "Could not read first frame")

    # set initial focus coordinates to centre of image
    focus = (int(frame_t0.shape[1] / 2),
             int(frame_t0.shape[0] / 2))

    # calculate block size and position
    block_info = get_block_info(frame_t0.shape)

    # read second frame
    success, frame_t1 = capture.read()  # assume this one works if t0 worked

    # create output array
    sample = np.zeros((1, block_info['total_number']
                          + block_info['fovea_number']))

    while success:
        # convert frames to greyscale
        f0_grey = cv2.cvtColor(frame_t0, cv2.COLOR_BGR2GRAY)
        f1_grey = cv2.cvtColor(frame_t1, cv2.COLOR_BGR2GRAY)

        # calculate difference between pixel brightness
        pixel_spikes = get_frame_diff_as_spikes(f0_grey, f1_grey)

        # get current block boundaries based on current focus area
        boundaries = get_boundaries(frame_t0.shape,
                                    block_info, focus)

        # get spikes for blocks and update focus
        spikes, focus = get_block_spikes(pixel_spikes, block_info,
                                         boundaries, focus)

        # get colour difference for focus
        fovea_pixel_spikes = get_fovea_spikes(frame_t0, frame_t1,
                                              boundaries[-1])

        # summarise the spikes into their blocks
        fovea_spikes = get_block_fovea(fovea_pixel_spikes,
                                       block_info)

        # add fovea row to spike row
        spikes = np.append(spikes, fovea_spikes, axis=1)

        # add spike row to final output
        sample = np.append(sample, spikes, axis=0)

        # draw block boundaries
        frame_t0 = draw_boundaries(frame_t0, block_info,
                                   boundaries)

        # update display windows
        cv2.imshow("Original", frame_t0)
        cv2.imshow("Pixel Spikes", pixel_spikes * 255)
        cv2.imshow("Fovea Spikes", fovea_pixel_spikes * 255)

        # stop when Esc key is pressed
        if cv2.waitKey(10) == 27:
            break
```

```python
        # slow down processing to enable observation
        if TEST_RUN:
            time.sleep(0.1)

        # move to next image
        frame_t0 = frame_t1
        success, frame_t1 = capture.read()

    # remove the first line that we filled with zeros when creating the sample array
    sample = np.delete(sample, 0, axis=0)

    save_result_file(filename, sample)

    # some stats to optimise thresholds based on spike rate
    periphery_spike_rate = np.mean(sample[:,
                                    :block_info['total_number']+1])
    fovea_spike_rate = np.mean(sample[:,
                                    block_info['total_number']+1:])
    print(f'Periphery {periphery_spike_rate:.6f}
            Fovea {fovea_spike_rate:.6f}')

    return (periphery_spike_rate, fovea_spike_rate)


def get_block_fovea(fovea_pixel_spikes, block_info):
    """
    Summarises the fovea spikes into blocks to reduce the number of inputs.
    Returns row of spikes from left to right, then top to bottom.
    """

    # if we only have one fovea block, we want to keep all original pixels
    if BLOCKS[-1] == [1, 1]:
        return np.reshape(fovea_pixel_spikes, (1, -1))

    # create output row
    spikes = np.zeros((1, block_info['fovea_number']))

    # keep track of the current item in the output spike row
    block_index = 0

    # set dynamic threshold
    # threshold = 2 * np.mean(fovea_pixel_spikes)

    # debug info
    fovea_block_width = block_info['widths'][-1]
    fovea_block_height = block_info['heights'][-1]

    for col in range(BLOCKS[-1][0]):
        col_start = (col * fovea_block_width)
        col_end = col_start + fovea_block_width

        for row in range(BLOCKS[-1][1]):
            row_start = (row * fovea_block_height)
            row_end = row_start + fovea_block_height

            # calculate block's spike rate
            block_mean = np.mean(fovea_pixel_spikes
                            [row_start:row_end, col_start:col_end])

            # check if we need to create a spike
            # if block_mean > threshold:
```

```python
            if block_mean > BLOCK_FOVEA_THRESHOLD:
                spikes[0, block_index] = 1

            # increase counter to keep track of next spike index
            block_index += 1

    return spikes


def get_block_info(frame_shape):
    """
    Calculates the sizes of the blocks based on the size of the frame.
    """

    height = frame_shape[0]
    width = frame_shape[1]

    # count number of pixels per block and total number of blocks
    block_widths = []
    block_heights = []
    total_number_of_blocks = 0

    for i in range(len(BLOCKS)):
        block_widths.append(int(width /
                             ((BLOCK_SCALING_FACTOR ** i) * BLOCKS[i][0])))

        block_heights.append(int(height /
                             ((BLOCK_SCALING_FACTOR ** i) * BLOCKS[i][1])))

        total_number_of_blocks += BLOCKS[i][0] * BLOCKS[i][1]

    # remove fovea block
    total_number_of_blocks -= BLOCKS[-1][0] * BLOCKS[-1][1]

    # build return dict
    block_info = {'widths': block_widths, 'heights': block_heights,
                  'total_number': total_number_of_blocks}

    # distinguish between pixel-wise and block-wise fovea
    a, b = 0, 0
    if BLOCKS[-1] == [1, 1]:
        a = block_heights[-1]
        b = block_widths[-1]
    else:
        a = BLOCKS[-1][0]
        b = BLOCKS[-1][1]

    block_info['fovea_number'] = a * b

    # debug info
    if TEST_RUN:
        print("Number of input channels required for periphery: ",
              block_info['total_number'])
        print("Number of input channels required for fovea:",
              a, "x", b, "=", block_info['fovea_number'])

        m = block_heights[-1] * BLOCKS[-1][0]
        n = block_widths[-1] * BLOCKS[-1][1]

        print("Size of the fovea in pixels:", m, "x", n, "=", m*n)

    return block_info
```

```python
def get_block_spikes(pixel_spikes, block_info, boundaries, focus):
    """
    Calculates which block emits a spike. Also determines the most active region.
    Returns row of spikes going from outermost to innermost block and from left to right then from top to bottom.
    Returns coordinates of focus.
    """

    # create output row
    spikes = np.zeros((1, block_info['total_number']))

    # keep track of the current item in the output spike row
    block_index = 0

    # set dynamic threshold
    # threshold = 2 * np.mean(pixel_spikes)

    # keep track of most active region
    highest_spike_rate = 0.0

    for i in range(len(BLOCKS) - 1):
        for col in range(BLOCKS[i][0]):
            col_start = (col * block_info['widths'][i]) + boundaries[i][0]
            col_end = col_start + block_info['widths'][i]

            for row in range(BLOCKS[i][1]):
                row_start = (row * block_info['heights'][i]) + boundaries[i][2]
                row_end = row_start + block_info['heights'][i]



                # calculate block's spike rate
                block_mean = np.mean(pixel_spikes[
                                     row_start:row_end,
                                     col_start:col_end])

                # check if we need to create a spike
                # if block_mean > threshold:
                if block_mean > BLOCK_THRESHOLD:
                    spikes[0, block_index] = 1

                # increase counter to keep track of next spike
                block_index += 1

                # find block with most activity
                if block_mean > highest_spike_rate:
                    highest_spike_rate = block_mean
                    # set actvity centre to centre of block
                    focus = (int((col_start + col_end) / 2),
                             int((row_start + row_end) / 2))

    return spikes, focus


def get_boundaries(frame_shape, block_info, focus):
    """
    Calculates the current outer boundaries of the blocks.
    The blocks must not go over the edges of the frame.
    """
```

```
    max_col = frame_shape[1]
    max_row = frame_shape[0]

    widths = block_info['widths']
    heights = block_info['heights']

    # calculate block position for drawing
    boundaries = []  # one entry per block level -
                       start_col, end_col, start_row, end_row
    boundaries.append([0, max_col, 0, max_row])  # level 0 always
                                                  covers the full frame

    for i in range(1, len(BLOCKS)):
        start_col = int(focus[0]
                      - ((BLOCKS[i][0] / 2) * widths[i]))
        if start_col < 0:
            start_col = 0

        end_col = start_col + (BLOCKS[i][0] * widths[i])
        if end_col > max_col:
            start_col -= (end_col - max_col)
            end_col = max_col

        start_row = int(focus[1]
                      - ((BLOCKS[i][1] / 2) * heights[i]))
        if start_row < 0:
            start_row = 0

        end_row = start_row + (BLOCKS[i][1] * heights[i])
        if end_row > max_row:
            start_row -= (end_row - max_row)
            end_row = max_row

        boundaries.append([start_col, end_col, start_row, end_row])

    return boundaries


def get_fovea_spikes(frame_t0, frame_t1, fovea_boundaries):
    """
    Calculates if the colours of the pixels in the fovea block
    differ enough to create a spike.
    For this, it uses the CIEDE 2000 Delta E method.
    Returns a two-dimensional matrix of spikes.
    """

    # transform fovea area to CIELAB colour space
    # fovea boundaries are in order start_col, end_col, start_row,
      end_row
    fovea0_lab = cv2.cvtColor(frame_t0[
                        fovea_boundaries[2]:fovea_boundaries[3],
                        fovea_boundaries[0]:fovea_boundaries[1],
                                   :],
                           cv2.COLOR_BGR2Lab)
    fovea1_lab = cv2.cvtColor(frame_t1[
                        fovea_boundaries[2]:fovea_boundaries[3],
                        fovea_boundaries[0]:fovea_boundaries[1],
                                   :],
                           cv2.COLOR_BGR2Lab)

    # calculate delta E for each pixel
    delta_e_array = cd.delta_E_CIE2000(fovea0_lab, fovea1_lab)
```

```
    # set all elements > fovea_threshold to 1
    spikes = cv2.threshold(delta_e_array, FOVEA_THRESHOLD, 1,
                           cv2.THRESH_BINARY)[1]

    return spikes


def get_frame_diff_as_spikes(frame0, frame1):
    """
    Calculates the difference between two frame arrays.
    frame0 is the first frame and frame1 is the frame following
    frame0.
    Returns a difference matrix that has the same shape as frame0
    and frame1 - if they are in greyscale, they have rows and
    columns, and if they are in colour, they also have channels.
    (OpenCV uses BGR!)
    """

    # calculate absolute differences
    diff = cv2.absdiff(frame0, frame1)

    # set all elements > pixel_threshold to 1
    pixel_spikes = cv2.threshold(diff, PIXEL_THRESHOLD, 1,
                                 cv2.THRESH_BINARY)[1]

    return pixel_spikes


def save_result_file(filename, sample):
    """
    Saves the sample that was incrementally created.
    The file name is equal to the video name.
    """

    # save file as csv
    np.savetxt(os.path.join(DIRECTORY2, filename[:-4]) + '.csv',
       sample, fmt='%1.0f', delimiter=',', newline='\r\n')


def show_debug_information(capture):
    """
    Prints out some information about the current video.
    """

    print("Frames per second:", capture.get(cv2.CAP_PROP_FPS))
    print("Number of frames:",
          capture.get(cv2.CAP_PROP_FRAME_COUNT))
    print("Frame width:", capture.get(cv2.CAP_PROP_FRAME_WIDTH))
    print("Frame height:", capture.get(cv2.CAP_PROP_FRAME_HEIGHT))



###################################################################
###                        PROGRAM LOGIC                        ###
###################################################################


periphery_spike_rates = []
fovea_spike_rates = []
```

```
with os.scandir(DIRECTORY) as directory:
    for item in directory:
        if item.is_file() and item.name.endswith(FILETYPE):
            periphery_spike_rate, fovea_spike_rate =
                                                  encode(item.name)
            periphery_spike_rates.append(periphery_spike_rate)
            fovea_spike_rates.append(fovea_spike_rate)


print("Mean periphery spike rate:", np.mean(periphery_spike_rates))
print("Mean fovea spike rate:", np.mean(fovea_spike_rates))

fig, (ax0, ax1) = plt.subplots(nrows=2, sharex=True)

ax0.set_title("Fovea Spike Rates")
ax0.hist(fovea_spike_rates, bins=40)

ax1.set_title("Periphery Spike Rates")
ax1.hist(periphery_spike_rates, bins=40)

fig.tight_layout()
plt.show()
```

LISTING VI UNPACKING_RETINOTOPY_DATA.M

```
%% using grayordinates

load("osf data\atlas.mat")
wang2015 = wang2015 + 1; % indices are still 0-based but the rest
                           is not
load("osf data\prfresults.mat", "allresults")

% allresults(values, measurement type, participant, fit)
ang_all = squeeze(allresults(:,1,184,1));

% turn the whole thing so that 0 is at the top
ang_all_top_is_0 = ang_all - 90;
ang_all_top_is_0(ang_all_top_is_0 < 0) =
                      ang_all_top_is_0(ang_all_top_is_0 < 0) + 360;
ang_all_top_is_0 = 360 - ang_all_top_is_0;

ecc_all = squeeze(allresults(:,2,184,1));

%% take the coordinates apart using the HCP workbench command

% in console:
% wb_command -surface-coordinates-to-metric
%         S1200_7T_Retinotopy181.L.white_MSMAll.32k_fs_LR.surf.gii
%         left.func.gii

% wb_command -surface-coordinates-to-metric
%         S1200_7T_Retinotopy181.R.white_MSMAll.32k_fs_LR.surf.gii
%         right.func.gii

% wb_command -metric-convert -to-nifti
%         left.func.gii
%         left.nii

% wb_command -metric-convert -to-nifti
%         right.func.gii
%         right.nii

left_nii = niftiread("left.nii");
right_nii = niftiread("right.nii");

left_xyz_nii = squeeze(left_nii(:,1,1,:));
right_xyz_nii = squeeze(right_nii(:,1,1,:));

% these actually contain too many points -
% but files
% S1200_7T_Retinotopy181.Fit1_PolarAngle_MSMAll.32k_fs_LR.dscalar
% and
% S1200_7T_Retinotopy181.Fit1_Eccentricity_MSMAll.32k_fs_LR.dscalar
% contain a list of vertices for the left and right hemispheres
% so copy those indices from one of the files (they're identical)

left_vertex_indices = [0;1;2;3;[...];32489;32490;32491];
left_vertex_indices = left_vertex_indices + 1; % because matlab...
left_xyz_91282 = left_xyz_nii(left_vertex_indices,:);

right_vertex_indices = [0;1;2;3;[...];32489;32490;32491];
right_vertex_indices = right_vertex_indices + 1; %because matlab...
right_xyz_91282 = right_xyz_nii(right_vertex_indices,:);

all_xyz_91282 = [left_xyz_91282; right_xyz_91282];
```

```
%% get V1 coordinates using Wang (2015) atlas
v1_wang_indices = find(wang2015 == 2 | wang2015 == 3); % 2 = V1v
                                                          3 = V1d
v1_wang_all_xyz = all_xyz_91282(v1_wang_indices,:);

%% get values for V1 using Wang (2015) atlas
v1_wang_ang = ang_all_top_is_0(v1_wang_indices);
v1_wang_ecc = ecc_all(v1_wang_indices);

v1_wang_xyz_ang_ecc = [v1_wang_all_xyz, v1_wang_ang, v1_wang_ecc];
save('v1_wang_xyz_ang_ecc.mat', 'v1_wang_xyz_ang_ecc');
```

LISTING VII GET_ANG_ECC_PIXEL.M

```
function [col, row] = get_ang_ecc_pixel(polar_angle, eccentricity,
                                        width, height)
%GET_ANG_ECC_PIXEL Returns the pixel coordinates for a given polar
angle and eccentricity relative to the centre of a pixel frame.

% set default values for frame size
if nargin < 3
    width = 176;
    height = 100;
end

% get central point
centre_col = width/2;
centre_row = height/2;

% convert to pixels
eccentricity = eccentricity * 100 / 8;

% calculate lengths of adjacent and opposite
delta_row = eccentricity .* cosd(polar_angle);
delta_col = eccentricity .* sind(polar_angle);

% adjust pixel positions from centre
row = round(centre_row - delta_row);
col = round(centre_col - delta_col);

end
```

LISTING VIII GET_INPUT_COORDINATES.M

```
function [ coord ] = get_input_coordinates(blocks)
%GET_INPUT_COORDINATES Returns a list of input coordinates for the
 blocks.

v1_data = load('v1_wang_xyz_ang_ecc.mat');
v1_data = v1_data.v1_wang_xyz_ang_ecc;

[cols, rows] = get_block_centres(blocks);
[ret_cols,ret_rows] = get_ang_ecc_pixel(v1_data(:,4),v1_data(:,5));
coord = zeros(size(cols,1), 3);

for i = 1:size(cols,1)
    [~, ind] = min(sqrt((ret_cols - cols(i)).^2
                  + (ret_rows - rows(i)).^2));
    coord(i,:) = v1_data(ind, 1:3);
end

coord = round(coord);

end
```

LISTING IX COMBINE_AUDIO_VISUAL_DATA.PY

```python
import matplotlib.pyplot as plt
import numpy as np
import os


def fill_array_centred(array, rows):
    front_rows = int(rows / 2)
    back_rows = rows - front_rows # in case of odd number
    cols = np.shape(array)[1]

    return np.concatenate((np.zeros((front_rows, cols), dtype=int),
            array, np.zeros((back_rows, cols), dtype=int)), axis=0)


def fill_array_front(array, rows):
    cols = np.shape(array)[1]

    return np.concatenate((array, np.zeros((rows, cols),
                                            dtype=int)), axis=0)


def generate_list_of_filenames():
    filenames = []

    for sam in range(1, 251):
        word = ''

        if sam < 51:
            word = 'bird'
        elif sam < 101:
            word = 'down'
        elif sam < 151:
            word = 'stop'
        elif sam < 201:
            word = 'tree'
        else:
            word = 'up'

        filenames.append('sam' + str(sam) + '_' + word + '.csv')

    return filenames


def remove_leading_rows_of_zeros(array):
    while True:
        # break if no rows left
        if np.shape(array)[0] == 0:
            return array

        # break if row with content is detected
        # but only if next row also has content
        if array[0,:].sum() > 0 and array[1,:].sum() > 0:
            return array

        # keep removing rows until either of the above conditions is met
        array = np.delete(array, 0, axis=0)
```

```
def remove_trailing_rows_of_zeros(array):
    while True:
        # break if no rows left
        if np.shape(array)[0] == 0:
            return array

        # break if row with content is detected
        # but only if next row also has content
        if array[-1,:].sum() > 0 and array[-2,:].sum() > 0:
            return array

        # keep removing rows until either of the above conditions is met
        array = np.delete(array, -1, axis=0)


def main(template_name):
    print('Combining', template_name)

    audio_src = os.path.join('speech', 'encoded_samples',
                             template_name)
    video_scr = os.path.join('videos', 'encoded_samples',
                             template_name)
    tar_dir = os.path.join('combined', template_name)

    filenames = generate_list_of_filenames()

    # some analysis
    audio_rows_before = []
    audio_rows_after = []
    video_rows_before = []
    video_rows_after = []

    # merge files from different folders
    for filename in filenames:
        audio_file = np.loadtxt(os.path.join(audio_src, filename),
                                dtype='int', delimiter=',')
        video_file = np.loadtxt(os.path.join(video_scr, filename),
                                dtype='int', delimiter=',')

        audio_rows_before.append(np.shape(audio_file)[0])
        video_rows_before.append(np.shape(video_file)[0])

        # remove rows of zeros
        audio_file = remove_leading_rows_of_zeros(audio_file)
        audio_file = remove_trailing_rows_of_zeros(audio_file)
        video_file = remove_leading_rows_of_zeros(video_file)
        video_file = remove_trailing_rows_of_zeros(video_file)

        # adjust video file length based on biological observations
        video_file = np.repeat(video_file, 3, axis=0)

        audio_rows_after.append(np.shape(audio_file)[0])
        video_rows_after.append(np.shape(video_file)[0])

        # calculate how many rows need to be padded
        row_diff = np.shape(audio_file)[0]
                   - np.shape(video_file)[0]

        # if video file is longer than audio file, fill audio file
        if row_diff < 0:
            audio_file = fill_array_front(audio_file,abs(row_diff))
```

```
        # if audio file is longer than video file, fill video file
        if row_diff > 0:
            video_file = fill_array_front(video_file, row_diff)

        # join files and save as one
        combined = np.concatenate((audio_file, video_file), axis=1)
        np.savetxt(os.path.join(tar_dir, filename), combined,
                   fmt='%d', delimiter=',')

    # some statistics
    print("Audio file length ranged from",
          np.min(audio_rows_before), "to",
          np.max(audio_rows_before), "with mean",
          np.mean(audio_rows_before))
    print("Audio file length now ranges from",
          np.min(audio_rows_after), "to",
          np.max(audio_rows_after), "with mean",
          np.mean(audio_rows_after))
    print("Video file length ranged from",
          np.min(video_rows_before), "to",
          np.max(video_rows_before), "with mean",
          np.mean(video_rows_before))
    print("Video file length now ranges from",
          np.min(video_rows_after), "to",
          np.max(video_rows_after), "with mean",
          np.mean(video_rows_after))

    fig, (ax0, ax1) = plt.subplots(nrows=2, sharex=True)
    ax0.set_title("Audio file length before")
    ax0.hist(audio_rows_before, bins=40)
    ax1.set_title("Audio file length after")
    ax1.hist(audio_rows_after, bins=40)
    fig.tight_layout()
    plt.show()

    fig, (ax0, ax1) = plt.subplots(nrows=2, sharex=True)
    ax0.set_title("Video file length before")
    ax0.hist(video_rows_before, bins=40)
    ax1.set_title("Video file length after")
    ax1.hist(video_rows_after, bins=40)
    fig.tight_layout()
    plt.show()


main('MNI_by_4')
main('TAL_by_8')
```

```
%% switch template to MNI
num_feat = "49"; %#ok<NASGU>
template = "mni_by_4"; %#ok<NASGU>
folder = "MNI_by_4"; %#ok<NASGU>


%% switch template to TAL
num_feat = "39";
template = "tal_by_8";
folder = "TAL_by_8";


%% set class ID - (1) bird (2) down (3) stop (4) tree (5) up
class_id = "1";
class_name = "Bird";


%% load common files
all_coord = csvread(folder + "/" + template + ".csv");
input_coord = csvread(folder + "/for_" + template + ".csv");


%% plot
c = csvread(folder + "/NeuCube_Pruned_Neuron_Coordinates_Classes_"
            + class_id + "_Features_1-" + num_feat + ".csv");
m = csvread(folder + "/NeuCube_Pruned_Weight_Matrix_Classes_"
            + class_id + "_Features_1-" + num_feat + ".csv");

figure('Name',class_name,'NumberTitle','off','Color','white')

scatter3(all_coord(:,1),all_coord(:,2),all_coord(:,3),
         50, [0.7 0.7 0.7], '.')
hold on
scatter3(input_coord(:,1),input_coord(:,2),input_coord(:,3),
         50, [0.3 0.3 0.3],'*')
scatter3(c(:,1),c(:,2),c(:,3), 50, 'b', 'o')
axis('equal')
axis off


for i = 1:length(c)
    for j = 1:length(c)
        if m(i,j) > 0
            line([c(i,1), c(j,1)],
                 [c(i,2), c(j,2)],
                 [c(i,3), c(j,3)],
                 'Color', 'b', 'LineWidth', m(i,j)*8)
        end
    end
end
```

APPENDIX B SUPPLEMENTARY MATERIAL

This appendix contains supplementary tables with detailed results and supplementary figures for the experiments described in the three application chapters that were referenced in the main text.

TABLE OF CONTENTS

Table B-1: Results for the preliminary experiments using the sound processing model with different properties of Izhikevich neurons and the TAL_by_10 template.

Property	Accuracy in % – 10 runs with 5-fold cross-validation										Mean
A	52.2	51.2	53.0	52.4	50.0	52.2	50.8	52.0	54.4	53.6	52.2
B	52.6	54.4	53.2	53.6	52.8	52.8	50.8	54.2	52.0	54.0	53.0
C	55.6	54.4	56.4	55.8	54.6	55.0	57.6	55.8	56.0	54.8	55.6
D	55.0	54.6	54.4	54.0	52.6	53.0	55.6	55.0	53.4	53.0	54.1
E	53.8	54.6	55.4	55.6	54.4	54.4	55.4	55.0	55.6	53.4	54.8
F	53.0	52.6	51.4	54.0	53.6	52.6	52.2	51.6	52.6	53.8	52.7
G	56.8	55.2	56.6	56.0	53.0	54.6	55.6	54.8	53.2	54.8	55.1
H	51.8	51.8	53.8	52.6	54.0	52.4	52.6	54.0	52.0	51.8	52.7
I	52.6	53.6	51.4	52.6	51.4	51.4	54.0	52.4	52.0	53.6	52.5
J	57.0	56.4	55.8	55.2	54.8	55.4	56.0	55.4	55.6	55.4	55.7
K	55.0	57.2	54.6	55.8	56.4	55.8	57.2	55.4	57.8	58.2	56.3
L	52.6	54.2	53.8	54.0	55.0	53.6	53.4	55.6	55.6	55.8	54.4
M	54.6	56.2	55.4	55.8	56.4	55.2	56.4	55.0	55.6	55.4	55.6
N	54.2	52.6	53.8	54.2	53.8	54.0	53.6	55.2	54.4	53.4	53.9
O	56.8	55.6	56.8	55.8	56.2	57.0	56.8	56.4	55.6	55.6	56.3
P	63.6	62.8	62.2	63.8	63.8	65.0	64.2	64.2	63.8	64.6	63.8
Q	64.4	62.4	65.2	64.2	65.4	65.0	65.0	62.4	63.4	65.0	64.2
R	59.2	58.4	59.0	60.2	58.2	60.6	59.8	59.0	60.2	58.4	59.3
S	58.6	56.4	59.0	57.0	59.4	56.4	58.8	59.6	56.2	59.2	58.1
T	out of memory										

TABLE B-2: RESULTS FOR THE PRELIMINARY EXPERIMENTS USING THE SOUND PROCESSING MODEL WITH DIFFERENT THRESHOLDS FOR LIF NEURONS.

Threshold	Accuracy in % – 10 runs with 5-fold cross-validation										Mean
0.01	66.2	66.0	68.6	67.4	63.2	64.6	66.8	65.0	66.2	66.0	66.0
0.02	67.4	65.4	69.4	67.8	68.6	70.0	67.2	68.0	66.0	65.6	67.5
0.03	68.8	65.6	68.4	67.2	68.0	66.2	68.6	68.0	71.6	70.0	68.2
0.04	70.8	65.4	65.4	67.4	67.6	67.4	67.8	66.0	66.4	66.6	67.1
0.05	65.0	67.0	66.8	65.8	68.0	65.4	69.4	67.0	66.2	66.6	66.7
0.06	68.4	69.0	66.6	65.4	64.6	70.4	69.8	62.6	68.0	67.6	67.2
0.07	63.8	64.0	64.6	63.2	64.2	64.8	64.4	65.0	67.4	65.2	64.7
0.08	63.4	64.2	65.4	64.0	62.2	63.6	62.8	64.2	62.8	62.8	63.5
0.09	63.2	63.8	63.6	63.4	64.2	64.0	64.2	64.2	64.2	63.0	63.8
0.1	64.4	65.2	63.0	64.2	63.6	64.4	63.0	63.4	65.0	63.0	63.9
0.15	63.8	64.0	64.6	64.0	64.2	65.2	63.0	62.8	64.2	64.2	64.0
0.2	63.6	63.0	64.2	63.2	64.6	62.0	62.4	62.4	66.6	64.2	63.6
0.25	65.2	63.8	64.8	65.2	65.6	65.4	63.8	59.6	64.4	64.2	64.2
0.3	64.0	65.0	65.2	65.4	65.0	65.4	64.0	63.4	62.8	65.8	64.6
0.35	63.6	63.8	63.4	64.8	63.4	62.8	64.6	62.6	65.4	64.4	63.9
0.4	63.0	64.4	63.2	64.2	65.2	62.8	65.2	63.8	63.0	64.2	63.9
0.45	63.8	63.6	63.2	65.2	64.2	63.4	64.8	64.4	65.8	64.6	64.3
0.5	64.6	63.8	63.4	64.4	64.2	65.4	65.0	64.0	65.2	64.0	64.4
0.55	63.8	62.8	66.0	65.0	64.6	65.4	65.0	64.4	64.4	64.6	64.6
0.6	63.6	64.8	63.4	63.2	64.6	63.6	65.0	63.6	63.4	63.4	63.9
0.65	64.2	64.4	64.8	63.8	65.0	64.2	65.2	65.0	64.6	64.4	64.6
0.7	64.4	64.4	63.2	63.2	64.2	64.4	65.0	64.0	63.8	64.4	64.1
0.75	63.6	63.0	64.4	63.2	62.4	64.6	65.2	63.6	65.4	64.2	64.0
0.8	65.0	63.6	63.4	63.6	64.8	65.2	64.8	64.6	64.0	64.2	64.3
0.85	64.8	64.8	63.4	64.8	65.0	63.4	63.4	62.8	63.4	63.4	63.9
0.9	65.6	65.6	63.6	63.2	64.8	64.4	63.8	64.0	64.0	65.6	64.5
0.95	62.8	64.0	63.2	64.0	63.6	64.8	64.4	64.2	64.0	64.8	64.0
1	65.0	65.0	63.0	64.8	63.6	64.4	65.6	65.4	63.2	63.6	64.4
1.5	65.8	63.4	64.2	63.2	63.4	62.6	64.4	65.0	63.4	63.0	63.8
2	62.4	65.2	64.2	64.4	62.6	61.8	65.4	63.0	62.4	65.2	63.7

Table B-3: Detailed results for the 30 runs of the cross-validation experiments using the sound processing system and the MNI_by_3 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	82.69	87.61	86.69	88.60	88.84	87.99	88.24	87.22	89.66	90.44	89.20	89.20	88.48	89.43	88.36	90.77	87.86	89.08
Run 2	80.27	86.82	88.72	92.50	90.66	86.42	91.60	91.19	88.48	89.66	89.89	90.66	91.40	88.48	90.22	89.31	89.20	88.96
Run 3	77.75	87.99	87.86	90.00	89.31	86.69	89.31	89.66	87.48	86.15	91.29	88.24	88.11	89.20	87.99	88.24	89.31	88.84
Run 4	80.63	87.86	89.31	88.24	87.73	89.31	89.43	88.24	88.72	87.09	89.31	88.24	90.55	89.20	89.43	89.43	88.11	89.20
Run 5	75.40	87.73	87.09	89.89	86.56	89.66	90.00	89.66	89.55	87.99	90.77	90.00	89.89	88.11	88.24	88.36	89.43	89.43
Run 6	79.52	87.09	89.66	90.55	89.43	89.43	87.09	88.60	87.73	90.66	89.55	86.42	88.60	89.43	88.60	90.44	89.08	89.43
Run 7	81.34	88.24	86.15	88.48	87.35	89.31	89.20	90.55	89.31	88.96	87.09	88.84	88.36	87.22	86.82	87.99	89.31	88.96
Run 8	77.75	88.48	89.89	88.11	89.55	89.77	90.33	88.48	87.22	90.00	87.99	91.19	88.84	90.00	87.86	89.66	88.48	87.99
Run 9	82.02	86.56	89.66	90.77	88.60	88.96	90.33	86.01	89.31	88.24	90.11	90.22	89.66	89.66	88.36	89.20	87.61	89.31
Run 10	78.95	87.22	88.36	88.60	90.87	89.55	87.61	89.08	90.33	87.48	88.60	87.22	88.24	88.96	89.77	89.20	89.08	88.84
Run 11	81.85	87.86	88.96	93.07	86.01	87.35	87.35	88.96	89.08	86.69	89.43	91.50	90.55	90.77	89.31	87.61	89.55	88.72
Run 12	80.81	86.42	88.60	86.82	91.09	92.01	89.08	88.36	90.11	87.22	89.20	90.11	88.11	89.43	89.55	89.31	88.84	88.24
Run 13	81.16	87.61	87.35	90.66	90.77	88.36	87.48	89.08	88.11	88.11	91.81	90.55	91.91	89.66	87.99	90.11	88.60	88.24
Run 14	81.34	86.69	90.11	91.50	88.96	88.60	88.36	90.55	90.00	89.66	89.31	90.33	89.55	88.36	88.84	89.77	88.84	89.20
Run 15	80.63	88.84	89.43	90.55	89.89	88.96	89.43	88.72	88.11	88.11	88.11	89.66	89.31	88.24	88.24	87.35	87.35	88.48
Run 16	81.16	86.15	90.66	89.43	89.43	89.43	89.08	88.48	87.73	89.08	90.44	89.89	89.08	89.66	90.22	88.96	90.66	87.99
Run 17	79.33	86.82	88.48	89.55	89.08	87.48	89.66	90.00	86.96	88.60	90.66	90.11	88.36	90.00	89.31	87.73	89.55	88.24
Run 18	78.95	86.15	89.66	88.48	87.22	90.22	89.66	90.00	89.55	88.48	89.31	90.55	89.66	87.86	88.84	91.19	88.48	88.60
Run 19	81.34	90.22	90.87	88.60	88.11	89.89	88.96	89.77	86.56	87.48	90.22	89.89	89.77	89.55	88.96	88.72	89.77	87.22
Run 20	80.99	88.96	90.77	87.35	86.42	90.22	90.55	88.36	90.87	87.73	90.44	89.43	88.11	88.96	88.60	89.20	88.36	89.77
Run 21	80.27	87.09	88.48	88.72	90.22	87.99	89.89	87.35	90.33	88.48	86.96	86.56	89.66	89.43	88.24	89.77	88.72	89.43
Run 22	77.96	87.48	91.40	89.08	89.31	89.66	88.96	89.43	86.15	89.08	90.11	89.43	89.20	89.20	88.36	90.22	89.31	89.55
Run 23	78.75	88.11	89.55	92.69	86.69	87.61	89.66	88.60	89.31	87.09	91.70	88.72	89.77	87.48	88.60	90.66	90.77	89.31
Run 24	80.45	88.24	87.09	90.77	90.87	88.84	89.31	88.72	87.99	88.60	89.31	89.77	88.72	89.55	88.60	87.73	89.08	88.60
Run 25	80.08	89.20	88.48	90.00	90.11	87.22	88.36	87.99	87.22	86.28	90.11	90.44	88.24	88.48	88.36	88.84	86.96	89.89
Run 26	83.49	87.22	86.69	83.17	85.15	88.60	89.08	88.60	85.15	89.20	89.31	88.72	90.33	89.55	89.31	88.84	89.43	88.11
Run 27	78.95	86.15	90.66	87.09	84.42	86.56	89.89	90.44	89.77	86.42	88.36	89.77	90.22	90.22	88.11	87.48	87.48	86.69
Run 28	77.34	89.89	88.60	85.73	85.30	87.35	87.09	87.61	90.55	88.24	90.33	88.84	89.89	89.55	87.99	88.84	88.96	88.11
Run 29	81.51	87.86	86.42	89.08	87.22	87.99	89.55	88.96	88.84	90.55	90.77	88.84	88.36	88.36	87.73	89.55	88.36	89.43
Run 30	78.16	90.11	89.77	89.43	88.72	87.61	89.08	90.66	90.44	87.35	89.89	90.00	89.31	89.08	89.31	88.96	87.48	88.96
Mean	80.03	87.75	88.85	89.25	88.46	88.63	89.12	88.98	88.69	88.30	89.65	89.45	89.34	89.10	88.67	89.12	88.80	88.76
Std dev	0.017	0.011	0.014	0.020	0.018	0.013	0.010	0.011	0.014	0.012	0.012	0.012	0.010	0.008	0.007	0.010	0.009	0.007

TABLE B-4: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE SOUND PROCESSING SYSTEM AND THE MNI_BY_4 TEMPLATE IN %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	81.68	89.66	88.11	86.82	87.35	85.87	85.15	87.73	86.56	88.96	88.72	89.08	87.35	86.42	86.69	85.44	86.15	88.96
Run 2	75.62	85.87	87.09	82.02	86.82	83.49	87.35	89.89	89.08	88.24	87.09	88.24	88.11	87.09	87.22	87.48	85.15	88.24
Run 3	80.08	85.01	88.72	89.31	89.55	85.44	89.20	89.89	86.56	90.55	87.22	86.56	87.09	85.15	86.96	86.69	86.15	86.69
Run 4	82.52	86.01	86.01	84.11	88.60	87.22	88.84	87.86	88.24	89.77	85.73	85.44	87.09	87.09	88.48	85.15	87.22	88.84
Run 5	80.27	84.71	86.69	89.43	86.69	83.17	87.22	88.24	89.55	89.20	87.73	87.48	86.56	86.01	87.61	86.01	87.22	85.44
Run 6	78.16	87.73	85.01	85.73	87.61	88.11	86.56	87.48	89.31	87.48	86.42	86.96	87.99	85.59	88.11	85.30	85.15	87.22
Run 7	77.55	88.24	86.96	88.11	87.09	85.73	89.55	87.99	87.73	89.43	87.73	85.87	87.22	86.82	86.42	85.87	86.96	87.22
Run 8	78.36	87.09	87.35	88.48	85.44	84.11	84.71	85.44	87.48	88.60	89.43	87.99	88.24	86.42	88.72	86.42	87.22	85.15
Run 9	78.16	89.89	89.43	83.49	88.72	89.89	86.69	87.48	89.31	87.35	86.42	88.84	85.73	86.56	87.99	86.28	86.69	87.35
Run 10	79.90	83.01	87.09	84.57	88.36	88.48	87.73	90.00	87.86	89.89	88.72	86.96	85.44	86.56	85.73	88.72	85.01	85.87
Run 11	78.56	85.44	84.71	88.36	89.89	88.72	85.73	88.60	87.48	88.36	88.48	86.69	87.09	86.15	85.44	87.61	87.09	85.30
Run 12	79.52	85.73	86.15	87.09	90.33	86.96	85.30	87.86	88.11	89.20	87.35	87.73	87.86	85.73	86.28	87.86	86.01	86.96
Run 13	81.85	86.82	88.11	88.36	86.69	87.09	86.96	88.11	90.66	87.61	86.96	88.72	86.15	85.87	86.42	85.87	87.73	86.15
Run 14	81.16	88.24	87.35	87.99	87.99	86.69	85.30	89.08	88.84	92.11	86.56	89.08	86.28	85.44	87.48	85.59	85.01	85.44
Run 15	80.81	89.43	85.15	88.48	87.35	88.48	90.00	87.35	88.48	85.87	89.66	87.22	85.59	86.15	86.69	85.59	88.24	87.22
Run 16	79.90	87.99	87.99	85.15	87.99	88.72	88.72	86.96	89.77	91.70	86.69	86.28	86.15	86.28	88.11	87.61	86.69	85.44
Run 17	79.90	84.71	87.73	87.99	84.42	85.15	87.22	90.66	87.73	86.82	87.61	87.35	89.77	86.96	89.08	86.01	84.26	86.15
Run 18	84.11	86.42	87.73	87.73	85.01	85.30	88.48	87.61	87.48	88.24	89.08	86.96	86.01	88.48	86.69	86.82	88.36	87.22
Run 19	77.55	87.22	87.22	86.96	88.60	85.73	89.55	87.09	89.77	88.48	86.28	86.28	89.08	86.56	87.86	87.73	87.73	85.30
Run 20	80.63	85.87	86.42	88.72	86.01	87.09	85.44	88.96	87.99	88.36	86.82	87.99	86.56	86.96	87.09	86.96	86.15	87.48
Run 21	81.16	86.15	87.61	88.72	88.11	87.99	88.60	86.82	88.36	87.61	89.20	87.48	87.48	87.73	85.73	85.15	86.82	88.36
Run 22	80.45	88.48	87.61	85.44	88.36	87.61	90.00	86.82	90.77	88.24	87.73	87.09	88.11	84.11	86.01	85.15	87.35	86.82
Run 23	81.16	85.30	88.48	87.48	89.55	88.48	87.86	86.42	88.72	88.96	86.82	87.22	87.09	87.48	90.00	87.99	87.35	86.69
Run 24	77.34	87.35	85.15	88.84	87.73	83.96	86.56	90.22	88.36	88.60	87.99	88.24	86.42	86.82	86.82	85.73	86.15	85.44
Run 25	79.71	86.15	86.69	88.84	84.11	86.28	88.11	89.31	87.35	87.61	89.55	87.99	86.96	89.55	86.82	87.73	85.44	86.15
Run 26	79.90	87.09	91.19	86.96	89.08	89.55	88.96	85.30	87.86	88.72	89.20	87.09	86.28	88.24	86.69	86.69	87.35	86.82
Run 27	82.19	86.96	86.96	87.22	88.84	90.00	87.73	89.55	87.48	89.31	86.96	85.59	87.99	87.99	87.61	86.69	87.73	88.24
Run 28	80.27	86.01	83.96	88.60	89.08	87.73	86.82	88.36	89.89	89.31	87.35	85.15	88.24	86.82	87.86	86.15	85.87	88.36
Run 29	77.96	85.87	89.20	87.35	86.69	87.22	86.96	87.61	89.55	87.35	89.55	87.61	88.96	86.82	86.28	87.35	85.73	86.28
Run 30	80.27	86.42	85.30	85.01	86.56	87.35	86.28	89.43	90.33	87.48	88.72	88.11	86.96	87.22	87.73	86.15	85.59	85.01
Mean	79.89	86.69	87.11	87.11	87.62	86.92	87.45	88.14	88.56	88.65	87.79	87.31	87.19	86.70	87.22	86.53	86.52	86.73
Std dev	0.0177	0.0154	0.0152	0.0184	0.0155	0.0181	0.0150	0.0134	0.0112	0.0131	0.0113	0.0102	0.0106	0.0104	0.0104	0.0097	0.0104	0.0115

TABLE B-5: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE SOUND PROCESSING SYSTEM AND THE MNI_BY_5 TEMPLATE IN %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	85.30	85.73	86.15	83.17	83.80	86.42	83.65	84.71	86.82	83.33	83.49	82.69	83.65	81.68	82.36	82.85	83.65	83.01
Run 2	86.69	85.44	86.28	84.26	86.15	86.01	82.19	84.86	85.44	83.33	84.42	82.36	82.52	83.17	82.52	85.30	81.68	82.52
Run 3	86.15	86.69	82.52	83.80	83.49	84.86	84.57	86.96	84.42	84.57	84.11	83.17	83.33	81.51	83.80	84.71	83.33	84.86
Run 4	83.65	83.65	84.26	86.82	83.49	84.26	84.86	89.31	84.11	85.73	84.71	84.71	84.57	81.85	83.96	84.11	81.85	84.11
Run 5	84.86	83.96	84.57	87.86	85.87	87.22	86.82	85.44	85.01	85.01	84.86	82.19	83.33	83.80	83.17	83.33	83.96	84.86
Run 6	80.99	84.57	87.61	83.33	82.52	86.56	84.42	86.96	86.96	84.26	84.71	84.57	83.80	82.02	83.49	83.49	82.85	84.11
Run 7	82.19	86.15	84.86	85.59	85.73	87.73	86.69	83.17	85.44	85.01	82.19	83.80	85.30	83.80	82.19	82.52	83.33	83.49
Run 8	84.26	83.96	85.73	84.86	84.86	86.15	84.71	87.09	88.36	86.01	81.51	81.85	83.33	83.96	82.52	84.86	82.85	82.85
Run 9	84.42	84.57	83.49	88.24	85.87	86.82	84.11	83.96	83.65	83.65	81.51	78.56	82.69	80.81	83.96	81.34	84.57	84.71
Run 10	83.01	85.15	85.44	85.01	87.48	85.73	87.35	83.49	86.01	84.11	84.86	86.15	81.85	85.59	80.99	85.30	83.33	83.80
Run 11	83.17	85.59	84.71	85.73	88.24	85.44	84.71	86.42	82.36	83.49	85.59	81.34	84.57	82.36	83.96	82.85	82.19	83.80
Run 12	86.42	85.44	84.26	82.69	83.33	82.69	83.01	85.01	86.01	82.85	86.01	83.49	83.17	83.33	86.15	85.15	85.59	82.36
Run 13	83.01	86.28	83.33	88.24	84.26	85.01	83.65	85.30	84.11	84.57	83.96	84.57	83.17	83.49	85.59	82.85	85.01	82.69
Run 14	83.96	85.73	83.65	87.73	85.87	86.96	85.30	85.59	84.26	85.30	81.85	82.19	83.80	83.80	82.36	82.85	85.01	82.52
Run 15	87.35	84.57	85.01	84.71	86.82	85.01	86.69	85.01	85.30	85.87	85.01	83.01	81.16	84.11	83.49	82.19	84.42	83.49
Run 16	81.16	85.44	86.28	86.96	87.22	82.52	86.96	86.28	84.71	83.49	83.49	84.57	82.85	85.59	83.33	81.68	83.80	82.85
Run 17	80.45	86.82	81.51	85.30	87.09	86.69	85.30	84.57	86.69	83.80	83.17	83.01	85.59	82.52	82.69	83.01	83.80	82.19
Run 18	83.17	85.30	85.15	85.01	86.42	87.48	85.73	86.42	85.87	85.87	85.15	85.30	83.17	85.73	82.19	85.59	83.96	82.69
Run 19	84.86	84.42	87.61	83.33	85.44	83.96	84.57	83.80	88.72	81.85	82.85	82.36	82.19	81.68	85.15	85.73	82.69	83.17
Run 20	82.52	81.85	85.30	86.15	84.42	87.22	87.22	84.11	82.52	85.15	86.01	85.01	84.11	85.44	83.01	81.34	83.80	83.01
Run 21	85.30	85.73	81.85	85.30	87.22	86.15	85.44	87.86	83.33	84.26	85.01	85.87	84.11	85.59	83.80	85.30	82.02	83.96
Run 22	83.33	85.59	87.86	86.82	86.69	86.01	86.96	85.59	82.52	86.42	84.57	83.49	83.49	84.86	84.11	83.49	82.52	84.71
Run 23	82.85	82.85	85.44	85.15	86.69	87.35	84.42	80.99	86.96	84.26	82.36	84.11	85.30	82.52	84.57	81.16	83.65	85.15
Run 24	86.82	84.26	83.65	85.15	84.57	84.57	86.69	86.28	83.65	84.26	84.86	84.71	80.81	82.52	84.26	82.69	82.36	82.36
Run 25	85.30	86.28	83.65	85.15	86.28	83.65	87.22	83.80	86.01	86.96	84.26	84.86	85.73	83.01	84.86	80.81	86.15	79.90
Run 26	83.96	87.22	85.73	85.59	84.86	86.28	83.01	84.71	82.69	85.01	84.26	85.44	83.65	83.65	84.11	83.96	82.85	84.26
Run 27	84.26	83.01	86.96	86.42	86.69	87.35	88.48	86.01	82.69	84.57	84.42	85.59	81.85	84.71	81.85	83.65	83.33	83.33
Run 28	83.49	84.42	83.80	83.17	84.26	86.28	86.15	86.01	86.56	83.17	85.87	84.42	83.17	82.85	83.96	85.30	83.33	80.99
Run 29	82.36	85.59	85.15	83.80	85.01	86.15	88.24	84.86	85.44	83.01	85.59	84.11	85.15	82.19	86.69	86.15	83.65	86.15
Run 30	83.65	85.59	86.15	84.42	85.30	84.86	85.01	85.44	86.69	84.42	84.57	82.69	83.80	84.11	84.57	83.17	83.96	85.01
Mean	83.96	85.06	84.93	85.33	85.53	85.78	85.47	85.33	85.11	84.45	84.17	83.67	83.51	83.41	83.66	83.56	83.52	83.43
Std dev	0.0170	0.0120	0.0158	0.0153	0.0140	0.0136	0.0158	0.0156	0.0171	0.0114	0.0128	0.0158	0.0121	0.0136	0.0127	0.0149	0.0105	0.0127

TABLE B-6: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE SOUND PROCESSING SYSTEM AND THE TAL_BY_6 TEMPLATE IN %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	76.06	78.36	74.01	76.06	80.99	85.15	85.15	85.87	85.30	87.61	87.73	88.48	88.96	90.33	89.31	87.73	89.66	88.24
Run 2	79.14	80.08	77.13	78.95	83.49	87.99	86.56	87.22	85.73	86.56	87.09	86.42	87.99	88.84	90.98	88.84	88.36	89.31
Run 3	75.40	78.75	78.75	81.34	76.92	84.86	85.87	85.87	88.72	90.00	86.69	88.24	89.31	90.11	89.20	89.66	89.31	90.66
Run 4	76.06	80.27	78.95	82.36	80.99	83.33	83.33	89.55	87.86	86.56	88.36	87.73	91.19	90.66	87.99	89.55	89.77	88.48
Run 5	74.94	81.51	74.71	76.50	77.34	84.26	86.96	89.20	88.60	86.69	88.96	88.24	87.86	90.98	89.43	89.20	87.86	88.96
Run 6	80.27	81.34	75.40	79.52	79.14	85.87	86.42	86.69	85.59	86.82	88.96	86.15	89.89	91.91	90.55	88.72	86.15	88.24
Run 7	80.27	81.34	75.84	80.81	78.75	84.42	87.22	84.86	87.35	86.56	87.35	89.20	89.31	90.00	89.66	89.66	88.48	89.89
Run 8	75.62	80.81	78.16	83.17	79.33	86.56	86.28	86.69	87.48	87.61	89.20	89.20	88.48	90.33	88.48	87.73	87.73	88.72
Run 9	77.96	78.16	77.75	77.13	79.14	84.42	84.26	84.11	85.87	88.72	88.11	89.66	89.66	89.66	89.08	90.33	90.66	89.66
Run 10	77.75	77.75	83.33	82.02	83.01	84.57	83.96	86.15	87.61	82.36	88.48	87.73	88.36	89.55	90.44	89.55	88.72	89.08
Run 11	75.40	78.75	75.40	79.14	79.90	85.73	87.09	85.73	88.11	88.24	90.55	88.60	90.66	90.00	88.24	88.84	89.77	89.31
Run 12	74.48	79.71	72.79	76.50	83.80	83.33	86.96	86.28	87.99	88.84	88.84	87.99	90.44	87.73	89.43	87.73	88.48	88.84
Run 13	77.75	77.13	75.62	82.02	78.95	84.86	85.15	85.44	89.31	88.84	87.48	90.44	88.48	89.89	86.82	89.20	87.09	88.60
Run 14	76.50	76.50	80.99	71.52	80.99	85.30	85.44	88.60	88.11	87.09	88.24	87.61	89.08	88.96	90.55	89.66	88.11	88.84
Run 15	77.55	78.75	78.75	80.45	78.75	83.96	86.42	83.65	88.72	87.35	88.11	88.24	88.36	89.55	88.11	88.48	88.48	88.11
Run 16	74.71	80.81	82.19	80.81	79.90	83.49	86.56	87.48	90.22	86.82	89.20	87.86	91.19	91.09	87.48	89.20	88.96	88.96
Run 17	73.04	80.45	79.90	81.34	77.34	85.87	87.48	84.86	86.42	88.96	85.30	89.20	89.20	90.22	89.66	89.31	87.35	88.48
Run 18	77.34	80.99	73.53	76.50	79.14	85.01	86.56	88.48	87.48	86.42	87.61	88.11	88.60	89.77	89.31	89.77	87.35	89.31
Run 19	74.94	77.13	73.29	76.92	76.50	82.85	87.86	86.01	88.36	87.99	88.96	87.61	89.43	89.43	89.66	89.55	87.99	88.36
Run 20	77.55	80.45	81.85	80.08	76.92	86.69	87.09	85.01	85.59	87.48	89.66	89.55	91.09	91.60	88.84	89.43	88.96	89.20
Run 21	79.90	80.81	73.29	83.01	76.71	85.30	82.36	88.48	87.99	86.01	89.08	88.36	89.55	87.73	88.96	89.66	89.89	88.36
Run 22	75.62	77.34	78.75	71.78	79.14	86.96	83.80	88.48	88.96	88.96	87.09	90.33	90.44	89.66	90.55	86.15	90.00	89.20
Run 23	79.52	80.63	77.13	82.69	80.63	82.69	87.35	85.73	89.66	87.48	90.00	86.82	91.40	90.22	89.20	88.84	90.55	88.48
Run 24	77.55	77.96	83.17	76.71	79.14	80.45	87.22	84.11	88.36	86.96	87.22	88.11	90.11	88.48	88.60	89.89	87.61	87.99
Run 25	75.17	80.45	80.63	78.95	75.62	85.87	85.30	88.48	87.61	85.59	89.55	87.09	86.28	89.55	88.36	89.31	88.96	88.48
Run 26	76.06	82.19	81.16	76.92	77.13	86.42	86.56	86.15	86.69	85.73	87.73	88.60	88.60	89.20	90.77	88.96	89.08	89.08
Run 27	80.45	79.52	78.75	81.34	72.03	83.33	87.48	85.87	89.20	87.61	88.84	88.36	90.44	89.66	87.35	87.73	88.60	88.84
Run 28	77.96	80.27	73.53	80.81	81.16	84.71	87.35	85.01	87.99	89.31	87.22	90.44	88.84	89.77	90.44	88.72	89.77	87.86
Run 29	75.62	77.75	78.36	78.75	74.48	84.57	86.56	85.30	85.15	89.66	89.08	87.86	87.61	88.72	88.84	87.48	88.84	90.44
Run 30	78.56	79.33	76.28	76.28	74.01	83.49	86.15	87.99	87.22	87.73	86.96	87.73	89.20	90.11	89.43	88.48	89.08	89.08
Mean	76.97	79.51	77.65	79.01	78.71	84.74	86.09	86.44	87.64	87.42	88.25	88.33	89.33	89.79	89.19	88.91	88.72	88.90
Std dev	0.0193	0.0152	0.0307	0.0298	0.0266	0.0150	0.0135	0.0158	0.0132	0.0148	0.0112	0.0106	0.0118	0.0094	0.0104	0.0089	0.0104	0.0065

TABLE B-7: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE SOUND PROCESSING SYSTEM AND THE TAL_BY_7 TEMPLATE IN %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	76.50	77.34	77.75	84.42	83.17	86.42	84.11	85.01	87.09	85.87	86.82	87.86	88.96	87.22	87.86	88.11	86.69	86.82
Run 2	81.16	80.99	77.55	75.62	83.65	81.51	83.65	88.11	87.35	86.82	89.55	90.44	88.60	87.61	88.11	87.61	88.24	86.69
Run 3	75.62	79.52	82.19	73.77	83.01	83.17	83.96	85.44	87.99	84.57	87.99	90.55	87.48	90.11	88.96	88.96	88.96	88.60
Run 4	77.75	81.85	73.29	79.52	81.34	86.01	85.87	85.87	86.82	88.24	85.73	88.11	88.48	87.99	86.28	89.43	87.61	87.35
Run 5	79.33	81.16	79.71	77.55	83.49	83.65	87.09	83.33	86.69	87.09	88.72	90.98	88.24	87.86	86.15	89.08	87.35	86.56
Run 6	80.81	79.33	75.40	78.36	77.13	87.22	88.48	83.49	88.24	86.56	87.22	89.66	88.11	86.15	86.42	88.11	88.48	87.99
Run 7	80.45	78.95	77.55	83.96	83.01	85.59	84.71	83.96	86.96	87.99	86.56	88.96	89.55	87.35	85.59	85.73	86.82	87.86
Run 8	78.95	82.02	77.13	75.40	76.71	87.22	83.17	86.82	85.87	85.87	88.36	90.22	88.11	86.56	87.35	87.99	89.20	87.86
Run 9	78.36	81.16	76.28	81.51	82.69	86.69	87.48	85.73	88.84	86.56	85.87	90.55	91.09	87.22	87.61	86.69	86.56	86.42
Run 10	78.36	78.95	81.34	76.92	82.19	83.65	86.69	84.26	84.26	85.59	87.48	90.98	87.22	87.09	88.72	87.09	87.86	86.42
Run 11	77.96	78.75	76.28	75.40	83.65	89.20	83.17	83.17	85.87	86.42	88.84	88.11	87.22	88.84	87.61	86.96	87.22	87.35
Run 12	77.34	81.85	74.71	80.99	83.17	83.33	84.26	87.09	88.48	87.09	89.55	90.33	89.77	85.87	87.61	86.56	87.61	86.42
Run 13	74.94	78.95	80.27	74.01	77.34	86.69	86.96	88.24	89.31	84.57	89.08	88.60	87.86	87.09	87.99	86.96	87.22	87.61
Run 14	78.56	74.25	77.55	80.63	84.86	86.56	85.01	86.82	87.22	89.31	86.15	88.60	85.73	88.36	88.84	84.71	85.87	87.86
Run 15	77.13	80.81	76.50	73.53	80.45	86.96	86.69	86.15	88.48	85.44	88.48	90.22	87.86	87.22	87.48	88.60	86.28	87.48
Run 16	76.71	73.04	82.52	77.34	81.85	85.59	83.33	90.00	86.42	85.87	86.01	89.77	87.09	87.09	86.01	88.96	86.96	86.82
Run 17	77.96	74.25	76.28	76.06	77.34	86.01	85.30	85.01	84.86	85.59	86.96	89.77	87.22	86.56	87.09	87.09	87.35	85.30
Run 18	79.71	80.81	75.62	78.16	79.52	85.01	82.85	87.61	83.80	89.31	88.11	90.77	87.22	88.24	89.89	87.35	87.73	85.15
Run 19	75.17	81.51	78.36	83.96	82.69	83.17	86.15	88.72	83.49	87.22	87.99	90.11	91.50	88.48	86.56	87.73	87.99	88.36
Run 20	79.52	79.71	75.17	73.77	81.34	85.87	87.61	86.96	85.59	87.48	89.89	89.55	88.11	88.60	87.09	87.48	86.56	84.57
Run 21	76.28	81.51	80.63	81.51	83.01	86.82	84.42	85.73	87.73	86.42	85.59	89.89	88.60	86.56	87.86	88.24	88.60	87.22
Run 22	77.34	80.63	74.71	73.04	83.80	87.61	84.26	84.71	89.08	87.73	87.35	90.00	90.87	86.15	86.42	86.28	86.96	85.59
Run 23	80.45	80.81	77.55	75.62	79.71	86.15	84.71	85.44	84.26	87.73	87.99	90.00	87.09	86.28	87.73	88.36	86.82	87.09
Run 24	79.14	70.99	75.40	72.79	81.16	86.69	83.80	86.82	86.56	83.33	86.96	90.33	89.55	86.96	86.15	87.09	87.61	88.72
Run 25	79.14	79.14	74.94	75.84	80.99	86.82	87.48	87.09	87.09	89.55	87.61	90.77	87.22	86.56	87.61	87.86	87.86	87.35
Run 26	77.75	79.14	79.14	81.51	82.02	87.22	84.71	85.73	88.11	87.22	87.22	89.43	88.96	87.48	88.11	85.87	88.24	87.35
Run 27	76.50	70.99	73.29	74.01	78.95	83.33	84.26	85.73	88.24	86.82	87.22	90.00	87.99	86.96	88.48	87.48	86.42	88.36
Run 28	79.71	80.63	82.52	80.81	81.34	87.09	81.85	85.15	86.56	86.69	87.48	89.77	87.48	89.55	86.28	87.61	88.60	86.69
Run 29	77.13	82.52	82.19	77.75	80.99	86.82	86.42	85.44	81.85	88.60	86.15	88.11	92.20	87.73	88.84	86.56	86.56	85.73
Run 30	75.62	80.45	76.06	73.53	83.01	88.36	83.33	84.71	85.01	87.61	89.20	89.20	90.98	87.99	88.11	87.35	88.48	88.72
Mean	78.05	79.07	77.60	77.58	81.45	85.88	85.06	85.95	86.60	86.84	87.60	89.72	88.55	87.46	87.49	87.46	87.49	87.08
Std dev	0.017	0.031	0.027	0.035	0.022	0.017	0.017	0.016	0.018	0.014	0.012	0.009	0.015	0.010	0.010	0.010	0.008	0.011

TABLE B-8: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE SOUND PROCESSING SYSTEM AND THE TAL_BY_8 TEMPLATE IN %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	80.27	81.51	81.68	80.63	80.81	83.01	82.69	84.26	86.82	88.72	87.99	90.98	85.59	87.09	88.48	87.61	87.22	89.20
Run 2	78.75	80.81	78.75	81.68	82.69	86.82	87.48	86.28	86.15	88.11	88.96	88.24	88.11	86.56	89.55	87.99	88.84	87.48
Run 3	75.84	82.52	76.50	81.68	82.52	84.11	87.61	85.30	86.15	85.44	89.89	88.60	87.22	88.24	87.22	88.72	86.96	87.99
Run 4	78.16	79.52	73.53	80.08	82.52	85.30	85.59	82.69	89.55	85.59	86.96	86.69	89.20	86.82	87.22	87.86	88.48	86.56
Run 5	80.63	78.75	76.28	80.63	81.85	87.99	85.59	86.28	89.43	85.01	87.61	89.31	87.09	86.96	86.56	87.35	87.35	89.20
Run 6	82.02	74.71	84.42	81.85	79.90	84.86	87.73	88.11	89.08	85.73	89.08	90.00	85.87	87.61	88.48	88.84	85.59	88.36
Run 7	78.16	78.16	80.63	79.90	82.36	84.71	85.15	87.09	86.96	87.35	90.22	86.96	89.89	87.35	85.73	87.22	86.96	86.15
Run 8	79.90	73.29	81.51	73.77	78.75	85.01	86.96	88.60	86.42	87.22	89.20	88.96	86.42	86.56	87.48	87.09	86.15	86.56
Run 9	80.99	80.81	76.28	79.52	81.68	86.01	85.87	87.09	88.96	89.66	90.44	89.77	88.84	86.42	90.00	88.11	88.36	87.48
Run 10	76.28	80.08	82.36	79.14	83.65	87.22	85.15	83.01	85.44	87.99	88.84	87.22	88.11	87.73	86.96	84.71	88.24	86.96
Run 11	75.17	80.81	75.40	81.16	78.16	85.30	87.99	88.84	86.56	85.87	88.24	89.08	85.15	86.42	89.08	86.82	87.35	87.22
Run 12	79.90	83.33	74.71	79.33	83.80	81.51	86.01	87.48	86.56	83.49	90.55	87.35	83.33	87.35	88.48	85.15	84.42	88.84
Run 13	79.33	83.33	79.14	78.75	80.08	84.86	86.15	86.69	85.44	86.15	88.96	87.86	88.36	87.99	86.82	87.35	86.42	85.87
Run 14	76.28	78.36	76.28	82.19	81.68	85.15	86.82	88.72	85.59	84.71	89.20	87.61	87.22	86.69	87.73	88.72	88.60	86.42
Run 15	81.51	76.06	80.45	75.84	77.96	86.01	86.82	88.36	86.69	88.11	89.43	88.96	86.56	88.60	87.22	88.96	86.96	87.09
Run 16	77.55	79.52	79.52	81.16	81.85	87.35	87.48	87.61	87.86	83.96	89.08	89.43	88.48	88.36	86.56	87.73	87.73	87.99
Run 17	75.40	76.50	82.19	79.14	81.68	86.28	85.15	85.44	87.61	89.20	89.08	91.09	87.35	86.15	88.96	86.96	90.33	86.15
Run 18	78.16	77.13	81.68	77.34	80.08	84.42	86.01	84.42	89.08	86.01	89.20	86.69	87.99	86.42	85.87	86.56	87.86	86.82
Run 19	75.84	80.27	81.68	74.94	79.52	84.11	86.42	89.77	86.69	84.42	90.11	88.24	88.11	86.56	88.72	86.69	89.55	87.61
Run 20	79.14	81.16	83.01	84.11	80.08	82.02	85.15	86.15	87.61	88.36	90.11	89.20	85.01	88.36	87.48	88.11	87.61	86.82
Run 21	76.06	76.71	82.69	77.96	80.63	83.49	88.96	90.00	87.61	86.96	88.24	87.35	89.08	87.99	88.36	86.28	85.30	86.96
Run 22	78.36	80.45	75.40	74.94	84.11	83.65	86.01	88.24	86.42	87.35	89.77	87.22	87.61	85.87	88.11	87.22	87.22	85.73
Run 23	79.52	78.95	77.75	80.63	79.71	84.11	86.15	85.01	88.11	86.96	90.00	87.22	87.61	85.30	87.99	88.60	85.44	87.61
Run 24	78.95	79.90	74.94	82.69	82.52	85.44	87.35	85.59	88.48	88.11	88.36	88.48	86.28	87.09	88.24	87.99	87.86	86.69
Run 25	83.49	80.45	76.50	75.62	79.90	82.02	85.59	85.87	86.56	86.28	87.86	87.73	88.24	87.22	88.48	86.28	88.11	88.24
Run 26	78.16	73.29	78.16	77.96	83.80	81.68	83.01	88.60	87.73	86.69	88.48	87.99	88.72	88.48	87.35	86.42	88.11	87.35
Run 27	75.40	79.71	82.69	82.19	82.02	85.15	85.87	83.65	85.73	85.15	87.73	88.24	86.28	88.36	89.55	86.56	87.73	87.48
Run 28	76.92	72.79	76.92	77.34	80.45	81.34	84.57	88.11	87.48	90.00	90.22	88.36	88.72	87.35	85.87	87.86	88.24	86.69
Run 29	77.96	79.14	78.75	76.28	80.08	85.73	86.82	86.15	88.96	86.01	88.84	90.11	86.01	86.42	87.22	87.09	87.09	86.42
Run 30	76.71	70.72	76.92	76.50	81.34	84.42	85.44	88.11	88.36	87.86	90.55	89.31	88.24	86.15	87.61	87.09	86.96	87.48
Mean	78.36	78.62	78.89	79.17	81.21	84.64	86.12	86.72	87.34	86.75	89.11	88.47	87.36	87.15	87.78	87.33	87.43	87.25
Std dev	0.0211	0.0311	0.0296	0.0261	0.0163	0.0172	0.0133	0.0190	0.0122	0.0164	0.0094	0.0118	0.0144	0.0086	0.0109	0.0100	0.0124	0.0090

Table B-9: Detailed results for the 30 runs of the cross-validation experiments using the sound processing system and the TAL_by_9 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	79.71	77.13	75.17	76.92	81.85	85.44	85.30	84.42	83.65	77.96	83.65	83.80	83.49	83.96	85.73	83.96	83.80	82.85
Run 2	76.28	77.55	80.27	74.94	81.85	84.42	83.80	86.96	82.69	82.36	84.71	82.85	83.49	80.81	80.99	82.85	83.49	83.96
Run 3	79.14	80.81	82.19	72.29	81.51	87.48	86.15	86.28	80.99	82.85	85.01	83.65	85.73	83.01	82.85	83.33	86.56	83.33
Run 4	80.63	79.33	77.55	74.94	84.86	86.56	88.11	85.15	83.33	78.36	82.52	85.87	84.42	86.42	81.68	83.80	83.49	81.51
Run 5	78.16	79.90	77.13	76.71	82.69	85.15	84.71	83.80	83.49	83.17	84.42	82.36	84.26	81.51	85.15	84.86	83.49	81.51
Run 6	76.71	77.34	77.13	78.16	79.90	82.85	85.15	83.17	83.96	77.13	83.96	83.65	83.80	84.57	83.01	82.52	80.81	83.80
Run 7	74.48	75.62	74.48	76.71	80.27	85.87	85.01	81.34	84.42	80.63	85.01	83.49	82.36	84.71	82.19	82.85	82.36	84.71
Run 8	75.40	78.95	77.55	76.28	85.87	83.17	84.42	85.73	80.63	78.36	83.17	83.17	82.02	85.15	85.30	84.11	85.01	81.68
Run 9	78.95	81.68	80.45	76.50	83.49	85.30	84.26	86.01	78.36	81.85	84.42	83.33	82.02	82.52	77.75	85.59	83.33	83.49
Run 10	76.50	74.01	73.04	75.17	80.99	87.61	87.09	83.17	82.85	77.13	82.52	82.36	83.96	83.96	81.85	82.85	83.49	83.49
Run 11	74.94	76.92	77.34	76.06	78.75	84.71	85.15	84.57	78.75	81.85	87.09	82.69	81.51	82.69	85.44	81.34	82.52	85.59
Run 12	77.55	73.53	76.92	72.79	83.17	83.65	84.57	86.96	84.26	76.28	84.26	83.33	84.86	84.71	83.01	82.52	83.96	81.34
Run 13	77.75	74.48	77.75	74.48	81.85	86.56	84.86	83.96	83.33	77.75	83.96	83.65	83.49	83.65	83.33	84.71	86.42	85.01
Run 14	74.94	76.28	72.03	80.45	76.92	85.59	87.86	86.69	80.99	82.85	85.01	84.42	83.33	85.01	82.02	82.19	83.96	85.73
Run 15	77.13	77.55	79.71	74.25	81.85	84.71	86.28	84.11	83.80	80.45	86.01	83.17	83.80	83.01	84.11	84.26	84.57	82.02
Run 16	76.28	77.13	74.94	73.29	83.65	86.42	85.44	84.11	81.68	84.86	83.33	84.26	81.85	85.44	82.36	81.51	84.42	82.36
Run 17	78.16	80.45	82.52	77.13	80.99	85.01	86.96	87.35	81.68	84.42	83.49	84.42	84.26	81.68	85.44	82.69	82.52	83.33
Run 18	75.17	78.16	74.71	76.92	82.85	87.35	84.11	84.57	82.02	83.01	84.57	84.26	83.01	83.96	83.96	81.68	84.11	83.96
Run 19	75.84	73.04	79.33	72.03	81.34	86.96	85.59	86.69	79.90	81.51	86.01	82.85	82.36	82.19	83.96	84.57	84.86	85.01
Run 20	79.14	77.96	80.08	79.33	85.01	88.48	86.01	86.82	83.65	82.69	84.86	82.69	82.19	82.69	81.51	84.11	84.86	83.65
Run 21	79.71	77.13	78.16	77.96	80.08	84.57	87.73	85.87	81.34	81.85	84.71	85.30	82.36	83.17	83.96	83.49	84.57	83.80
Run 22	76.92	78.75	75.17	74.48	82.02	83.65	86.96	83.80	83.17	83.17	83.01	83.96	82.85	82.85	85.59	84.11	82.02	81.51
Run 23	74.48	72.54	78.75	79.71	84.11	87.09	86.69	86.96	80.81	81.85	81.16	83.65	83.65	84.26	84.71	81.68	83.33	84.57
Run 24	78.75	74.01	80.27	72.54	81.68	83.17	85.59	85.15	82.36	80.81	85.15	80.27	83.80	85.01	80.99	83.80	82.36	85.01
Run 25	74.94	79.90	76.28	72.29	81.51	84.71	85.59	85.87	82.69	82.36	82.19	85.73	85.30	84.26	83.01	84.26	83.65	83.65
Run 26	78.95	73.53	78.56	75.84	82.52	86.28	86.01	84.42	80.99	82.85	85.73	83.80	82.85	82.69	82.85	85.87	83.49	83.80
Run 27	75.40	74.25	74.71	75.17	85.15	85.44	87.61	84.57	83.49	83.01	86.96	81.16	86.01	83.65	83.33	82.02	82.69	83.96
Run 28	75.84	72.79	80.08	75.62	83.65	84.57	83.80	84.26	82.69	82.02	83.80	83.96	84.42	82.52	83.33	81.85	83.01	83.01
Run 29	75.62	74.94	73.77	74.71	86.01	86.15	87.22	85.73	83.96	78.95	84.26	83.17	83.65	86.15	83.80	85.15	83.96	84.42
Run 30	76.06	71.25	77.75	73.53	82.02	86.82	86.15	85.59	82.36	84.26	85.87	83.80	82.36	82.85	81.85	85.15	83.33	84.86
Mean	76.98	76.56	77.46	75.57	82.28	85.52	85.81	85.14	82.28	81.22	84.36	83.50	83.45	83.64	83.17	83.46	83.68	83.56
Std dev	0.0175	0.0273	0.0261	0.0219	0.0201	0.0143	0.0122	0.0140	0.0155	0.0235	0.0134	0.0113	0.0114	0.0134	0.0171	0.0127	0.0118	0.0124

Table B-10: Detailed results for the 30 runs of the cross-validation experiments using the sound processing system and the TAL_by_10 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	78.95	76.92	81.16	81.34	77.34	78.56	77.13	80.99	79.71	81.16	78.75	76.06	71.78	75.62	74.94	75.84	77.13	74.71
Run 2	79.33	75.62	81.85	79.33	77.34	75.84	80.63	80.99	79.52	77.13	76.71	77.34	76.92	77.13	77.13	73.53	72.79	74.25
Run 3	79.33	74.25	82.36	81.85	79.14	75.17	75.84	79.71	78.75	78.36	76.71	77.75	72.54	75.84	77.34	73.29	72.29	74.48
Run 4	76.50	80.45	80.27	84.71	79.14	73.29	78.36	78.75	79.33	75.84	76.28	77.55	74.71	74.25	73.29	73.04	75.40	76.28
Run 5	78.75	75.40	80.81	80.45	78.16	75.40	78.36	78.36	82.69	78.16	79.71	77.34	76.92	75.17	75.84	77.13	74.94	76.06
Run 6	76.06	66.98	79.90	84.26	78.56	77.96	79.33	81.51	81.85	80.99	79.71	76.28	72.79	75.40	77.96	76.92	72.79	78.36
Run 7	74.71	78.56	82.19	83.80	79.14	76.50	79.33	77.96	79.52	79.33	78.75	77.96	74.25	76.71	75.17	75.84	76.92	75.17
Run 8	78.36	74.25	79.90	81.68	80.45	78.56	78.36	78.75	80.27	83.33	74.25	73.29	74.71	70.72	76.50	76.92	74.25	72.79
Run 9	77.75	72.54	80.27	82.19	80.99	77.34	80.08	80.08	80.45	79.33	77.13	75.84	72.54	71.52	75.62	73.77	76.50	73.04
Run 10	77.13	77.34	80.45	82.02	78.36	75.17	77.75	77.75	82.85	79.90	74.48	75.62	75.84	76.71	75.17	73.77	76.92	74.25
Run 11	78.36	77.13	81.51	81.16	75.40	76.71	77.34	80.27	80.27	79.14	78.16	77.34	75.62	75.84	78.95	76.71	73.53	73.77
Run 12	75.17	77.13	80.63	83.17	79.90	81.16	79.33	76.28	78.56	79.14	77.55	76.92	79.90	79.52	76.92	74.71	75.62	76.06
Run 13	78.56	74.48	80.81	83.33	78.16	77.55	80.63	76.28	80.45	79.14	78.36	75.40	77.75	78.56	76.92	76.06	77.13	76.28
Run 14	79.90	72.79	80.08	84.86	80.45	78.75	77.55	78.95	81.68	81.51	79.14	78.75	78.75	75.62	76.92	78.36	75.62	75.84
Run 15	77.96	74.94	80.08	82.36	78.16	77.75	79.52	79.90	79.52	78.56	77.75	76.28	76.92	78.36	76.71	73.29	76.06	75.17
Run 16	79.33	76.92	79.33	82.69	80.45	78.16	77.13	79.90	80.63	81.51	75.17	76.06	75.62	77.13	72.54	76.71	76.92	76.50
Run 17	80.81	78.36	81.34	82.19	80.63	77.96	79.71	81.34	76.92	81.68	71.52	76.28	75.17	77.34	74.48	75.17	74.94	73.77
Run 18	76.92	77.34	78.56	81.85	80.81	78.95	79.33	78.56	80.27	80.08	78.36	75.40	77.13	72.29	75.62	74.48	74.25	77.75
Run 19	79.33	72.79	77.55	81.51	78.75	77.13	77.13	78.16	80.81	79.33	74.94	76.92	74.25	72.79	75.84	74.94	76.71	75.17
Run 20	75.62	74.71	80.81	84.26	80.08	79.71	75.62	77.13	82.19	79.71	73.77	73.53	74.25	74.71	76.71	77.55	77.55	72.79
Run 21	81.16	74.25	79.33	81.51	80.08	80.81	79.90	83.17	80.45	77.75	73.77	78.16	75.84	78.56	74.94	75.17	70.72	73.29
Run 22	76.28	79.90	78.95	86.69	80.45	80.63	77.75	75.40	77.34	78.36	79.33	74.25	78.36	74.01	74.71	72.29	77.13	74.94
Run 23	79.14	72.29	80.99	81.16	76.50	77.13	78.95	78.36	80.99	79.33	77.55	77.55	74.01	74.94	73.04	75.17	76.28	75.17
Run 24	76.28	79.14	82.36	82.85	78.56	79.90	76.92	79.52	78.75	80.27	76.92	72.03	73.04	76.71	77.55	76.50	74.48	71.52
Run 25	78.56	77.75	80.63	79.71	77.75	76.06	78.36	77.34	78.95	80.81	79.14	77.13	74.94	73.53	80.63	73.29	75.40	77.75
Run 26	77.96	78.36	82.36	82.85	77.96	76.92	78.95	80.27	81.16	77.96	80.27	80.08	75.40	74.48	73.04	74.01	76.28	72.03
Run 27	79.14	70.45	80.27	82.02	79.14	80.08	75.62	78.75	79.71	79.33	78.36	76.28	76.92	74.25	75.40	79.71	74.71	77.55
Run 28	78.56	74.25	81.16	83.49	78.36	76.71	78.36	77.55	80.81	81.85	76.71	76.92	74.01	73.04	76.92	77.34	75.17	77.75
Run 29	78.95	77.13	78.95	82.02	78.56	80.99	80.99	76.50	78.36	77.55	78.95	75.40	76.28	76.50	76.71	71.78	75.84	76.06
Run 30	76.50	76.28	77.55	84.26	78.95	79.52	77.34	81.51	79.33	79.90	78.36	77.34	74.01	78.56	76.50	76.50	76.50	74.01
Mean	78.05	75.62	80.41	82.52	78.92	77.88	78.39	79.00	80.07	79.55	77.22	76.44	75.37	75.53	76.00	75.33	75.36	75.09
Std dev	0.0158	0.0287	0.0126	0.0156	0.0132	0.0192	0.0143	0.0180	0.0140	0.0158	0.0209	0.0163	0.0193	0.0216	0.0173	0.0186	0.0163	0.0176

TABLE B-11: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM WITH DYNAMIC BLOCK THRESHOLDS AND THE MNI_BY_3 TEMPLATE IN %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	37.61	41.76	39.30	42.95	39.72	37.18	34.09	43.73	34.54	36.31	37.61	41.35	38.04	34.99	37.18	35.87	36.75	36.75
Run 2	45.27	42.16	41.76	41.76	35.43	34.54	38.88	43.73	44.50	36.75	38.04	28.44	37.61	35.43	38.04	36.75	38.04	38.88
Run 3	34.99	39.30	38.88	38.04	39.30	38.04	40.54	37.18	37.18	34.09	36.31	38.46	39.72	34.09	41.35	38.04	37.61	32.26
Run 4	46.02	41.35	33.64	40.54	40.13	41.35	35.87	37.61	38.04	40.13	36.31	41.76	39.72	35.87	37.18	32.26	34.54	41.76
Run 5	39.72	42.95	40.13	41.35	41.35	44.50	38.46	34.09	40.13	32.72	36.31	36.31	38.04	39.72	33.18	34.99	38.46	38.04
Run 6	43.34	40.54	40.13	37.18	40.95	39.30	37.61	37.61	40.13	36.75	38.04	36.31	38.88	38.46	41.35	37.61	35.87	36.31
Run 7	40.54	41.76	42.55	33.64	42.95	41.35	42.95	40.95	35.43	41.35	34.09	36.31	34.09	37.18	39.30	42.55	37.61	39.30
Run 8	41.35	34.54	43.73	38.46	40.13	40.54	38.04	40.54	35.43	37.61	39.30	35.87	38.88	35.43	42.95	35.43	38.46	39.72
Run 9	42.95	42.55	41.35	37.18	43.73	38.88	31.32	36.31	34.99	39.72	36.75	38.46	37.18	38.04	40.95	36.75	40.13	38.04
Run 10	42.95	40.95	44.50	43.73	38.04	39.72	37.61	38.04	36.31	37.61	37.18	42.16	41.35	40.95	37.61	32.26	37.61	35.87
Run 11	42.55	42.55	43.34	43.34	37.61	39.72	37.18	31.79	35.87	36.75	39.30	39.72	34.99	40.54	42.16	40.54	41.35	39.72
Run 12	41.76	34.54	43.34	37.61	40.95	38.04	37.18	40.13	44.12	34.54	34.54	44.12	35.43	33.64	36.31	36.31	37.18	39.30
Run 13	41.76	45.27	38.88	40.13	38.46	38.46	39.30	34.54	36.75	40.54	38.04	38.88	36.75	38.88	39.72	36.75	37.18	35.87
Run 14	34.54	42.95	38.04	40.95	39.72	42.95	39.72	42.95	36.31	40.13	38.46	35.43	36.31	37.61	36.75	38.04	34.99	34.99
Run 15	41.35	40.13	41.35	38.46	47.50	32.72	37.61	38.04	33.18	40.13	42.16	35.87	37.61	35.87	33.64	40.54	38.88	33.64
Run 16	38.46	40.54	38.88	46.02	44.50	38.46	31.79	39.30	39.30	40.13	43.73	40.13	41.76	35.87	33.64	32.26	31.32	37.18
Run 17	43.34	39.30	41.76	40.13	40.95	40.95	40.54	40.54	40.95	40.13	34.09	34.99	36.31	38.46	40.95	37.18	40.13	41.76
Run 18	40.95	42.55	41.35	37.61	40.13	32.26	34.09	34.99	38.04	38.88	33.64	40.13	42.55	36.31	38.04	39.30	35.43	40.13
Run 19	41.76	40.95	42.55	39.30	39.30	38.04	38.46	38.46	38.46	41.35	34.09	34.99	41.76	40.54	38.46	42.16	38.04	40.54
Run 20	40.13	35.43	41.76	44.50	44.89	34.09	38.88	38.04	39.30	42.95	34.09	37.61	39.72	41.35	35.87	38.46	38.04	38.88
Run 21	41.35	44.50	38.04	45.64	40.95	37.18	39.30	42.95	38.46	33.18	39.30	37.61	27.45	30.85	34.99	38.04	36.75	37.18
Run 22	40.13	40.54	38.04	40.13	46.02	38.88	43.34	37.18	39.72	38.04	33.64	33.64	41.35	40.54	40.13	39.72	36.75	42.95
Run 23	41.76	42.55	41.76	42.55	40.54	38.88	42.95	39.30	41.76	41.76	39.72	38.04	40.54	37.18	42.95	40.54	38.46	38.88
Run 24	43.73	42.55	33.64	38.04	40.54	33.64	38.88	38.88	39.30	32.26	38.88	35.43	39.72	39.72	37.61	37.18	32.26	35.43
Run 25	39.72	39.30	42.55	42.95	35.43	35.87	41.76	34.09	37.61	35.43	32.72	42.95	41.76	37.18	39.72	37.61	35.87	40.54
Run 26	41.35	37.61	40.54	35.43	29.89	43.34	36.75	36.31	42.95	38.04	37.61	34.99	40.13	36.75	37.18	38.04	37.18	35.43
Run 27	39.72	40.95	41.76	40.54	38.46	38.46	39.72	34.54	42.16	34.99	38.46	36.31	39.30	41.35	40.13	39.30	35.87	33.64
Run 28	35.43	35.43	41.76	40.54	37.61	36.31	40.13	38.88	34.09	38.46	39.30	38.46	33.64	38.46	39.72	40.54	38.88	36.31
Run 29	38.88	40.54	40.95	37.18	38.46	40.54	42.16	35.87	32.72	40.13	40.95	34.99	40.95	37.18	40.13	35.43	40.95	38.88
Run 30	34.54	36.75	40.13	39.30	39.72	41.35	40.95	34.99	38.46	41.35	39.72	40.54	41.35	32.26	37.61	33.64	41.35	42.95
Mean	40.60	40.43	40.55	40.17	40.11	38.52	38.54	38.05	38.21	38.07	37.41	37.68	38.43	37.36	38.49	37.47	37.40	38.04
Std dev	0.0288	0.0276	0.0251	0.0291	0.0333	0.0299	0.0294	0.0299	0.0302	0.0286	0.0266	0.0317	0.0314	0.0263	0.0261	0.0268	0.0229	0.0270

Table B-12: Detailed results for the 30 runs of the cross-validation experiments using the video processing system with dynamic block thresholds and the MNI_by_4 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	38.46	36.31	37.18	40.95	40.54	35.43	44.12	50.35	38.46	39.30	30.85	43.73	36.31	42.55	40.13	40.13	40.95	35.87
Run 2	38.46	39.72	43.73	36.75	40.54	42.16	44.12	43.34	39.72	44.12	39.30	42.16	45.64	42.55	43.34	43.34	37.18	41.35
Run 3	37.61	35.87	35.43	38.04	36.75	43.73	38.04	44.50	40.54	40.13	43.34	40.95	38.46	41.76	44.89	35.87	43.73	39.72
Run 4	34.54	40.54	38.04	42.16	39.72	45.64	44.12	38.46	42.55	34.99	44.12	35.43	40.13	47.50	37.61	46.76	35.87	38.46
Run 5	33.18	35.43	43.73	38.46	38.46	44.50	39.30	46.39	37.18	41.76	44.12	44.89	42.55	44.89	43.34	44.89	39.30	41.76
Run 6	45.27	43.73	38.46	37.18	35.87	43.73	41.76	41.76	35.87	38.04	37.18	38.88	38.46	42.95	44.50	40.95	40.54	44.89
Run 7	38.04	40.13	40.54	39.30	39.72	40.54	35.43	41.35	35.87	39.30	34.54	33.64	46.39	40.13	36.75	42.55	39.72	44.12
Run 8	42.95	36.75	43.34	41.76	38.04	43.34	37.61	44.50	40.54	40.13	40.95	40.13	48.22	40.13	34.99	42.16	39.30	42.95
Run 9	41.76	38.88	42.55	40.95	40.54	34.99	42.16	42.95	42.55	41.76	47.86	41.35	39.30	40.54	43.34	38.88	44.50	40.54
Run 10	40.54	33.18	40.54	37.18	32.26	34.09	42.95	36.31	37.18	38.88	41.35	38.46	44.50	41.35	37.61	40.54	37.18	40.95
Run 11	37.61	41.76	42.16	42.16	42.95	43.34	35.43	38.46	38.88	40.54	37.18	46.02	38.88	38.88	45.27	41.76	42.16	42.55
Run 12	37.61	40.54	43.34	37.18	40.13	39.72	38.88	42.55	44.50	41.76	42.95	46.76	39.72	40.95	44.89	42.16	44.89	38.04
Run 13	34.99	34.09	39.30	43.73	40.13	41.76	39.30	45.64	38.88	38.04	40.13	38.04	44.89	38.04	40.13	44.89	34.09	38.04
Run 14	35.43	37.61	35.43	36.75	36.31	39.72	34.09	37.18	42.95	40.95	37.61	41.76	45.27	44.12	46.76	34.09	43.73	41.76
Run 15	38.04	41.35	44.50	43.73	40.13	41.35	39.30	42.95	38.04	38.46	39.72	42.55	40.13	40.54	39.72	43.73	35.87	38.88
Run 16	41.35	36.31	45.64	40.13	40.54	39.72	43.34	40.95	42.16	44.50	43.34	43.34	40.13	43.73	45.64	42.55	40.13	42.16
Run 17	38.88	33.64	36.75	38.46	38.46	39.72	42.55	39.30	37.18	40.13	42.16	37.18	46.02	36.75	44.12	44.89	45.64	43.73
Run 18	40.13	37.61	39.72	43.73	30.37	38.46	43.34	33.64	42.95	38.04	40.13	39.30	43.34	44.89	42.95	43.73	42.95	39.30
Run 19	33.64	40.95	36.75	35.87	40.95	38.04	42.16	39.72	46.02	40.54	39.30	41.35	40.95	42.16	45.27	39.30	40.54	39.30
Run 20	38.46	37.61	40.13	42.95	37.18	41.76	47.86	38.46	35.87	42.55	41.35	40.13	42.55	41.35	38.04	42.95	43.34	41.76
Run 21	43.34	41.35	42.95	40.54	39.72	41.76	40.95	38.88	39.72	35.87	42.95	43.34	40.13	40.54	43.73	43.73	40.54	36.75
Run 22	40.54	39.72	34.99	34.09	40.95	40.54	41.76	35.43	40.13	42.95	41.35	40.54	41.35	40.54	42.95	38.04	42.16	41.76
Run 23	33.18	35.43	36.31	40.95	43.73	37.18	44.12	40.13	44.12	37.61	44.12	39.72	40.95	37.61	38.88	44.12	43.73	39.72
Run 24	42.95	45.64	40.95	42.16	38.88	41.76	40.95	40.95	35.87	40.13	41.35	42.55	42.55	38.88	42.55	40.95	44.12	45.64
Run 25	37.18	35.43	34.99	43.73	41.35	40.13	38.88	42.95	39.30	38.04	41.35	39.72	42.95	44.89	44.50	46.39	42.16	41.76
Run 26	40.54	35.87	35.87	39.72	43.73	46.76	43.73	38.88	37.18	41.35	38.46	39.72	44.89	38.04	41.76	46.39	42.16	40.95
Run 27	44.12	37.61	36.31	43.34	41.35	40.54	40.95	43.34	37.18	39.30	35.87	43.73	44.12	43.34	35.87	39.30	44.89	39.30
Run 28	38.04	40.54	36.75	44.50	42.16	41.76	40.95	36.31	43.34	38.88	40.54	42.16	38.88	42.16	42.95	38.04	42.55	40.95
Run 29	44.50	39.30	37.18	47.13	40.54	42.95	38.88	35.43	38.88	39.72	44.89	39.30	35.87	40.13	45.64	45.64	40.54	47.86
Run 30	40.13	42.55	40.13	43.73	42.55	41.35	44.12	38.46	45.64	41.35	44.50	46.76	42.16	38.46	46.39	44.12	43.73	42.95
Mean	39.05	38.52	39.46	40.58	39.49	40.88	41.04	40.65	39.98	39.97	40.76	41.12	41.86	41.34	42.15	42.09	41.27	41.12
Std dev	0.0330	0.0303	0.0319	0.0304	0.0294	0.0291	0.0303	0.0365	0.0300	0.0213	0.0341	0.0301	0.0301	0.0248	0.0329	0.0308	0.0293	0.0258

TABLE B-13: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM WITH DYNAMIC BLOCK THRESHOLDS AND THE MNI_BY_5 TEMPLATE IN %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	34.09	40.54	37.18	43.34	40.13	37.18	40.95	38.88	44.50	48.58	39.30	40.13	37.61	46.02	44.50	45.27	37.61	43.73
Run 2	42.16	37.61	40.54	43.34	38.04	45.27	34.09	42.16	37.18	43.34	41.76	44.12	40.54	44.89	42.16	33.64	40.13	41.76
Run 3	41.35	45.64	38.88	39.72	38.04	38.88	43.34	40.54	38.88	41.76	44.89	39.30	42.16	44.50	38.04	42.55	38.04	38.04
Run 4	44.12	36.31	41.35	42.16	43.73	43.73	36.75	38.46	44.12	40.95	38.04	43.73	44.12	31.79	45.64	38.46	44.12	42.95
Run 5	39.30	33.64	42.55	44.50	42.55	41.76	44.50	42.95	47.50	45.64	41.35	47.13	42.55	39.72	34.09	39.72	42.16	41.76
Run 6	46.76	41.35	39.30	39.30	40.13	36.75	41.76	45.64	47.13	47.50	37.18	35.43	38.46	44.89	42.16	38.88	46.02	42.95
Run 7	34.99	37.61	42.95	42.95	39.30	41.35	39.72	39.30	42.16	38.88	42.16	43.73	40.13	43.34	36.75	34.54	44.89	42.55
Run 8	42.16	34.09	41.35	39.30	39.30	39.30	40.13	44.12	40.54	40.95	34.09	40.54	42.55	42.55	38.46	42.16	37.61	41.76
Run 9	42.95	33.64	42.16	36.31	43.34	40.54	34.09	47.13	40.13	45.27	39.72	39.30	42.95	37.61	40.13	42.55	34.99	43.34
Run 10	34.99	39.30	40.54	39.72	37.18	44.12	41.35	42.16	44.50	40.13	35.87	36.31	43.73	42.55	40.54	42.55	38.04	39.30
Run 11	37.18	40.54	42.95	44.89	42.95	39.72	46.76	46.02	43.73	44.50	40.95	39.30	42.16	38.46	39.30	34.99	40.54	40.13
Run 12	39.72	37.18	40.13	40.95	38.88	43.34	38.46	38.04	37.18	39.72	38.88	39.72	41.35	45.27	40.54	40.95	38.04	42.55
Run 13	41.35	40.95	36.31	39.72	39.30	38.04	40.54	41.76	39.30	40.13	37.18	41.35	38.88	38.04	38.88	40.13	44.89	46.02
Run 14	38.04	32.72	41.35	36.75	41.76	36.31	40.54	43.73	43.73	41.35	43.73	38.04	41.35	41.35	43.34	44.50	38.04	38.88
Run 15	31.79	36.31	42.95	44.12	36.31	37.61	44.50	42.16	40.95	46.39	42.95	38.04	40.95	42.95	44.12	44.12	42.95	41.35
Run 16	35.87	37.18	39.72	42.55	40.95	41.76	42.55	42.16	40.95	40.95	40.13	46.02	45.27	43.73	47.13	41.35	38.04	40.95
Run 17	34.99	41.35	40.54	38.88	50.00	40.54	34.54	38.04	42.16	36.75	40.13	40.54	35.87	37.61	44.89	43.34	45.64	41.76
Run 18	42.95	41.76	37.18	40.95	40.13	38.88	38.88	46.76	40.13	39.30	44.50	44.50	38.04	41.35	44.50	39.72	42.16	40.13
Run 19	43.34	36.75	36.31	41.35	42.55	43.34	43.73	44.89	38.46	35.43	42.95	40.13	37.18	36.31	37.61	38.04	42.16	43.73
Run 20	37.61	44.50	40.54	39.30	39.72	43.34	45.64	44.50	40.95	40.54	42.55	40.13	38.88	42.16	42.16	36.75	41.76	43.73
Run 21	37.18	35.87	41.35	42.95	42.55	42.55	43.73	43.73	37.61	35.87	45.27	39.72	44.89	38.46	40.95	40.13	40.95	37.18
Run 22	37.61	37.18	39.30	36.31	43.34	44.12	42.16	40.13	44.89	43.34	42.55	41.76	41.35	40.13	39.72	40.54	40.13	42.16
Run 23	39.30	34.09	37.18	41.35	40.13	33.64	41.35	43.73	40.13	38.04	38.88	42.55	40.13	36.31	38.88	44.89	35.87	42.16
Run 24	40.95	38.46	40.13	34.99	40.13	36.31	40.54	40.54	42.55	38.88	41.76	40.95	42.95	40.13	46.02	40.54	41.35	44.50
Run 25	38.04	39.30	37.61	42.95	44.50	46.39	42.16	37.61	38.04	43.73	38.04	41.35	34.99	44.50	42.16	42.95	35.43	41.35
Run 26	35.43	40.13	37.18	42.16	41.76	41.35	36.75	38.88	45.27	43.73	47.50	38.88	41.76	41.35	35.87	38.04	39.30	43.34
Run 27	43.34	34.99	36.75	37.18	36.75	38.04	43.34	45.27	40.13	34.99	38.04	43.73	42.95	40.54	37.61	40.95	41.76	43.34
Run 28	40.13	38.04	45.64	39.72	38.46	40.54	37.61	42.95	38.46	45.64	41.76	38.46	38.04	43.34	38.88	39.30	41.35	42.16
Run 29	40.13	34.99	39.72	48.94	45.27	45.27	42.55	44.89	42.95	42.95	40.13	46.02	42.55	32.26	44.12	48.94	42.16	44.50
Run 30	40.95	42.16	42.16	41.76	39.72	38.88	42.16	34.54	40.13	42.16	38.88	41.76	38.04	40.95	40.13	38.46	41.35	40.95
Mean	39.29	38.14	40.06	40.95	40.90	40.63	40.84	42.05	41.48	41.58	40.70	41.09	40.75	40.77	40.98	40.63	40.58	41.97
Std dev	0.0343	0.0323	0.0231	0.0295	0.0283	0.0309	0.0323	0.0306	0.0282	0.0346	0.0293	0.0278	0.0260	0.0355	0.0319	0.0329	0.0293	0.0193

Table B-14: Detailed results for the 30 runs of the cross-validation experiments using the video processing system with dynamic block thresholds and the TAL_by_5 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	35.43	38.88	36.75	41.35	40.54	42.95	37.61	38.88	39.72	46.02	39.30	39.30	43.34	44.12	47.13	44.12	46.02	46.39
Run 2	33.64	44.12	40.95	46.39	44.50	44.50	40.95	46.76	44.12	43.34	42.95	38.46	41.35	40.54	42.55	40.54	42.95	45.64
Run 3	41.35	39.30	43.73	39.30	44.50	44.50	37.61	42.95	40.95	41.35	40.95	40.95	42.16	41.76	46.76	40.54	46.39	38.04
Run 4	39.30	40.95	43.34	39.72	43.34	42.95	38.46	46.39	42.95	39.30	42.16	41.76	38.88	46.39	42.95	42.16	44.50	38.88
Run 5	37.61	42.95	42.95	38.46	46.02	42.55	42.95	46.76	43.73	40.95	41.35	42.16	44.50	40.95	45.27	43.34	47.13	47.13
Run 6	42.16	44.50	44.89	41.76	44.12	40.13	41.76	41.76	40.13	41.35	44.50	41.35	45.64	43.34	38.88	43.73	48.94	41.76
Run 7	34.09	39.30	40.54	42.55	43.34	35.87	44.12	40.13	40.95	42.16	39.30	42.16	40.13	46.76	42.95	47.13	37.61	44.12
Run 8	40.54	43.73	35.43	39.72	42.16	42.55	43.34	45.64	40.54	44.89	46.39	42.55	38.46	45.64	46.02	45.64	35.87	45.64
Run 9	38.46	42.95	40.54	44.89	41.35	42.55	38.88	40.54	48.94	45.64	42.95	40.54	38.46	47.13	37.18	43.73	42.95	43.73
Run 10	43.73	39.72	40.54	43.73	45.27	44.50	40.54	46.39	42.55	37.61	38.88	45.27	42.55	40.13	40.54	36.75	37.18	42.55
Run 11	34.99	44.89	42.16	41.76	46.39	42.16	36.75	45.27	43.34	40.95	34.99	44.12	44.50	43.73	39.72	44.89	43.34	47.86
Run 12	33.64	34.09	35.43	42.95	42.16	39.30	39.30	42.55	38.88	38.88	39.72	42.95	41.35	37.61	43.73	42.55	41.35	44.12
Run 13	38.88	46.02	43.73	41.35	38.46	41.76	40.54	44.50	44.50	41.35	44.12	42.16	39.30	46.39	48.22	47.13	46.39	44.50
Run 14	37.18	41.76	41.35	44.50	40.54	41.76	38.46	38.88	48.22	38.46	42.55	42.55	45.27	41.76	38.46	43.34	43.73	42.95
Run 15	35.87	44.12	43.73	37.18	40.54	42.16	41.76	39.72	42.95	39.30	42.16	42.55	38.88	43.34	36.31	43.73	38.88	48.22
Run 16	36.31	46.39	40.95	37.18	42.95	38.04	42.16	38.04	42.95	41.76	38.88	40.95	39.72	41.35	41.35	46.39	44.50	50.00
Run 17	38.88	40.54	49.30	42.16	42.16	44.50	42.95	40.95	43.73	42.55	44.12	40.54	35.87	39.72	38.04	43.34	47.13	50.70
Run 18	42.16	41.76	45.27	45.27	45.64	40.13	44.50	40.95	44.12	43.73	44.12	44.50	42.16	48.22	40.54	47.13	38.88	44.89
Run 19	37.18	37.18	39.72	46.02	44.89	34.09	43.34	42.95	47.50	44.12	45.64	43.73	37.61	42.95	37.61	41.76	37.61	50.00
Run 20	38.88	39.72	38.04	40.13	34.09	43.34	43.34	44.12	43.73	45.64	39.30	42.95	41.35	44.12	41.76	37.61	46.76	44.50
Run 21	36.31	33.64	42.16	39.30	37.61	42.55	38.04	39.72	40.54	45.27	43.73	38.88	43.34	39.30	45.64	40.13	44.50	37.18
Run 22	43.73	38.46	42.55	41.35	36.31	41.76	46.39	42.55	45.27	41.35	48.22	46.39	42.95	39.30	42.55	42.55	42.95	41.76
Run 23	40.54	39.72	44.12	42.95	40.13	35.43	41.76	45.64	43.34	42.16	40.95	39.30	42.16	40.54	41.76	41.76	43.34	50.00
Run 24	40.95	43.34	38.88	42.95	44.50	38.88	39.72	42.55	40.95	42.55	42.55	43.34	40.54	46.76	46.02	39.30	46.02	46.39
Run 25	39.30	43.73	43.73	38.04	45.27	44.50	41.35	41.35	46.02	43.73	47.13	43.34	47.86	43.34	46.76	44.89	44.89	44.12
Run 26	38.04	46.02	43.34	43.34	44.89	42.95	41.76	47.86	42.95	44.12	39.30	41.35	47.13	41.76	38.46	42.95	45.64	41.76
Run 27	36.75	37.18	47.86	43.73	36.75	38.88	36.75	43.34	42.95	41.35	42.16	47.50	36.75	46.76	39.30	40.54	45.27	40.13
Run 28	38.46	35.87	42.95	34.99	40.13	45.64	41.76	43.34	41.35	42.95	40.54	45.27	47.13	42.95	42.55	43.73	42.95	44.89
Run 29	38.46	42.16	37.61	36.75	39.30	38.88	41.76	44.50	41.76	44.12	42.16	39.72	44.50	43.34	43.34	41.76	48.58	41.76
Run 30	38.88	46.02	39.72	38.88	40.13	40.13	45.27	41.76	38.04	47.50	41.76	42.95	41.76	40.95	39.72	42.16	44.12	44.50
Mean	38.39	41.30	41.74	41.29	41.93	41.33	41.13	42.89	42.92	42.48	42.09	42.32	41.85	43.03	42.07	42.84	43.55	44.47
Std dev	0.0270	0.0346	0.0319	0.0287	0.0313	0.0285	0.0250	0.0264	0.0254	0.0237	0.0277	0.0215	0.0307	0.0273	0.0328	0.0255	0.0342	0.0344

Table B-15: Detailed results for the 30 runs of the cross-validation experiments using the video processing system with dynamic block thresholds and the TAL_by_6 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	39.72	39.30	36.75	41.76	35.87	36.31	43.73	38.46	44.12	40.13	38.88	36.75	37.18	46.39	38.46	40.95	41.76	41.76
Run 2	43.34	39.30	38.88	34.09	40.95	41.76	40.13	44.50	45.27	43.34	39.72	36.31	40.13	40.95	39.30	44.12	37.61	42.55
Run 3	39.72	42.55	36.31	38.04	39.30	39.30	40.95	42.16	34.09	40.95	42.16	42.95	37.61	35.87	39.72	37.61	43.34	43.73
Run 4	39.30	39.72	44.89	39.72	41.76	41.76	38.88	38.88	39.30	39.72	42.95	44.12	36.75	42.55	44.12	41.35	43.34	35.43
Run 5	43.34	44.12	48.22	40.95	40.95	40.54	36.31	41.35	42.55	39.72	37.61	38.04	41.76	35.87	40.54	42.55	39.72	36.75
Run 6	38.46	44.89	42.95	45.64	38.88	40.54	44.89	40.95	40.54	35.87	34.09	40.95	37.61	38.46	38.46	38.04	40.95	38.04
Run 7	34.99	44.89	39.72	42.16	42.95	40.54	38.46	42.55	35.43	44.89	44.50	40.54	40.95	45.64	34.54	42.16	45.64	42.55
Run 8	40.95	42.55	35.87	35.87	39.30	41.35	40.13	40.95	40.13	39.72	45.27	40.54	40.13	39.72	41.76	38.46	33.64	36.31
Run 9	41.35	39.72	42.55	42.16	39.30	40.54	44.50	39.72	45.27	40.13	41.76	41.35	38.88	37.61	38.04	43.73	41.35	40.95
Run 10	39.30	37.61	35.87	36.75	41.76	35.87	34.99	36.31	38.88	37.18	39.30	37.18	40.54	40.13	37.61	39.72	44.50	34.99
Run 11	40.13	42.16	34.09	41.35	40.95	42.16	46.39	39.30	37.18	44.12	39.72	38.46	39.72	44.12	34.09	39.30	45.64	39.30
Run 12	35.87	40.13	42.55	43.34	39.30	42.16	43.73	40.13	41.35	43.73	40.95	38.88	41.76	40.54	38.04	37.61	39.72	42.55
Run 13	44.12	41.76	37.61	42.55	42.55	40.13	39.72	42.16	39.30	44.12	34.09	41.76	45.64	35.87	38.46	36.31	41.35	40.13
Run 14	44.12	46.39	39.72	41.35	45.64	37.61	44.89	40.13	36.75	46.02	38.04	42.95	44.12	38.46	42.95	37.61	43.73	37.61
Run 15	35.87	38.46	40.95	48.22	35.87	40.54	37.18	42.55	34.09	42.16	36.31	37.18	39.30	35.87	41.76	43.34	42.16	38.46
Run 16	42.16	42.16	40.54	40.54	43.73	42.16	40.13	40.95	42.55	38.88	44.12	42.55	44.12	39.30	40.54	42.55	33.64	40.13
Run 17	38.46	39.30	38.04	42.55	44.12	40.95	38.88	38.46	40.13	40.95	40.95	42.16	41.35	40.13	40.54	44.12	42.16	38.04
Run 18	40.54	39.30	40.95	39.30	42.95	43.34	39.72	38.04	38.04	39.72	43.34	40.95	40.54	40.54	40.95	42.95	40.95	38.04
Run 19	41.76	39.30	41.76	38.46	44.12	41.35	44.89	46.39	38.88	40.95	40.13	43.73	43.73	41.35	42.55	37.61	38.46	40.54
Run 20	39.72	43.34	45.27	46.02	40.54	38.88	40.13	45.27	42.55	34.54	42.16	43.73	35.43	43.34	36.75	36.75	42.55	41.35
Run 21	38.46	42.55	38.04	45.64	43.34	36.75	38.88	49.65	43.34	40.54	36.75	38.88	35.87	40.54	31.79	40.54	43.34	37.61
Run 22	35.87	42.95	42.16	44.89	45.27	46.02	41.35	36.75	43.73	39.72	31.32	40.13	38.46	37.18	34.99	40.54	37.61	36.31
Run 23	39.72	45.64	44.50	41.76	39.30	39.30	38.04	36.75	42.16	38.88	36.31	43.73	39.30	38.88	42.16	44.50	39.72	42.16
Run 24	38.88	40.95	42.16	38.46	41.35	42.55	38.88	38.88	40.13	41.76	44.12	36.75	41.35	38.04	35.87	46.02	38.46	42.95
Run 25	38.46	36.31	39.72	38.46	43.34	43.34	43.73	37.18	41.76	40.95	38.46	42.16	44.89	36.31	43.73	38.88	42.55	38.46
Run 26	34.09	36.75	41.76	44.50	38.46	41.35	39.72	43.73	40.13	44.50	39.72	41.35	44.89	41.35	40.13	38.88	37.18	40.13
Run 27	42.16	42.95	40.95	44.89	44.12	35.87	35.43	42.55	40.54	38.46	39.72	38.04	43.34	42.95	37.18	37.18	44.50	38.88
Run 28	37.18	38.04	38.88	44.12	39.30	39.72	42.55	40.54	38.88	38.46	42.16	38.04	40.95	40.95	46.02	37.61	37.61	36.75
Run 29	43.34	41.76	42.16	41.76	41.35	38.46	42.16	39.72	40.54	36.75	41.76	35.87	41.76	40.95	39.72	42.16	40.13	36.75
Run 30	40.54	39.30	40.54	40.95	38.46	44.12	40.54	41.35	44.89	39.72	36.31	45.64	40.95	38.04	38.46	40.95	40.54	36.31
Mean	39.73	41.14	40.48	41.54	41.17	40.51	40.66	40.88	40.42	40.55	39.76	40.39	40.63	39.93	39.31	40.47	40.80	39.18
Std dev	0.0266	0.0260	0.0309	0.0318	0.0249	0.0239	0.0291	0.0296	0.0297	0.0268	0.0333	0.0266	0.0268	0.0279	0.0315	0.0267	0.0303	0.0246

TABLE B-16: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM WITH DYNAMIC BLOCK THRESHOLDS AND THE TAL_BY_7 TEMPLATE IN %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	41.76	38.04	38.88	34.99	42.16	40.54	44.89	40.54	43.34	46.76	36.75	44.12	48.94	43.73	37.18	42.16	44.50	40.95
Run 2	38.88	37.61	47.13	40.54	40.54	47.50	42.55	38.04	46.39	46.39	39.72	42.16	37.61	42.95	43.34	34.99	40.95	44.50
Run 3	39.72	42.55	45.27	35.87	42.55	42.16	46.39	41.76	47.50	44.89	45.27	44.89	40.13	39.30	43.73	38.04	41.76	40.95
Run 4	36.75	37.18	44.12	38.88	42.95	39.30	47.50	45.64	43.34	44.12	39.72	37.18	38.46	42.16	46.39	39.30	38.46	38.04
Run 5	43.34	40.13	44.12	39.72	34.54	37.61	40.54	43.73	46.39	39.72	40.54	40.95	43.34	40.54	46.39	40.13	43.34	37.18
Run 6	39.30	36.75	39.30	39.72	41.76	43.73	39.30	37.61	39.30	42.95	45.64	41.35	42.16	45.64	42.16	42.55	42.16	47.13
Run 7	33.64	40.13	42.55	38.46	42.16	37.61	41.35	47.13	46.76	44.50	38.46	38.88	43.34	45.64	38.46	37.61	42.16	48.94
Run 8	41.35	36.75	38.04	42.16	42.95	45.64	34.99	39.72	44.12	44.50	37.18	47.50	43.34	47.50	43.34	40.95	37.61	43.34
Run 9	40.95	31.79	43.34	41.76	42.16	39.72	40.95	38.88	42.95	44.12	42.16	42.95	44.50	38.04	38.46	48.22	38.46	37.18
Run 10	32.72	35.43	40.13	38.46	40.54	39.30	40.95	43.34	46.39	39.72	36.31	39.72	38.88	37.61	37.18	44.89	40.95	40.13
Run 11	42.55	40.13	44.12	40.13	42.16	38.04	45.27	40.54	46.76	44.12	42.95	40.95	39.30	42.55	41.76	46.76	42.95	42.55
Run 12	36.31	36.75	40.95	42.16	42.16	39.72	46.39	39.72	41.76	42.16	41.35	42.95	42.16	44.12	39.30	33.18	35.43	38.04
Run 13	38.46	36.75	38.46	46.02	43.34	40.95	42.16	41.35	40.95	39.72	45.27	41.35	45.27	40.95	44.12	35.87	39.72	38.88
Run 14	34.09	39.30	43.34	39.72	37.18	46.02	42.16	44.50	50.35	39.72	44.89	38.46	39.72	46.39	43.34	39.30	46.76	42.16
Run 15	38.88	39.30	42.95	40.13	45.27	47.50	43.34	43.34	42.55	37.61	44.12	40.54	40.95	37.18	46.02	42.16	42.55	38.88
Run 16	40.54	35.87	37.61	37.61	40.95	42.95	34.09	39.30	41.76	31.79	42.55	39.72	42.55	41.76	38.46	39.30	41.35	39.72
Run 17	41.76	36.75	38.04	37.61	38.88	40.54	40.13	41.35	38.46	41.35	40.13	43.73	40.13	43.34	43.73	43.34	43.34	35.43
Run 18	42.55	38.46	38.46	42.95	38.46	46.39	39.72	40.54	46.39	44.89	39.30	42.55	37.18	47.50	32.72	46.76	46.39	45.27
Run 19	40.54	39.72	41.35	39.30	41.35	39.30	44.89	42.16	39.72	47.50	38.04	44.89	41.35	40.54	42.55	44.89	42.55	41.76
Run 20	39.72	35.87	43.34	44.12	36.75	41.35	40.95	43.73	43.34	42.95	39.72	39.30	42.55	46.76	47.86	43.34	37.18	42.95
Run 21	36.75	34.99	41.76	36.75	40.95	44.89	42.55	44.50	43.34	40.54	43.34	34.99	36.75	44.89	39.30	41.76	44.12	37.18
Run 22	35.87	33.18	39.72	37.18	40.54	38.46	45.27	40.95	49.65	43.73	40.95	39.72	49.65	41.76	48.22	41.35	40.95	40.95
Run 23	37.61	31.79	41.76	38.46	36.31	42.16	41.76	46.39	44.50	36.31	46.02	39.72	39.30	42.16	42.55	36.31	43.73	38.46
Run 24	39.72	34.99	38.04	42.16	42.95	38.88	42.16	46.39	44.12	48.58	40.54	41.35	45.27	46.39	39.30	43.34	43.73	41.76
Run 25	41.76	40.13	42.95	42.95	44.89	42.16	38.88	40.13	44.12	38.46	42.95	39.72	45.64	40.54	42.16	43.34	45.27	36.31
Run 26	41.35	37.61	37.61	45.64	42.16	43.34	40.54	41.76	46.02	43.73	44.12	37.18	44.50	41.35	41.35	43.73	41.35	39.72
Run 27	38.04	42.95	38.04	42.55	40.95	42.55	40.13	44.12	42.16	40.54	37.18	42.55	43.73	44.89	43.34	44.50	38.88	41.35
Run 28	38.46	37.18	45.27	38.04	36.75	42.95	42.55	39.30	38.04	42.95	40.13	36.31	42.55	39.30	45.64	44.89	41.76	36.75
Run 29	39.30	39.30	40.54	41.35	37.61	39.30	40.13	42.95	44.12	42.16	41.35	39.72	42.55	47.50	42.16	43.34	41.35	39.30
Run 30	40.95	39.72	35.87	40.95	37.61	44.50	39.72	40.13	38.46	39.30	37.18	42.16	43.34	42.16	35.87	46.02	40.54	41.35
Mean	39.12	37.57	41.10	40.21	40.65	41.83	41.74	41.98	43.77	42.19	41.13	40.92	42.17	42.84	41.88	41.74	41.68	40.57
Std dev	0.0267	0.0266	0.0282	0.0266	0.0262	0.0290	0.0297	0.0252	0.0311	0.0350	0.0285	0.0268	0.0308	0.0297	0.0360	0.0365	0.0261	0.0313

TABLE B-17: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM WITH DYNAMIC BLOCK THRESHOLDS AND THE TAL_BY_8 TEMPLATE IN %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	37.61	39.72	44.50	37.18	43.34	39.30	40.54	36.31	38.46	39.30	38.04	38.04	42.16	34.54	40.13	39.30	36.31	37.61
Run 2	40.54	38.88	42.95	36.75	39.30	37.18	37.18	41.35	38.04	42.55	46.39	42.16	37.18	43.73	45.64	42.16	40.13	43.34
Run 3	38.46	43.34	34.09	38.04	43.34	34.99	39.72	43.73	44.50	33.18	42.16	38.88	40.54	39.30	40.54	41.76	42.16	43.73
Run 4	37.61	44.89	37.18	45.64	40.54	38.46	36.75	35.43	37.18	37.61	38.46	40.13	39.72	40.54	39.30	38.46	40.54	38.04
Run 5	33.18	37.61	40.95	42.16	43.73	40.54	40.95	34.54	41.76	40.54	39.72	40.54	38.46	31.32	36.75	39.30	34.09	38.46
Run 6	38.04	38.46	36.31	37.18	43.34	43.34	37.61	40.13	37.61	40.54	42.55	38.88	44.12	46.76	45.27	38.88	37.18	45.27
Run 7	35.87	40.13	38.46	40.54	43.34	36.75	40.95	42.16	35.43	40.95	39.72	42.55	36.31	42.55	40.54	40.13	43.34	35.87
Run 8	45.64	45.27	40.13	42.55	36.31	37.18	39.72	42.95	44.12	38.04	38.88	40.13	42.55	44.12	37.18	39.30	37.61	37.61
Run 9	41.35	37.61	38.04	34.99	38.88	41.35	45.64	35.87	39.30	42.16	36.75	40.95	38.88	39.72	34.99	41.76	40.13	42.95
Run 10	38.88	38.88	39.30	38.46	37.61	42.55	43.73	35.87	38.04	38.04	39.72	38.88	41.76	39.72	38.04	39.30	43.73	42.16
Run 11	40.54	45.27	42.16	40.95	39.72	41.35	42.95	30.85	40.54	35.87	43.34	37.61	35.43	44.89	36.75	40.13	32.72	40.54
Run 12	36.75	38.04	39.72	35.87	42.16	41.35	42.55	40.13	38.46	42.16	41.35	37.61	38.46	42.55	41.76	40.54	34.54	32.26
Run 13	42.16	42.95	41.35	41.76	37.61	42.95	37.18	38.46	39.30	31.32	35.43	41.35	36.75	38.04	38.04	42.16	38.88	37.61
Run 14	35.87	44.12	42.55	40.54	38.88	40.13	36.31	41.35	36.75	40.13	39.72	37.61	42.95	41.76	42.95	38.88	43.73	38.04
Run 15	42.16	40.54	39.30	37.18	40.13	35.87	37.18	40.54	38.46	34.99	39.72	41.35	37.61	43.34	34.99	43.34	38.04	43.34
Run 16	41.76	35.87	38.04	40.54	33.18	49.30	39.30	35.87	38.46	46.39	38.04	38.04	40.54	41.35	39.72	37.18	36.75	39.30
Run 17	46.39	40.54	44.50	38.88	35.87	39.30	39.30	37.61	38.46	37.18	44.50	38.46	42.55	42.95	36.75	39.30	39.30	36.31
Run 18	41.35	37.18	43.34	43.34	31.32	39.72	44.12	40.13	42.16	44.50	44.50	38.88	38.88	40.95	38.88	39.30	34.09	47.86
Run 19	32.72	44.50	36.31	40.13	43.73	33.18	43.34	40.54	40.13	29.89	38.46	39.72	41.76	38.46	39.30	41.76	40.13	43.73
Run 20	35.87	40.54	41.76	38.46	40.13	38.88	38.46	37.61	34.54	43.34	46.76	37.18	41.76	43.34	42.55	44.50	42.55	37.18
Run 21	41.35	40.13	40.95	37.18	41.35	29.41	44.50	39.72	35.43	36.31	45.27	40.95	41.35	44.50	42.55	32.26	45.64	40.13
Run 22	42.55	46.39	44.12	36.31	39.72	35.87	41.35	35.43	37.61	38.46	38.88	44.89	45.64	38.88	45.64	37.61	40.54	40.95
Run 23	41.35	43.73	37.61	42.55	40.95	34.54	42.16	38.04	44.50	42.95	37.18	37.61	38.88	40.95	41.35	36.75	37.61	39.30
Run 24	48.22	34.99	40.13	40.95	34.99	41.35	38.46	41.76	44.89	40.13	39.72	38.88	42.55	42.55	40.13	42.55	40.95	44.89
Run 25	38.46	32.26	34.54	39.72	38.04	42.95	40.95	36.75	36.75	37.18	35.87	41.35	44.50	44.50	35.43	38.04	35.87	39.30
Run 26	39.30	40.13	39.72	43.34	39.72	40.13	44.12	35.43	39.30	39.30	46.39	41.76	39.30	38.04	40.54	38.88	38.04	41.35
Run 27	42.16	43.34	39.30	41.35	35.87	39.30	40.54	38.88	37.61	38.04	40.95	46.02	43.73	45.27	36.31	40.54	40.95	36.75
Run 28	38.88	36.31	39.72	35.87	43.34	38.04	32.72	39.72	38.46	46.02	40.54	43.73	39.72	37.18	40.95	34.54	42.16	43.34
Run 29	40.13	40.13	45.27	40.54	39.30	37.18	39.30	38.04	38.88	34.09	33.64	38.46	40.95	46.39	38.88	40.95	37.18	44.89
Run 30	35.43	39.72	42.95	34.09	39.72	36.75	38.46	40.13	37.18	38.88	45.27	38.04	38.46	42.55	37.61	43.34	40.54	38.04
Mean	39.69	40.38	40.17	39.44	39.52	38.97	40.20	38.51	39.08	39.00	40.60	40.02	40.45	41.36	39.65	39.76	39.18	40.34
Std dev	0.0353	0.0338	0.0288	0.0277	0.0315	0.0365	0.0289	0.0284	0.0267	0.0390	0.0342	0.0223	0.0253	0.0338	0.0293	0.0255	0.0316	0.0343

TABLE B-18: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM WITH DYNAMIC BLOCK THRESHOLDS AND THE TAL_BY_9 TEMPLATE IN %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	34.99	41.35	40.13	44.12	45.27	44.12	40.95	42.55	42.55	42.95	40.95	40.54	42.16	46.76	40.54	41.35	39.30	48.58
Run 2	42.95	40.54	37.18	35.43	43.73	44.89	40.95	38.04	42.16	44.89	38.04	40.54	41.35	38.46	38.88	41.76	39.30	40.95
Run 3	40.95	36.31	43.34	40.95	43.34	40.54	43.34	44.12	38.46	40.95	38.04	41.35	43.73	36.75	42.55	38.04	40.95	40.95
Run 4	40.13	36.31	46.39	42.55	41.76	38.88	38.46	35.87	47.13	40.95	36.31	42.95	44.12	40.13	35.87	39.30	35.87	43.34
Run 5	35.43	38.46	42.16	46.39	42.16	43.34	45.64	37.61	42.16	41.35	39.30	40.95	42.55	38.46	41.35	43.34	38.04	41.76
Run 6	37.61	42.16	38.88	41.76	42.55	42.55	38.88	38.04	43.34	39.30	38.88	42.16	38.04	45.64	38.88	42.95	39.30	40.95
Run 7	34.54	42.95	43.73	44.12	40.54	41.35	42.55	38.88	39.30	38.46	40.13	37.18	40.54	36.75	39.72	40.95	40.54	42.55
Run 8	34.99	44.12	42.16	40.54	40.95	38.88	43.34	38.46	40.13	39.72	39.72	45.27	47.13	39.72	40.13	36.75	41.35	44.12
Run 9	37.61	35.87	38.46	44.89	38.04	44.50	44.50	40.95	44.89	42.55	35.87	42.95	42.55	35.43	42.16	39.72	38.88	39.72
Run 10	38.04	43.73	42.55	44.50	44.89	46.39	46.76	42.95	42.55	45.27	43.73	44.12	40.95	42.55	35.87	38.46	39.72	44.89
Run 11	36.31	39.72	43.34	41.35	43.73	34.54	41.35	38.46	45.27	42.95	40.13	37.61	45.64	36.31	46.02	41.76	39.72	41.76
Run 12	38.88	39.72	34.54	44.12	43.34	38.04	41.76	36.75	37.18	43.34	47.86	39.72	34.54	40.13	47.50	43.34	46.02	40.54
Run 13	42.95	41.35	34.54	43.73	40.13	39.30	38.04	38.46	40.95	40.95	43.34	39.72	46.02	42.55	38.46	41.35	44.12	46.39
Run 14	38.46	42.55	35.87	44.12	40.13	40.13	48.94	40.54	34.99	41.76	42.95	39.72	44.50	40.54	36.75	34.09	41.76	39.30
Run 15	34.99	44.12	37.18	39.72	43.34	40.54	45.64	44.12	41.76	39.30	44.89	42.16	38.04	42.55	42.16	37.18	39.30	37.61
Run 16	37.61	42.16	42.95	40.13	38.46	39.30	43.73	40.54	42.16	43.34	45.27	38.46	43.73	40.95	41.76	37.18	38.04	45.27
Run 17	34.09	37.18	46.39	40.95	45.64	45.64	40.13	35.87	41.76	42.16	44.12	34.99	42.16	40.13	42.16	38.04	40.95	39.72
Run 18	36.75	42.55	41.76	44.89	42.55	42.16	44.12	40.95	40.95	39.72	46.39	41.35	41.35	43.34	44.50	37.18	38.88	45.27
Run 19	39.72	38.88	39.72	48.22	38.88	44.12	37.18	41.35	41.76	40.13	38.88	42.55	42.55	40.54	38.04	43.34	44.50	46.02
Run 20	31.32	39.30	35.87	47.13	37.18	45.64	43.73	43.34	35.87	38.88	33.64	40.95	40.95	37.18	40.54	37.18	43.34	42.16
Run 21	28.93	38.04	42.16	41.35	34.09	43.34	43.73	36.75	38.04	41.35	41.35	41.35	46.02	39.30	37.61	38.88	38.46	40.54
Run 22	34.09	41.35	35.87	40.95	47.50	38.46	41.35	36.31	42.95	44.50	42.16	36.31	41.35	36.75	43.73	34.54	42.55	40.95
Run 23	39.30	47.86	42.95	42.16	40.54	39.72	41.76	44.89	36.31	33.64	38.88	40.13	46.02	41.76	41.35	45.27	42.55	39.30
Run 24	37.61	45.27	40.54	42.55	44.50	35.43	42.95	47.50	45.64	40.54	38.46	43.34	41.35	40.95	43.73	38.46	42.55	47.13
Run 25	36.31	39.30	45.27	44.12	42.16	43.73	38.04	38.04	38.46	41.76	44.89	40.54	40.54	41.76	44.12	40.54	34.99	43.34
Run 26	33.18	44.50	37.61	37.61	43.73	46.76	45.27	40.54	41.76	39.72	41.35	40.13	40.54	35.43	40.95	42.16	36.31	42.55
Run 27	36.31	40.95	48.58	42.95	48.22	40.13	39.72	44.50	38.46	39.72	42.16	40.54	41.76	42.95	41.35	41.35	45.64	46.39
Run 28	38.46	46.02	41.76	39.30	46.02	41.76	39.72	40.13	41.35	43.73	44.50	40.54	40.13	44.50	40.13	40.13	44.50	40.13
Run 29	36.31	40.13	40.13	39.72	43.73	40.95	41.35	36.31	43.34	41.76	42.16	40.54	44.12	42.16	36.31	40.13	40.54	40.13
Run 30	38.46	39.72	39.72	40.13	44.12	42.55	43.34	40.54	42.95	44.12	40.13	44.89	35.87	41.76	44.50	40.54	40.95	42.55
Mean	36.91	41.08	40.72	42.35	42.37	41.59	42.24	40.11	41.15	41.32	41.15	40.78	42.01	40.41	40.92	39.84	40.63	42.49
Std dev	0.0301	0.0293	0.0358	0.0273	0.0304	0.0302	0.0275	0.0302	0.0290	0.0233	0.0323	0.0230	0.0288	0.0288	0.0292	0.0263	0.0273	0.0268

Table B-19: Detailed results for the 30 runs of the cross-validation experiments using the video processing system with dynamic block thresholds and the TAL_by_10 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	34.54	38.88	39.30	43.34	37.61	36.31	38.88	43.73	44.89	40.13	42.16	40.54	38.04	42.16	44.89	40.54	46.02	45.27
Run 2	34.99	28.44	38.46	42.16	41.76	40.54	40.54	46.02	42.95	40.13	34.99	39.30	41.76	42.16	40.95	42.95	45.27	38.04
Run 3	36.31	40.95	36.75	43.73	36.31	39.30	39.72	40.54	40.54	46.02	36.75	39.72	43.34	40.95	42.16	40.54	38.04	45.64
Run 4	40.54	41.76	33.64	41.76	44.12	41.76	43.34	36.31	40.13	42.16	47.13	46.76	45.27	45.27	44.50	34.99	38.04	44.12
Run 5	36.75	40.13	36.31	40.95	47.13	42.95	38.88	38.04	40.13	40.13	46.02	42.16	38.88	43.73	48.22	37.61	44.89	42.95
Run 6	36.31	37.18	40.54	40.95	41.35	46.02	44.89	44.89	42.16	42.95	34.54	39.30	42.95	43.34	38.88	38.88	50.00	48.58
Run 7	40.13	37.61	40.54	48.58	40.13	43.73	45.64	39.30	38.88	42.95	44.12	44.50	44.12	42.55	45.64	44.50	38.04	39.30
Run 8	30.85	44.12	37.61	39.30	44.89	44.12	44.12	43.34	36.75	40.54	40.54	42.55	42.16	40.54	43.73	42.16	36.75	42.55
Run 9	40.54	39.30	39.30	44.12	42.16	47.86	39.30	44.89	44.50	38.46	37.18	41.76	41.76	41.76	44.50	34.54	42.55	41.76
Run 10	36.31	35.43	40.54	40.95	39.30	46.76	40.13	45.27	40.54	41.35	35.43	40.54	41.76	43.34	38.46	45.64	41.35	43.73
Run 11	38.04	42.16	44.50	40.95	41.76	43.73	43.73	39.72	39.72	47.13	43.73	42.55	41.76	41.35	42.16	42.95	44.12	42.16
Run 12	36.31	36.75	37.61	35.43	39.72	46.39	41.76	44.50	35.87	38.46	45.27	45.27	46.76	44.50	43.73	45.27	43.73	43.34
Run 13	38.46	40.54	43.34	42.55	40.13	42.55	44.89	34.54	39.30	40.54	38.88	42.16	46.76	36.75	45.27	44.50	40.95	40.95
Run 14	35.43	34.54	37.18	47.50	40.95	39.30	40.54	37.18	38.88	40.13	44.12	47.50	42.16	42.16	37.18	43.73	37.18	48.22
Run 15	37.18	33.18	45.27	36.31	46.02	45.27	41.76	42.55	40.95	45.27	42.55	44.50	40.54	39.30	42.16	41.35	38.04	43.34
Run 16	37.18	36.75	37.18	42.55	40.13	40.95	39.72	43.73	40.95	37.61	42.95	44.50	44.50	41.35	40.95	43.34	43.34	43.34
Run 17	38.04	39.72	38.04	43.73	38.88	38.04	44.50	44.89	45.64	39.30	42.95	35.43	42.55	45.64	42.95	42.55	42.16	36.31
Run 18	35.87	36.31	38.88	44.12	43.73	38.46	46.02	40.95	34.99	44.89	40.54	38.88	42.55	38.04	41.76	39.72	46.39	45.27
Run 19	36.31	33.64	40.13	40.13	40.13	37.18	43.34	38.46	49.65	42.55	44.50	39.72	46.76	42.55	44.50	44.50	39.72	50.35
Run 20	35.43	38.88	38.46	42.16	43.34	37.18	44.12	49.65	45.64	44.50	38.04	38.46	44.12	39.72	44.50	44.50	42.95	41.76
Run 21	41.76	34.54	35.87	40.95	42.16	42.95	39.30	44.12	42.55	40.95	41.76	42.95	41.35	46.76	44.89	44.12	39.30	43.73
Run 22	35.43	42.55	36.75	39.72	41.76	47.86	44.12	42.55	44.89	42.55	44.89	37.61	38.04	41.76	47.13	39.30	42.16	43.34
Run 23	44.12	34.99	41.76	43.34	45.27	42.55	46.76	42.55	41.35	40.95	44.89	49.30	44.12	42.16	44.50	42.95	44.50	44.12
Run 24	39.72	33.18	42.16	42.95	36.31	36.31	40.13	40.13	43.34	40.13	38.88	40.13	43.34	46.02	41.35	41.76	41.76	39.30
Run 25	32.26	38.88	40.54	38.46	40.54	45.27	43.73	40.95	40.13	43.34	40.95	33.64	33.64	43.34	41.76	41.35	41.76	40.95
Run 26	34.54	37.18	40.95	40.54	40.95	38.46	41.76	44.50	45.64	41.76	38.88	39.30	37.18	47.86	41.35	41.35	41.76	40.95
Run 27	42.55	33.64	40.95	46.76	43.73	36.75	34.99	44.50	41.76	44.89	41.35	42.55	36.75	46.76	43.34	36.31	45.64	44.89
Run 28	35.87	37.18	40.54	39.72	42.55	41.35	40.95	44.12	44.50	37.18	35.43	43.34	45.27	42.95	44.89	42.16	40.13	47.86
Run 29	34.99	37.18	41.35	42.95	38.46	43.34	36.31	41.35	44.89	38.46	38.46	40.54	42.16	44.50	38.04	43.73	43.34	37.61
Run 30	41.35	40.13	38.46	43.73	41.35	40.95	41.35	48.22	45.64	39.30	43.34	42.16	42.95	44.89	40.13	45.27	43.34	45.64
Mean	37.27	37.52	39.43	42.01	41.42	41.81	41.84	42.38	41.93	41.49	41.04	41.59	42.11	42.80	42.82	41.77	42.11	43.18
Std dev	0.0295	0.0336	0.0255	0.0281	0.0260	0.0347	0.0281	0.0337	0.0324	0.0253	0.0351	0.0332	0.0305	0.0252	0.0259	0.0293	0.0307	0.0321

TABLE B-20: DETAILED SPIKE RATES FOR THE SECOND SET OF EXPERIMENTS PERFORMED ON THE VIDEO DATA.

Template	Peripheral spike rate	Foveal spike rate	Block levels	Accuracy
MNI_by_2	0.2488	0.3788	5	--
MNI_by_3	0.3441	0.5293	4	45.13%
MNI_by_4	0.5056	0.6204	4	45.07%
MNI_by_5	0.5889	0.6487	3	41.14%
TAL_by_3	0.1983	0.2968	5	--
TAL_by_4	0.2878	0.3976	5	--
TAL_by_5	0.3221	0.4842	4	45.67%
TAL_by_6	0.3805	0.5466	4	44.16%
TAL_by_7	0.4746	0.5963	4	45.18%
TAL_by_8	0.4680	0.6587	3	43.60%
TAL_by_9	0.5473	0.6693	3	42.69%
TAL_by_10	0.5803	0.6837	3	42.32%
Average	**0.4122**	**0.5425**	**--**	**43.88%**

Table B-21: Detailed results for the 30 runs of the cross-validation experiments using the video processing system with a peripheral pixel threshold of 25, a foveal pixel threshold of 5, block thresholds of 0, and the MNI_by_3 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	34.09	40.54	45.64	47.13	41.35	49.65	44.89	49.65	45.64	50.00	43.34	42.95	46.39	46.76	42.16	50.70	46.02	40.95
Run 2	36.75	46.02	41.76	46.02	38.88	49.65	49.65	45.27	44.89	48.94	41.76	47.86	46.39	49.30	40.95	48.94	46.39	44.89
Run 3	43.34	41.76	45.64	49.30	48.22	45.27	48.94	47.13	44.89	41.35	48.94	43.73	41.35	45.64	48.22	46.39	42.16	49.65
Run 4	39.30	44.12	51.04	47.13	46.76	40.13	46.76	50.00	45.64	46.02	41.35	45.27	45.64	44.89	50.00	47.13	44.89	50.35
Run 5	40.54	40.95	47.13	45.27	42.55	44.89	44.12	46.39	47.13	45.64	46.39	49.30	48.94	44.50	43.34	48.22	47.50	45.27
Run 6	42.16	36.31	44.89	44.50	44.50	51.04	47.13	46.39	40.54	50.00	44.89	44.50	43.34	42.55	48.58	46.39	44.89	39.72
Run 7	38.04	43.34	42.55	46.02	46.76	46.76	46.02	42.95	51.04	47.50	48.22	47.50	43.73	46.76	44.89	45.64	50.35	45.27
Run 8	31.79	39.72	45.27	46.39	47.13	42.55	44.12	45.27	45.64	42.16	41.76	40.54	51.04	45.27	44.12	42.95	42.55	42.16
Run 9	42.55	44.89	41.35	45.27	44.50	45.64	44.50	44.89	46.76	46.02	44.50	43.73	47.13	47.13	51.38	46.39	44.89	46.39
Run 10	34.54	40.54	41.76	44.50	45.27	50.70	49.65	47.13	45.27	47.50	47.50	47.50	44.12	49.30	47.50	48.58	39.72	47.13
Run 11	34.09	40.54	44.12	45.27	47.86	40.95	47.50	46.39	42.95	46.39	44.89	48.58	43.34	45.27	41.76	47.86	50.35	44.89
Run 12	37.61	42.16	44.12	47.86	46.02	45.27	50.35	42.55	48.22	44.50	42.95	42.16	49.65	47.86	46.76	49.30	43.34	48.22
Run 13	44.12	36.75	45.27	47.13	51.72	48.94	45.64	48.22	48.22	47.50	40.13	48.22	47.50	42.55	38.88	47.50	47.50	46.02
Run 14	46.76	44.12	47.13	43.73	39.72	44.89	42.55	48.22	45.27	43.73	42.55	45.64	46.39	43.34	43.73	46.39	47.13	46.76
Run 15	42.55	45.64	42.16	46.39	46.02	47.13	48.94	45.64	48.22	46.76	46.02	45.27	52.06	53.40	47.13	46.39	43.34	40.95
Run 16	40.95	43.73	47.13	44.89	45.27	42.95	45.64	48.94	50.00	43.73	40.13	44.89	38.46	43.73	44.12	48.58	47.50	45.27
Run 17	41.35	38.04	46.76	45.64	47.13	44.89	44.12	44.50	44.12	45.27	45.64	48.22	45.64	44.89	43.73	44.12	45.64	43.73
Run 18	37.18	38.46	45.64	48.22	44.89	44.12	50.00	50.00	42.95	52.40	44.89	49.65	42.55	51.72	50.70	47.50	46.76	46.02
Run 19	37.61	39.72	45.27	48.22	40.95	48.22	40.54	50.35	48.58	50.35	39.30	42.16	48.94	46.02	45.27	42.95	46.02	42.55
Run 20	35.43	46.02	46.39	37.18	47.86	49.65	46.76	44.89	41.35	47.50	42.55	48.22	42.55	48.94	41.76	44.12	44.89	44.12
Run 21	37.18	36.31	47.50	46.39	48.22	47.50	39.72	45.64	47.50	49.65	45.64	43.34	47.86	43.73	44.89	44.89	45.27	43.73
Run 22	39.72	40.95	48.58	42.55	48.22	46.39	46.39	42.16	46.39	46.76	46.76	42.16	49.30	49.65	50.70	43.34	48.94	50.35
Run 23	37.18	44.12	43.34	39.30	50.00	47.50	45.27	47.50	49.65	43.34	44.50	42.95	47.86	45.64	44.50	44.50	45.64	43.73
Run 24	39.72	40.54	45.64	41.76	44.89	48.94	51.04	44.89	46.02	47.50	43.34	46.76	44.50	41.76	46.76	50.00	47.86	44.89
Run 25	41.35	37.61	43.73	46.76	44.89	40.13	42.95	45.27	46.76	47.86	44.12	43.73	44.50	45.64	46.39	50.70	43.73	47.50
Run 26	43.73	41.35	44.12	45.27	46.02	47.13	46.39	44.89	45.64	47.13	49.30	44.89	50.35	46.76	48.94	48.22	45.64	46.39
Run 27	38.46	40.13	43.34	46.76	42.16	42.55	42.16	44.89	48.94	43.34	46.39	45.64	44.50	48.58	38.04	44.50	46.39	42.95
Run 28	43.73	40.54	46.76	43.73	42.95	48.58	47.86	47.50	44.50	46.39	48.94	42.16	44.12	46.02	48.58	44.12	47.86	43.34
Run 29	40.95	38.88	44.12	44.12	43.73	40.54	47.13	45.27	40.54	44.89	40.13	48.94	43.34	49.30	44.50	40.54	44.50	46.76
Run 30	38.46	42.55	44.89	47.13	46.02	42.95	46.02	45.64	44.89	43.34	48.58	50.35	47.13	42.55	50.35	45.64	42.95	47.13
Mean	39.37	41.21	45.10	45.33	45.35	45.85	46.09	46.28	45.94	46.45	44.51	45.56	45.95	46.31	45.62	46.42	45.69	45.23
Std dev	0.0345	0.0277	0.0214	0.0253	0.0288	0.0315	0.0279	0.0212	0.0259	0.0256	0.0284	0.0266	0.0304	0.0277	0.0345	0.0245	0.0233	0.0260

TABLE B-22: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM WITH A PERIPHERAL PIXEL THRESHOLD OF 25, A FOVEAL PIXEL THRESHOLD OF 5, BLOCK THRESHOLDS OF 0, AND THE MNI_BY_4 TEMPLATE IN %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	45.27	39.72	42.16	46.76	46.39	42.95	44.50	40.54	45.27	48.22	45.64	49.65	47.86	40.13	41.35	47.86	41.35	44.89
Run 2	42.55	43.34	45.27	42.95	41.35	46.02	46.02	42.95	46.76	42.16	48.22	51.38	44.89	48.58	46.39	47.13	46.76	45.27
Run 3	44.89	47.86	46.02	44.12	42.55	47.86	46.02	47.86	50.35	47.13	41.35	45.27	44.12	45.64	42.95	46.76	41.35	46.02
Run 4	40.54	49.30	46.76	46.02	43.73	45.64	44.89	39.72	46.76	46.39	47.13	40.95	46.76	44.89	49.30	42.16	37.18	47.50
Run 5	42.16	49.65	47.13	51.04	46.02	45.64	43.73	50.70	48.22	46.02	50.70	46.76	46.39	45.27	42.55	42.55	46.76	44.89
Run 6	44.89	46.39	40.54	46.76	48.22	47.13	46.02	44.12	48.22	45.27	45.64	43.73	45.27	40.13	45.64	48.22	45.27	49.30
Run 7	45.64	45.27	47.13	44.12	43.34	51.04	44.89	47.50	43.34	47.86	49.65	41.76	42.55	46.02	42.95	47.50	46.02	40.95
Run 8	44.50	41.76	46.02	42.95	47.50	45.64	46.02	44.50	46.76	40.54	44.12	43.34	43.73	40.54	46.39	41.76	41.76	42.55
Run 9	40.54	43.34	47.50	46.39	42.95	44.89	47.13	47.86	46.39	44.89	46.39	43.73	40.54	51.38	44.50	43.73	47.50	46.39
Run 10	46.02	38.88	42.16	48.58	50.00	42.55	42.16	48.22	44.12	41.76	47.50	46.76	45.64	49.30	47.13	46.39	47.50	46.76
Run 11	40.13	48.94	43.73	45.27	45.27	47.13	47.50	49.30	47.86	48.22	45.64	46.02	45.64	46.39	47.50	43.73	42.16	48.58
Run 12	42.16	46.76	41.35	48.22	41.76	45.64	49.65	49.30	43.34	45.64	48.22	41.35	43.34	45.27	45.27	48.22	42.95	42.55
Run 13	39.72	42.95	47.50	41.76	43.34	45.64	48.94	47.50	44.50	50.00	46.76	38.46	47.86	44.50	46.39	44.12	42.16	48.94
Run 14	42.16	44.12	45.27	44.89	52.73	47.13	42.95	40.13	48.58	43.34	45.27	42.16	44.89	42.95	44.12	43.34	40.54	42.95
Run 15	40.54	42.55	44.50	46.02	40.54	44.89	46.76	44.50	47.86	44.50	51.38	45.27	42.95	43.34	44.89	48.22	46.02	46.39
Run 16	45.64	38.88	44.89	41.35	44.12	45.64	46.39	50.35	50.00	48.22	44.50	42.95	47.13	45.27	47.13	44.89	40.54	46.39
Run 17	43.34	42.16	42.95	48.58	44.89	44.50	46.02	49.30	48.22	42.95	49.65	44.89	46.02	45.27	46.76	40.95	49.30	43.73
Run 18	43.34	42.95	45.64	42.16	44.50	44.89	48.58	44.89	44.50	46.76	48.58	45.64	42.55	46.02	50.00	48.58	48.58	42.95
Run 19	39.30	48.22	45.64	46.39	40.95	46.39	42.55	41.35	45.64	46.76	46.02	44.12	38.88	42.95	36.75	42.95	45.64	41.76
Run 20	47.86	40.13	40.54	49.65	43.73	48.94	42.55	44.50	45.64	48.22	51.38	43.73	38.46	46.39	46.76	47.86	45.27	42.16
Run 21	39.72	45.64	42.16	41.76	47.50	43.34	44.50	47.50	42.16	39.72	47.50	46.39	47.50	45.27	45.64	43.73	38.04	45.27
Run 22	40.13	46.02	43.34	44.50	47.13	43.73	42.16	46.76	44.89	38.04	45.27	46.76	45.64	44.89	45.27	41.76	49.65	45.64
Run 23	45.64	47.13	41.76	49.65	46.02	43.73	45.27	42.95	44.89	50.00	48.58	45.64	47.50	46.02	46.02	44.12	39.72	48.58
Run 24	36.31	46.39	42.16	42.95	48.58	47.50	45.64	44.12	47.50	42.95	46.39	42.16	44.89	43.34	44.12	45.64	44.50	47.13
Run 25	43.73	43.73	40.54	43.73	42.95	50.00	43.34	44.50	46.02	49.65	42.95	44.50	41.76	44.50	45.27	44.89	47.86	44.50
Run 26	45.27	42.16	46.39	46.76	47.50	47.13	42.16	50.35	45.64	45.27	41.76	42.95	45.27	45.27	43.34	42.95	39.30	42.95
Run 27	44.12	48.22	44.12	43.73	49.65	39.72	44.89	43.73	46.39	44.50	46.76	47.13	46.02	43.34	44.12	46.39	43.34	44.89
Run 28	43.73	48.22	44.12	48.58	44.12	51.72	52.06	44.89	51.38	39.72	44.12	50.35	42.95	39.30	42.95	46.39	46.76	41.76
Run 29	37.61	44.89	47.13	42.55	44.50	39.72	45.64	48.22	50.00	41.76	44.89	50.00	42.95	43.34	45.64	46.76	46.76	40.54
Run 30	40.13	52.06	46.39	41.76	42.16	51.04	47.50	39.72	46.39	45.27	47.13	43.73	42.95	47.50	47.86	42.95	47.86	47.13
Mean	42.59	44.92	44.36	45.33	45.13	45.93	45.55	45.59	46.59	45.06	46.64	44.92	44.43	44.77	45.16	45.08	44.28	44.98
Std dev	0.0274	0.0333	0.0224	0.0269	0.0289	0.0285	0.0234	0.0326	0.0218	0.0317	0.0250	0.0289	0.0243	0.0261	0.0249	0.0229	0.0342	0.0243

Table B-23: Detailed results for the 30 runs of the cross-validation experiments using the video processing system with a peripheral pixel threshold of 25, a foveal pixel threshold of 5, block thresholds of 0, and the MNI_by_5 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	33.64	42.16	46.02	38.88	38.46	45.64	40.95	40.54	39.30	45.27	44.12	43.34	40.95	40.54	37.18	39.72	42.16	37.61
Run 2	42.95	40.13	40.95	40.95	41.76	42.95	38.46	44.12	43.73	38.46	39.72	37.61	35.87	44.89	39.72	38.46	42.95	42.16
Run 3	40.54	37.61	41.35	38.46	40.13	44.12	40.95	42.55	37.18	39.72	38.04	44.50	41.76	46.02	43.73	38.04	40.95	44.89
Run 4	42.16	43.34	38.46	44.50	44.12	43.34	38.46	40.54	44.50	42.55	39.30	42.95	42.16	42.55	41.35	40.54	44.12	38.88
Run 5	38.88	40.13	35.43	42.55	39.72	45.27	36.75	39.72	39.30	40.13	39.72	40.13	47.50	42.55	42.55	42.16	43.34	38.04
Run 6	38.04	41.76	43.34	43.73	35.87	43.73	38.04	38.46	40.54	41.76	39.72	38.46	40.13	39.72	42.16	43.73	42.95	47.13
Run 7	40.54	34.54	37.18	42.55	44.89	40.13	42.95	44.12	38.88	42.55	38.88	40.13	46.02	40.95	37.61	40.13	40.95	45.64
Run 8	38.04	38.88	34.09	42.95	44.89	40.13	45.27	42.16	37.61	40.95	42.16	41.76	35.43	42.55	37.61	44.89	41.35	40.13
Run 9	41.76	36.31	41.76	38.46	42.16	39.72	44.50	45.27	44.50	42.95	42.55	43.73	34.99	42.55	42.55	40.54	45.64	42.95
Run 10	44.12	42.95	34.09	40.13	40.95	47.50	44.50	49.30	44.89	47.13	36.75	37.18	40.54	38.88	40.95	37.61	41.76	39.72
Run 11	40.54	39.30	43.73	40.95	48.58	43.34	38.04	38.46	42.16	42.55	40.54	40.95	36.75	42.55	41.35	48.58	37.61	38.04
Run 12	38.46	43.73	39.30	46.02	40.95	44.50	42.16	40.13	39.30	41.35	40.54	41.76	37.61	41.35	40.13	43.34	43.34	42.16
Run 13	40.95	41.35	42.16	34.99	40.13	43.73	44.12	39.30	41.35	42.16	41.76	38.88	37.61	49.30	33.18	46.39	42.16	39.30
Run 14	40.13	43.73	40.13	37.61	39.72	38.88	36.31	42.55	44.50	41.35	34.54	36.75	39.30	40.95	36.31	42.55	42.95	39.72
Run 15	40.54	34.54	40.95	41.76	40.13	37.18	41.76	42.95	46.02	42.16	40.13	37.61	40.54	44.89	31.79	44.50	40.54	42.95
Run 16	39.72	39.30	40.95	33.18	43.73	41.35	44.12	44.50	42.95	38.46	39.30	41.76	38.88	36.31	37.18	38.88	44.12	42.16
Run 17	43.34	43.73	41.35	43.34	39.30	40.13	39.72	44.50	40.95	40.13	45.64	39.72	45.64	38.88	43.73	42.95	42.55	46.02
Run 18	39.72	39.72	40.54	41.35	39.30	45.27	37.18	41.35	44.89	40.54	44.12	38.46	42.55	42.55	44.50	42.16	35.43	34.54
Run 19	35.87	38.88	40.13	40.95	35.87	38.88	42.16	39.30	42.16	49.65	44.50	43.73	45.64	34.54	41.35	40.54	41.35	44.89
Run 20	41.76	42.55	39.72	39.72	42.16	37.18	38.88	40.95	40.13	43.34	38.88	40.54	42.16	35.87	43.34	42.95	48.58	39.30
Run 21	43.34	32.72	39.72	43.73	46.02	40.13	39.72	45.64	40.13	44.12	37.18	38.88	42.95	44.12	42.16	42.95	41.76	42.95
Run 22	35.43	39.72	38.88	43.73	42.95	40.95	43.73	40.13	40.54	40.95	47.13	42.16	40.13	38.04	48.58	41.35	38.88	41.76
Run 23	39.30	34.54	44.50	40.95	43.73	42.95	38.88	43.73	38.88	40.13	40.13	37.61	39.30	44.89	39.30	39.72	41.76	44.89
Run 24	44.50	37.61	38.88	37.61	44.12	40.13	43.73	38.88	36.75	42.95	39.72	39.30	46.02	43.73	42.95	41.35	40.95	41.76
Run 25	39.30	39.72	40.95	42.16	37.61	41.35	44.12	40.54	45.27	41.76	46.02	33.64	38.46	44.50	38.04	42.55	46.39	44.50
Run 26	34.54	42.95	36.75	42.55	39.30	41.76	40.54	45.64	39.72	40.95	44.89	38.46	41.35	44.89	43.73	41.35	39.72	40.54
Run 27	37.61	40.13	42.55	43.34	43.34	38.88	42.16	41.76	41.76	40.54	43.34	41.35	43.34	35.87	43.34	43.73	46.76	40.54
Run 28	43.73	40.95	42.95	39.72	43.34	38.88	42.95	42.16	41.35	37.61	42.16	46.02	36.31	40.13	40.54	46.02	43.73	37.61
Run 29	35.43	38.88	36.31	37.18	45.64	44.12	46.76	33.64	42.95	40.95	45.27	36.31	37.61	44.12	40.54	40.95	38.04	46.76
Run 30	40.13	38.88	44.50	41.35	39.30	33.18	42.95	44.89	42.95	39.30	45.27	41.35	40.54	44.50	43.34	41.35	38.46	41.76
Mean	39.83	39.69	40.25	40.85	41.61	41.51	41.36	41.93	41.50	41.75	41.40	40.17	40.60	41.77	40.69	42.00	42.04	41.64
Std dev	0.0284	0.0293	0.0294	0.0283	0.0296	0.0301	0.0277	0.0297	0.0252	0.0245	0.0306	0.0272	0.0333	0.0336	0.0343	0.0248	0.0276	0.0304

Table B-24: Detailed results for the 30 runs of the cross-validation experiments using the video processing system with a peripheral pixel threshold of 25, a foveal pixel threshold of 5, block thresholds of 0, and the TAL_by_5 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	40.54	39.72	44.12	44.89	48.58	52.73	48.22	50.70	46.76	45.27	51.04	47.86	47.50	42.16	48.58	36.31	47.13	42.95
Run 2	46.39	38.46	42.55	47.13	46.39	48.94	47.86	47.50	48.22	49.30	47.13	46.02	43.34	43.73	46.39	48.22	43.73	40.54
Run 3	40.95	38.46	41.76	44.50	49.30	54.05	52.06	47.13	46.76	50.00	46.76	41.76	43.34	42.55	46.39	46.39	44.50	46.76
Run 4	40.54	39.72	46.02	46.02	50.00	47.86	47.86	46.02	47.86	44.12	44.50	44.12	46.02	42.55	46.02	42.95	47.86	41.35
Run 5	40.54	40.95	42.95	40.95	47.50	41.35	48.58	45.27	46.02	48.22	50.35	45.27	42.16	46.39	41.35	46.76	47.50	48.58
Run 6	41.76	42.16	48.22	45.64	46.39	44.12	50.00	52.06	50.00	45.27	49.65	47.50	43.73	47.86	44.89	45.64	45.64	46.02
Run 7	37.18	40.13	40.54	44.89	48.58	52.73	47.50	44.50	44.50	49.65	39.72	44.50	42.95	44.50	43.73	46.02	42.16	47.86
Run 8	45.64	37.18	42.16	49.30	47.50	52.06	46.02	43.73	50.35	49.30	49.65	50.00	42.16	43.73	54.05	45.64	45.64	44.50
Run 9	41.35	45.64	39.30	42.55	46.76	50.70	47.13	51.38	45.64	47.13	45.64	46.02	45.64	50.00	50.70	46.76	45.64	48.94
Run 10	45.27	40.13	42.95	42.55	45.64	49.65	46.76	49.30	47.13	48.58	47.50	47.86	42.16	47.86	45.27	43.34	40.95	41.35
Run 11	41.35	37.18	38.88	42.55	39.72	50.00	52.06	48.94	50.35	51.72	45.27	48.22	45.64	50.70	47.86	47.86	44.50	43.34
Run 12	40.54	40.54	34.54	45.27	50.00	48.94	48.58	49.65	47.50	47.13	46.02	49.30	41.76	46.76	48.58	42.95	46.02	40.54
Run 13	38.04	42.95	38.88	50.00	52.40	49.30	51.72	48.94	49.30	47.50	48.22	40.95	42.95	44.89	46.39	40.95	46.02	41.76
Run 14	44.89	36.31	38.88	46.02	52.06	49.30	51.04	49.30	48.94	46.02	46.02	47.13	42.55	43.73	44.12	49.30	46.76	42.16
Run 15	42.55	41.76	44.12	47.86	51.04	48.94	51.38	48.22	49.65	46.39	46.76	42.55	45.27	43.73	46.39	43.73	46.76	46.02
Run 16	42.55	39.30	38.46	48.22	51.38	46.02	47.50	46.39	49.30	48.58	50.00	45.64	48.94	46.02	40.13	45.27	43.34	46.76
Run 17	39.30	38.88	46.02	46.02	51.04	50.35	50.35	48.94	45.64	42.95	47.13	42.55	47.13	43.73	46.02	45.27	42.16	44.50
Run 18	34.99	41.35	33.18	44.89	48.22	45.64	49.65	43.34	44.50	52.40	48.58	49.65	46.39	42.16	52.06	46.02	45.64	46.02
Run 19	37.61	43.73	40.54	47.13	52.40	45.27	48.58	45.64	47.86	45.27	46.76	48.22	44.50	44.50	45.64	43.73	39.30	49.65
Run 20	40.13	40.54	38.04	48.58	48.94	47.86	45.27	48.94	48.94	53.40	50.00	52.06	45.27	44.89	45.64	48.58	49.30	48.94
Run 21	42.55	40.13	43.34	47.50	46.76	48.22	47.13	49.65	46.76	44.89	46.39	48.94	43.34	48.94	47.13	45.64	47.50	38.46
Run 22	43.34	42.16	41.35	41.76	44.89	48.22	48.22	49.30	49.65	40.95	50.00	45.27	47.50	40.54	42.95	44.12	41.35	49.65
Run 23	45.64	42.16	41.35	46.76	45.64	50.35	46.76	49.65	47.13	42.16	46.76	46.76	46.76	44.89	47.86	44.89	45.27	44.89
Run 24	40.95	40.95	42.95	46.02	49.65	48.94	51.38	47.86	46.02	48.58	43.73	45.64	40.13	47.13	45.64	40.95	43.73	49.65
Run 25	38.46	38.88	46.02	45.64	45.64	46.76	48.58	49.30	49.30	48.58	49.65	42.95	44.12	47.86	45.64	44.50	43.34	45.64
Run 26	44.12	39.72	47.13	51.04	44.89	54.05	44.50	48.58	46.76	46.02	49.30	46.02	43.34	42.55	44.89	39.30	48.94	48.58
Run 27	40.95	40.13	42.95	46.39	48.94	49.65	46.39	50.00	51.72	42.16	50.00	44.89	44.12	42.95	42.55	44.89	48.58	50.00
Run 28	39.30	42.16	45.64	48.22	46.39	49.65	44.89	49.65	50.70	45.27	49.65	42.16	44.89	43.34	43.73	41.35	42.55	45.27
Run 29	38.46	39.72	37.61	46.02	47.13	44.89	52.06	49.65	44.12	49.30	40.13	47.50	44.50	44.12	47.50	47.86	42.95	45.27
Run 30	40.13	40.95	42.55	46.02	54.38	46.76	50.70	51.04	49.65	44.50	47.13	43.34	50.35	45.27	36.31	44.12	42.55	44.50
Mean	41.20	40.40	41.77	46.01	48.27	48.78	48.63	48.35	47.90	47.02	47.31	46.02	44.61	45.00	45.81	44.64	44.91	45.35
Std dev	0.0271	0.0195	0.0348	0.0234	0.0292	0.0285	0.0220	0.0217	0.0198	0.0300	0.0272	0.0268	0.0225	0.0244	0.0334	0.0281	0.0249	0.0314

TABLE B-25: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM WITH A PERIPHERAL PIXEL THRESHOLD OF 25, A FOVEAL PIXEL THRESHOLD OF 5, BLOCK THRESHOLDS OF 0, AND THE TAL_BY_6 TEMPLATE IN %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	40.13	37.61	40.95	44.12	42.95	45.27	44.50	38.04	42.16	47.86	47.50	48.22	42.55	46.02	46.39	44.50	47.13	41.76
Run 2	39.30	43.73	44.50	39.72	42.16	48.94	43.73	44.12	44.50	45.27	47.13	50.00	44.89	42.55	50.35	43.73	47.13	39.30
Run 3	35.43	39.72	36.31	39.72	46.76	45.64	44.12	46.02	42.16	42.95	40.95	50.00	44.50	49.30	47.50	48.22	50.70	45.27
Run 4	38.04	40.13	40.13	41.35	43.73	42.16	44.50	44.89	40.95	44.89	45.27	50.00	44.12	42.16	47.13	48.94	46.76	45.27
Run 5	38.46	38.46	41.76	44.12	37.61	49.65	45.27	48.58	42.95	44.89	47.50	54.70	50.70	42.16	46.39	46.76	42.95	47.13
Run 6	38.88	34.09	41.35	38.04	46.39	44.50	44.89	42.55	38.88	40.54	53.40	42.16	48.94	45.27	43.73	46.02	47.86	44.12
Run 7	38.04	38.04	38.88	45.27	44.89	46.76	49.30	45.64	44.12	41.76	47.50	49.30	43.34	43.73	46.76	41.35	44.89	45.64
Run 8	42.16	37.18	40.54	47.50	45.27	43.34	41.35	41.35	41.76	42.55	47.13	49.65	50.00	47.86	44.50	48.58	41.35	48.94
Run 9	39.72	41.76	39.72	42.55	45.27	47.86	45.64	44.12	42.55	42.95	46.02	50.00	45.27	46.76	48.94	45.27	44.89	44.12
Run 10	42.16	42.16	38.46	42.55	44.50	42.55	35.87	45.27	43.34	46.39	44.12	47.13	46.76	41.35	46.76	45.27	41.35	47.13
Run 11	39.30	38.46	38.04	44.50	46.02	46.76	47.50	45.64	44.12	41.76	44.50	46.02	41.76	46.02	45.64	46.76	52.73	52.40
Run 12	29.89	34.54	38.04	38.46	50.35	43.34	47.13	38.88	43.34	51.72	46.39	45.64	48.22	43.34	43.73	48.94	43.34	43.34
Run 13	37.61	38.46	36.75	41.35	42.95	50.35	43.34	42.16	45.64	43.73	45.64	48.58	47.50	46.39	45.64	44.89	48.94	47.13
Run 14	35.43	37.18	39.30	44.12	43.34	46.39	42.16	44.50	46.39	43.73	42.55	42.55	46.76	46.02	44.12	46.76	42.95	45.27
Run 15	38.88	39.72	39.72	42.55	46.76	38.04	42.55	50.35	50.35	42.16	46.02	44.12	44.89	37.18	49.30	47.13	41.35	48.22
Run 16	39.72	40.95	43.73	39.30	43.73	44.12	46.02	43.34	44.50	46.76	47.13	51.04	41.76	51.38	44.89	43.34	43.73	44.89
Run 17	37.61	33.18	43.73	45.27	46.02	40.54	43.34	43.34	39.30	44.50	46.02	42.16	48.94	47.13	47.50	46.02	52.40	42.95
Run 18	38.04	39.30	36.31	40.13	40.95	42.95	45.27	38.88	46.02	42.16	45.64	44.89	48.58	46.39	44.89	49.30	41.76	46.76
Run 19	42.16	42.16	38.46	42.16	46.39	39.72	40.13	43.73	46.39	45.27	47.13	47.13	48.22	46.39	50.35	53.07	44.89	41.76
Run 20	40.95	40.54	40.95	41.76	44.89	47.13	44.89	42.16	49.65	43.34	44.89	47.86	44.89	49.30	49.30	46.76	48.22	47.13
Run 21	38.46	39.72	33.64	39.72	44.89	48.22	42.95	44.12	45.27	47.86	41.35	45.27	48.58	44.12	50.35	43.34	50.00	42.55
Run 22	42.55	40.95	32.26	43.34	42.16	44.50	44.89	42.55	38.04	47.50	45.27	47.13	47.13	40.13	51.38	49.30	46.39	43.73
Run 23	34.99	38.88	38.88	41.35	44.50	49.30	47.50	46.39	42.95	44.50	43.34	42.95	50.00	47.86	51.04	51.72	46.02	48.94
Run 24	35.87	37.18	44.89	47.50	47.86	42.55	46.76	42.16	44.12	44.89	50.70	47.13	45.64	48.22	42.16	44.50	44.50	47.50
Run 25	38.46	34.54	40.54	38.88	46.02	44.12	42.16	44.89	39.30	41.76	41.76	44.89	47.50	47.50	52.06	47.86	49.30	44.50
Run 26	39.72	40.95	44.89	42.16	41.76	48.94	44.12	41.76	41.35	47.13	45.64	42.95	47.50	49.65	47.86	49.30	44.89	46.76
Run 27	36.31	35.43	40.13	38.04	47.50	42.16	39.30	45.27	41.35	47.50	48.22	42.95	49.30	44.89	41.35	48.94	46.39	46.02
Run 28	38.04	41.76	42.95	44.89	42.55	47.50	48.22	46.02	47.13	40.54	46.02	44.50	44.50	40.13	46.76	46.76	45.64	42.16
Run 29	41.76	40.54	38.04	42.95	46.76	40.95	44.89	43.34	39.30	48.22	42.16	51.04	47.50	44.50	43.34	47.86	40.95	43.73
Run 30	40.54	38.04	39.30	40.54	44.12	42.16	44.89	44.12	46.39	42.16	46.76	44.12	51.04	48.94	48.22	48.58	45.64	47.86
Mean	38.62	38.84	39.77	42.13	44.63	44.88	44.24	43.81	43.48	44.57	45.79	46.80	46.71	45.42	46.94	46.99	45.83	45.41
Std dev	0.0261	0.0259	0.0299	0.0256	0.0241	0.0315	0.0272	0.0258	0.0299	0.0262	0.0258	0.0315	0.0255	0.0321	0.0277	0.0251	0.0317	0.0265

Table B-26: Detailed results for the 30 runs of the cross-validation experiments using the video processing system with a peripheral pixel threshold of 25, a foveal pixel threshold of 5, block thresholds of 0, and the TAL_by_7 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	38.46	43.34	40.54	44.12	46.39	43.34	42.95	45.27	46.02	43.34	46.39	44.12	41.35	47.50	42.95	48.94	46.76	47.50
Run 2	37.61	42.16	44.89	40.95	47.13	43.73	47.50	44.50	44.12	46.76	43.34	42.55	41.35	51.38	47.50	47.13	52.73	44.50
Run 3	38.04	44.12	46.76	42.55	44.89	46.39	46.02	43.73	47.13	50.70	47.13	49.30	49.30	44.50	49.65	44.89	49.30	48.58
Run 4	38.88	43.34	48.94	44.12	47.86	46.76	41.35	47.86	46.39	45.64	44.50	46.39	46.02	46.76	46.02	46.02	44.12	44.12
Run 5	40.54	38.88	40.54	33.64	50.00	44.89	52.06	41.35	45.64	47.13	42.55	42.55	44.50	51.72	46.02	50.70	45.64	46.39
Run 6	41.35	39.30	44.50	46.76	50.70	46.02	42.55	46.39	48.94	47.86	42.16	46.02	48.94	43.34	46.76	46.02	40.95	43.34
Run 7	40.54	42.16	42.95	43.34	47.50	42.16	49.65	47.50	48.22	48.58	44.89	45.27	46.02	46.02	50.00	47.13	40.95	47.50
Run 8	41.35	39.72	43.73	46.39	46.02	48.58	51.72	47.50	46.76	48.22	45.64	46.39	50.70	47.86	45.64	44.89	49.65	46.02
Run 9	38.04	40.54	38.04	42.55	47.50	48.58	46.39	49.65	41.35	42.95	47.50	44.89	48.58	47.86	50.00	45.27	43.34	42.16
Run 10	38.88	43.34	44.50	45.27	42.55	42.55	44.89	46.39	48.22	45.64	40.54	43.34	47.13	49.30	46.39	46.76	43.73	45.64
Run 11	41.76	39.72	45.64	46.76	46.76	52.06	44.50	46.02	47.86	44.50	43.73	45.27	50.35	45.64	44.12	46.02	41.35	42.16
Run 12	38.04	39.30	40.95	50.00	44.89	41.76	43.73	48.58	38.46	50.35	46.76	46.39	45.27	53.07	47.50	44.12	46.02	46.76
Run 13	43.34	38.88	38.04	41.76	47.50	49.65	46.76	44.50	47.13	46.76	44.12	47.50	44.12	48.22	44.12	43.73	48.22	48.58
Run 14	43.73	38.88	40.95	40.95	42.95	50.70	47.86	38.46	47.50	46.76	47.13	39.72	44.12	46.76	40.54	49.30	46.02	41.76
Run 15	44.12	41.76	43.34	53.07	40.95	42.16	47.13	46.02	47.13	47.86	47.86	47.13	44.89	46.76	47.50	47.13	47.13	47.86
Run 16	41.76	38.46	41.35	46.76	43.73	45.64	42.16	43.34	44.50	43.34	48.94	46.39	43.73	46.39	50.00	47.50	44.89	48.58
Run 17	40.13	38.88	45.27	44.50	42.95	43.73	45.64	46.76	46.76	44.12	44.50	48.94	44.89	49.65	44.12	47.13	43.73	45.64
Run 18	39.72	46.39	37.18	42.55	52.06	44.89	48.22	45.27	50.00	44.89	44.12	47.86	46.39	45.64	46.76	44.12	46.39	50.00
Run 19	38.04	38.04	44.50	45.64	47.86	41.76	44.89	44.50	42.55	47.50	46.02	45.27	46.02	43.73	49.30	46.02	46.39	46.02
Run 20	43.34	41.76	38.88	45.64	48.22	41.35	46.02	48.58	48.22	44.12	40.95	46.39	51.38	49.30	51.38	46.39	46.39	44.89
Run 21	37.18	33.64	42.16	45.27	49.30	48.94	43.34	44.89	42.55	42.95	41.76	44.12	46.02	48.22	45.27	47.13	46.02	47.13
Run 22	36.75	38.46	42.55	47.13	42.95	53.73	44.89	44.50	47.86	48.58	39.30	42.55	41.76	45.64	42.55	52.40	46.02	46.02
Run 23	43.73	42.55	38.04	50.35	48.94	45.64	48.58	51.72	39.72	47.50	48.22	45.64	47.50	47.50	45.64	46.02	48.22	41.35
Run 24	38.46	38.88	42.95	43.73	46.76	46.39	46.02	42.95	45.27	49.30	48.22	50.35	44.50	39.72	45.27	48.94	45.27	47.50
Run 25	33.64	44.12	42.55	40.95	50.35	48.94	43.34	48.22	47.50	51.72	43.34	47.13	51.04	49.65	42.55	41.76	48.22	43.73
Run 26	41.35	42.95	34.99	44.50	42.95	47.86	48.22	44.50	47.13	44.89	46.39	47.50	45.27	49.65	46.02	44.89	46.76	50.70
Run 27	38.04	41.76	43.73	46.76	48.94	51.38	42.55	43.73	46.02	46.02	47.50	41.35	44.12	44.12	46.39	48.94	50.00	45.64
Run 28	41.76	35.43	39.72	42.16	45.27	39.72	50.00	41.35	46.76	46.76	47.50	49.30	46.76	43.73	44.12	46.76	45.27	50.35
Run 29	39.30	40.54	40.13	47.13	45.27	49.65	47.86	52.40	49.65	47.13	44.89	41.76	43.34	46.02	50.35	47.50	47.50	47.86
Run 30	39.30	30.37	43.73	47.13	44.89	47.86	44.12	48.58	47.13	42.95	47.86	45.27	47.86	44.50	44.89	46.76	44.12	48.94
Mean	39.91	40.26	42.07	44.75	46.47	46.23	46.03	45.83	46.08	46.49	45.12	45.55	46.11	47.01	46.31	46.68	46.04	46.24
Std dev	0.0241	0.0322	0.0305	0.0350	0.0272	0.0351	0.0274	0.0295	0.0271	0.0235	0.0252	0.0250	0.0271	0.0278	0.0261	0.0211	0.0261	0.0251

TABLE B-27: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM WITH A PERIPHERAL PIXEL THRESHOLD OF 25, A FOVEAL PIXEL THRESHOLD OF 5, BLOCK THRESHOLDS OF 0, AND THE TAL_BY_8 TEMPLATE IN %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	39.72	44.50	41.76	39.72	43.73	49.65	44.12	45.27	43.34	42.55	39.72	44.50	41.76	44.89	42.55	42.95	42.95	41.35
Run 2	38.04	40.54	46.76	40.13	42.95	47.50	40.95	47.86	44.12	39.30	47.86	45.27	42.55	42.95	44.89	43.73	38.04	43.73
Run 3	36.75	48.94	41.35	36.75	43.73	45.27	43.73	39.30	44.89	48.94	42.16	45.27	47.50	43.34	44.12	46.39	37.61	36.31
Run 4	42.16	38.88	39.72	43.34	41.35	45.27	47.50	43.73	42.95	44.50	44.12	39.30	42.16	43.34	46.39	40.95	43.34	43.73
Run 5	38.04	35.87	36.31	44.12	44.89	44.12	41.76	51.72	44.89	47.13	42.16	44.89	42.16	46.76	44.50	45.64	41.35	46.76
Run 6	38.88	44.89	40.54	41.76	46.02	37.61	44.50	40.13	44.89	47.50	47.86	39.72	42.55	42.95	44.50	47.13	42.16	46.39
Run 7	44.50	40.13	43.73	47.50	41.35	39.72	44.89	44.50	41.35	44.50	43.34	42.55	44.50	47.86	44.12	39.30	39.30	45.27
Run 8	42.95	39.30	41.76	45.64	37.18	42.95	40.95	43.34	47.86	40.95	48.58	46.39	42.95	41.76	41.76	44.89	43.73	43.73
Run 9	44.12	38.04	44.50	44.12	44.89	45.64	48.58	44.89	40.13	49.65	46.39	42.16	39.30	46.39	38.46	41.35	38.46	43.34
Run 10	42.16	45.64	45.64	43.34	42.55	48.22	50.70	37.18	48.22	51.04	41.35	40.95	44.50	46.76	46.39	43.34	47.13	42.95
Run 11	39.72	37.61	45.27	42.16	50.70	40.95	47.50	43.34	50.70	49.65	49.65	46.39	43.73	42.95	45.27	45.27	46.76	43.73
Run 12	40.13	42.95	43.73	43.34	42.16	37.18	43.34	44.12	47.86	46.39	51.72	43.34	42.55	41.35	43.73	38.88	42.95	46.02
Run 13	39.30	39.30	45.64	40.95	49.65	48.22	45.64	43.73	46.39	44.89	46.39	44.89	42.95	45.64	42.55	40.95	40.54	39.72
Run 14	38.04	42.16	32.72	38.46	44.50	44.12	44.50	46.39	46.02	45.27	48.94	43.34	42.16	46.39	46.02	37.61	45.27	41.76
Run 15	38.88	35.43	41.35	48.58	41.35	42.95	45.27	43.34	45.64	40.95	44.89	44.12	42.95	44.50	44.12	47.13	43.34	45.27
Run 16	41.35	40.95	45.27	32.72	41.76	44.12	42.95	48.94	47.13	45.64	46.02	48.22	44.89	43.73	39.30	44.12	46.02	47.13
Run 17	39.72	39.30	37.18	38.46	42.55	44.12	42.16	47.86	42.95	40.54	42.95	42.95	42.55	40.95	46.76	40.54	45.64	38.46
Run 18	37.61	40.95	39.30	38.88	46.39	41.76	48.94	46.76	44.12	47.50	42.55	48.94	42.16	47.50	39.72	41.35	44.50	42.95
Run 19	44.89	40.95	43.34	42.16	42.55	40.95	45.27	43.73	42.95	44.89	46.76	45.27	48.58	46.39	35.43	43.34	42.95	46.02
Run 20	40.13	42.95	48.94	45.64	40.95	42.16	41.35	45.64	44.89	44.12	43.73	44.50	40.95	41.76	46.76	48.94	47.50	43.34
Run 21	38.46	41.76	44.12	42.95	42.55	49.30	49.30	48.22	46.76	46.02	44.50	40.13	44.89	43.34	42.16	45.64	47.13	40.54
Run 22	42.55	41.35	35.87	46.39	43.34	45.27	45.64	43.73	44.50	45.64	42.55	46.76	48.94	49.30	41.35	39.72	39.72	42.16
Run 23	42.95	40.54	46.76	44.12	38.04	46.76	44.12	44.89	41.76	43.34	48.22	47.86	44.89	46.39	41.76	46.02	43.73	40.54
Run 24	42.16	43.73	41.35	45.27	45.27	44.12	42.95	47.50	47.50	39.30	48.58	45.27	44.89	44.50	46.02	48.58	42.95	40.13
Run 25	35.87	41.76	40.54	40.54	40.54	49.65	47.86	51.38	45.27	47.50	47.13	38.46	43.34	42.55	47.86	46.76	45.27	44.50
Run 26	39.72	45.64	40.95	40.95	45.64	42.95	42.55	42.55	42.95	42.95	44.89	41.76	46.02	42.95	41.76	46.76	45.27	43.34
Run 27	38.04	40.54	41.76	39.30	46.02	46.76	45.64	41.35	47.86	45.64	48.58	44.50	46.39	44.89	42.16	47.50	42.55	49.30
Run 28	42.95	43.73	44.50	44.12	54.38	46.02	50.00	39.72	42.95	47.13	42.55	46.02	44.50	40.13	40.95	39.30	44.50	45.64
Run 29	36.31	36.75	45.64	44.89	41.35	44.12	41.35	42.95	42.16	43.73	43.73	47.13	41.35	45.27	41.76	40.95	42.55	44.89
Run 30	40.13	45.64	41.76	42.16	43.73	44.12	47.86	48.94	41.35	42.95	47.13	35.43	43.34	34.09	49.30	39.72	44.12	48.22
Mean	40.21	41.36	42.27	42.28	43.74	44.38	45.06	44.77	44.81	45.00	45.50	43.88	43.73	44.05	43.41	43.49	43.24	43.57
Std dev	0.0242	0.0309	0.0354	0.0331	0.0343	0.0315	0.0279	0.0336	0.0244	0.0297	0.0285	0.0305	0.0215	0.0287	0.0296	0.0319	0.0267	0.0286

Table B-28: Detailed results for the 30 runs of the cross-validation experiments using the video processing system with a peripheral pixel threshold of 25, a foveal pixel threshold of 5, block thresholds of 0, and the TAL_BY_9 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	39.30	39.30	39.30	41.76	44.89	48.22	42.95	40.95	46.02	42.55	44.50	42.55	44.50	37.61	40.54	50.00	38.46	38.46
Run 2	31.79	41.76	34.54	41.76	48.58	41.35	40.95	43.73	47.13	42.16	36.75	47.13	46.76	42.16	40.13	39.72	41.35	38.88
Run 3	38.46	45.64	43.34	40.54	45.27	46.39	46.02	42.55	44.12	46.02	42.16	46.39	40.95	40.95	46.39	41.35	40.13	38.46
Run 4	41.35	45.27	38.88	38.88	38.46	45.64	48.58	38.04	44.12	48.22	47.13	43.73	40.54	38.88	46.76	42.55	45.27	44.89
Run 5	37.61	46.39	41.76	42.16	41.35	45.64	50.70	43.34	42.55	39.72	42.55	45.64	36.31	45.27	45.64	39.30	44.89	38.04
Run 6	39.30	47.86	38.88	42.55	46.02	43.34	42.16	40.54	42.55	46.02	43.73	45.64	42.16	36.75	37.61	46.39	42.95	43.34
Run 7	40.54	42.55	41.76	43.73	42.95	52.06	46.39	40.95	45.27	43.34	42.16	48.22	43.73	44.50	37.18	45.64	47.13	44.89
Run 8	35.87	41.35	38.88	42.95	40.13	40.95	44.89	43.73	51.04	40.13	39.72	42.55	40.54	42.95	43.73	40.54	44.12	45.27
Run 9	38.88	34.09	46.02	43.73	40.54	43.34	47.13	43.73	46.39	45.64	46.02	44.50	41.76	42.16	40.13	42.95	38.88	41.35
Run 10	40.13	41.35	45.27	39.72	42.16	44.89	46.39	44.50	47.13	46.02	41.35	46.76	38.46	36.31	38.88	42.95	41.35	42.95
Run 11	32.72	38.46	44.50	43.34	39.30	47.86	42.95	42.55	39.72	44.89	42.16	43.34	40.54	45.64	40.54	41.35	44.89	43.73
Run 12	40.95	43.34	44.12	42.16	45.27	39.30	44.50	47.13	48.94	46.39	45.27	37.18	38.46	42.95	41.35	44.89	41.35	50.00
Run 13	43.34	43.73	37.18	36.75	42.95	43.34	40.13	40.54	44.50	42.16	42.95	47.50	38.04	49.30	42.16	39.30	38.04	38.46
Run 14	43.73	38.88	46.02	39.30	42.55	46.02	44.50	41.76	38.88	44.50	37.18	52.73	46.02	47.50	39.72	47.86	38.88	39.30
Run 15	38.46	46.39	38.46	40.13	44.12	46.39	42.55	45.27	46.76	44.12	42.16	49.30	40.54	42.95	50.35	44.89	40.95	35.43
Run 16	38.04	40.13	37.18	47.50	40.13	42.55	40.95	44.12	44.89	42.55	45.64	40.95	41.35	41.35	40.13	42.55	40.95	41.35
Run 17	44.12	42.16	41.35	39.72	42.16	40.95	46.39	42.55	39.30	43.34	38.04	45.64	44.89	41.35	45.27	42.55	46.02	43.34
Run 18	42.95	44.12	39.30	44.50	43.34	45.27	43.34	45.64	44.89	46.39	45.27	40.54	46.39	41.35	44.12	45.64	42.55	44.89
Run 19	38.46	44.89	37.18	39.30	40.13	45.27	44.12	46.39	44.89	46.76	46.76	49.30	43.73	38.46	47.50	42.16	45.64	45.64
Run 20	41.35	40.13	38.88	41.35	42.16	47.13	48.58	43.34	47.13	42.95	43.73	43.34	42.16	46.76	44.50	44.12	37.61	43.73
Run 21	40.95	41.35	42.95	40.54	44.89	43.73	49.30	42.55	45.27	42.16	37.61	40.13	42.55	41.76	42.16	40.54	39.72	37.61
Run 22	43.73	41.76	39.30	38.46	44.50	46.02	43.73	42.95	42.95	43.34	49.65	44.50	43.34	40.54	42.16	44.89	42.55	46.02
Run 23	35.43	37.61	40.95	42.16	40.54	42.55	47.50	44.50	44.12	44.89	45.27	40.54	39.72	44.89	38.88	42.55	47.13	38.88
Run 24	40.95	43.73	40.95	40.54	46.02	47.13	34.99	40.54	41.76	45.64	47.13	39.72	44.89	42.55	42.16	41.76	43.34	41.76
Run 25	37.61	45.27	40.13	44.50	44.50	44.50	43.73	44.89	40.95	42.95	44.89	41.35	36.75	44.12	43.73	46.39	34.99	42.16
Run 26	42.95	40.95	45.64	43.34	43.34	42.55	43.34	45.64	39.30	44.50	42.95	38.88	45.64	41.76	42.95	36.31	44.12	44.50
Run 27	32.72	37.61	42.95	39.30	35.87	40.95	42.55	43.34	41.76	41.35	44.50	43.34	47.13	41.76	46.02	40.95	40.95	37.61
Run 28	43.34	40.54	41.35	32.72	41.76	42.16	45.27	42.55	44.50	38.04	36.31	40.13	51.04	42.55	42.95	40.54	45.64	41.35
Run 29	39.72	39.30	42.55	45.64	44.89	43.34	46.76	42.95	47.13	42.55	44.89	42.16	38.88	44.50	43.73	46.02	39.72	38.88
Run 30	43.34	42.95	38.88	45.64	42.16	42.16	46.39	44.89	43.73	45.64	40.95	48.22	40.95	44.12	42.16	47.13	42.16	46.02
Mean	39.60	41.96	40.95	41.49	42.70	44.37	44.59	43.21	44.26	43.83	42.98	44.06	42.29	42.46	42.65	43.13	42.06	41.91
Std dev	0.0333	0.0305	0.0290	0.0289	0.0260	0.0267	0.0311	0.0196	0.0288	0.0227	0.0331	0.0357	0.0334	0.0296	0.0302	0.0295	0.0300	0.0334

Table B-29: Detailed results for the 30 runs of the cross-validation experiments using the video processing system with a peripheral pixel threshold of 25, a foveal pixel threshold of 5, block thresholds of 0, and the TAL_BY_10 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	33.18	39.30	40.54	43.34	41.76	38.04	41.76	45.27	43.34	44.12	44.12	46.02	40.54	48.58	37.61	38.04	42.16	44.89
Run 2	41.35	42.55	42.16	46.76	39.30	43.34	40.95	41.76	43.34	34.99	38.46	40.54	44.89	34.54	46.76	38.46	41.76	40.95
Run 3	39.72	44.50	38.46	40.54	44.89	41.76	44.50	36.75	42.16	38.46	47.13	38.04	43.34	38.04	44.50	41.76	46.02	40.13
Run 4	38.04	45.27	37.61	41.35	47.50	51.04	41.35	48.58	38.88	42.95	40.54	36.75	38.88	43.34	39.30	42.55	46.39	42.16
Run 5	41.35	40.95	42.55	43.73	41.76	37.18	47.86	48.94	39.72	44.89	42.55	38.46	46.39	46.39	40.54	42.16	44.89	45.64
Run 6	48.94	39.30	43.73	39.72	44.50	38.88	38.88	42.16	41.76	38.04	41.35	38.04	46.02	44.12	42.16	43.34	38.46	45.64
Run 7	42.16	38.46	37.61	41.76	44.12	40.13	41.35	42.16	48.22	42.95	42.16	42.55	41.76	40.95	43.73	41.35	40.54	43.34
Run 8	41.35	38.46	47.13	38.88	39.72	44.12	46.02	42.55	44.12	44.12	40.95	41.76	42.55	50.00	42.16	46.76	44.50	40.54
Run 9	39.30	43.34	38.04	42.16	45.64	44.50	39.30	44.89	37.61	38.46	44.89	43.34	40.13	41.35	39.72	46.02	38.88	48.22
Run 10	40.13	45.64	41.35	43.34	37.61	43.73	44.50	38.46	44.50	38.04	38.46	42.55	42.55	41.76	48.22	43.34	40.95	42.55
Run 11	35.87	44.89	40.95	40.95	45.64	39.72	46.02	39.72	43.73	42.16	47.50	38.88	35.43	39.30	47.50	43.34	42.95	42.16
Run 12	41.35	40.95	38.88	39.30	43.34	40.13	46.02	44.89	39.72	46.76	44.50	42.95	41.35	47.13	43.73	43.73	46.02	40.54
Run 13	46.02	44.50	40.54	44.89	43.73	43.73	43.73	42.55	44.12	41.76	43.73	42.95	41.35	37.18	43.73	39.30	43.34	40.54
Run 14	36.75	40.13	42.95	45.64	44.89	44.50	34.54	37.18	39.30	46.02	44.12	41.35	40.13	40.95	43.34	44.50	45.64	44.89
Run 15	38.46	43.34	39.72	43.73	42.16	38.04	39.72	44.89	40.54	45.64	39.72	44.50	39.72	37.61	42.95	42.55	42.16	44.89
Run 16	40.95	42.16	50.35	41.76	38.88	35.87	45.27	40.95	46.02	37.18	46.02	44.89	40.95	46.76	44.89	43.34	46.39	39.72
Run 17	37.61	43.73	34.99	42.55	40.13	43.73	38.88	39.30	40.13	47.13	46.02	42.95	44.89	42.16	44.89	42.16	41.35	37.18
Run 18	34.54	38.04	38.04	44.12	42.95	46.76	40.54	44.89	43.73	39.72	44.12	39.72	43.73	41.76	46.02	36.75	43.73	38.88
Run 19	37.18	46.76	47.86	42.95	47.50	39.30	45.64	42.95	39.72	44.89	39.72	38.46	47.13	40.13	48.22	44.89	40.95	42.95
Run 20	38.46	44.50	41.35	46.76	43.73	36.31	45.64	43.73	40.54	38.46	40.13	44.50	39.30	51.04	44.12	43.73	36.75	38.04
Run 21	41.35	45.64	40.95	40.95	42.95	39.72	44.50	38.88	42.16	44.89	36.31	43.34	50.00	37.18	39.72	41.35	43.34	44.12
Run 22	38.04	35.43	41.76	39.30	46.76	47.50	44.12	41.35	40.95	52.40	46.39	47.13	42.16	44.50	39.72	41.76	45.64	39.72
Run 23	47.86	39.30	37.18	44.12	41.76	43.34	40.13	41.35	44.50	40.54	47.50	41.76	38.88	47.86	42.95	42.55	36.75	44.89
Run 24	37.61	46.39	38.04	41.35	40.13	43.73	41.35	44.50	41.76	43.34	37.61	44.89	48.22	43.34	43.73	45.64	37.61	38.46
Run 25	40.54	43.73	44.12	45.27	38.04	42.55	39.30	41.35	42.16	34.54	47.86	45.27	36.31	38.46	43.73	46.02	45.27	48.22
Run 26	42.55	40.95	46.02	39.30	40.95	42.95	43.73	42.55	38.88	38.46	46.76	42.16	42.95	43.34	38.88	43.73	47.86	41.76
Run 27	42.95	51.72	41.76	40.54	43.73	39.72	45.64	39.72	44.89	43.73	41.76	39.72	42.16	38.04	48.94	47.13	40.54	42.95
Run 28	40.54	32.26	45.27	42.16	48.58	42.55	44.50	40.95	41.76	38.46	46.39	38.04	43.34	43.73	42.55	46.76	42.95	44.12
Run 29	40.95	45.64	38.04	42.95	41.76	42.95	44.12	44.89	41.35	42.95	42.55	39.72	39.72	42.95	43.34	46.39	38.88	41.76
Run 30	39.72	44.50	41.76	39.72	45.64	43.73	44.50	46.02	42.16	44.50	40.95	42.55	33.64	46.76	42.16	51.38	42.55	47.13
Mean	40.16	42.41	41.32	42.33	43.00	41.99	42.81	42.47	42.06	42.02	43.01	41.79	41.95	42.64	43.33	43.36	42.51	42.57
Std dev	0.0339	0.0380	0.0347	0.0219	0.0284	0.0335	0.0296	0.0294	0.0235	0.0393	0.0322	0.0268	0.0357	0.0409	0.0285	0.0297	0.0301	0.0288

TABLE B-30: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM FOR OPTIMISING THE PERIPHERAL BLOCK THRESHOLD ON THE MNI_BY_5 TEMPLATE IN %. THE THRESHOLD VALUE TESTED HERE WAS 0.0.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	37.61	38.04	46.39	40.95	46.39	42.95	40.13	37.61	38.88	37.18	43.73	39.72	42.55	44.12	44.50	44.50	50.00	38.88
Run 2	42.16	40.54	39.72	40.95	42.95	38.04	35.43	39.30	46.02	40.54	39.30	40.95	40.95	44.12	44.50	40.13	42.55	40.95
Run 3	38.04	43.73	44.89	44.89	43.34	40.13	42.16	43.34	38.04	40.13	39.72	41.76	35.43	36.75	45.64	38.04	38.46	47.13
Run 4	40.54	42.95	44.12	38.04	40.95	40.95	42.16	33.18	38.46	44.12	39.30	40.13	44.50	47.50	44.50	42.95	38.46	44.89
Run 5	40.54	40.95	44.89	38.46	37.61	43.34	42.95	40.54	41.76	39.30	40.95	40.54	38.04	38.04	41.35	39.72	36.75	43.34
Run 6	43.73	43.73	40.54	33.18	33.64	40.54	41.76	40.13	34.09	39.30	39.72	40.54	42.55	36.75	40.95	38.04	40.95	41.35
Run 7	40.95	43.34	38.46	44.12	39.72	40.13	38.46	40.13	46.02	38.46	37.61	41.76	43.34	43.34	41.76	45.27	41.76	40.95
Run 8	36.31	37.18	36.75	40.54	40.54	44.50	40.95	31.79	42.16	35.87	37.61	42.55	37.18	36.75	44.12	42.95	45.27	41.35
Run 9	38.46	39.72	39.30	38.88	41.35	40.13	38.46	38.04	40.13	39.30	30.37	46.39	42.55	41.76	38.04	43.34	42.95	42.16
Run 10	33.64	41.76	39.30	39.30	41.35	32.26	39.72	38.46	40.95	35.43	34.99	46.39	45.64	42.95	43.73	39.72	40.95	42.55
Run 11	43.73	42.55	37.61	41.35	42.95	44.50	41.35	40.95	41.76	42.95	38.46	44.50	38.04	42.55	41.35	35.43	45.27	42.95
Run 12	37.18	44.12	41.76	42.16	48.58	38.46	41.35	37.61	43.73	40.13	44.89	43.73	42.55	41.76	37.18	40.95	38.46	37.18
Run 13	40.95	40.54	40.13	34.54	44.89	40.13	45.27	46.76	37.61	40.95	40.95	39.30	38.88	42.16	45.64	35.43	45.27	42.16
Run 14	39.72	40.54	35.43	42.16	44.89	40.95	42.16	39.30	38.46	42.16	33.64	42.55	35.87	37.18	39.72	38.04	40.54	39.30
Run 15	38.46	40.54	39.72	39.30	42.55	40.95	42.16	39.30	43.73	37.18	36.75	36.75	38.88	32.72	35.43	41.35	35.43	44.12
Run 16	33.18	44.50	38.88	37.61	40.54	44.89	38.46	42.55	39.30	40.54	38.46	38.04	45.27	36.75	39.30	41.76	45.64	39.72
Run 17	41.35	37.61	37.18	45.27	39.72	35.43	36.75	37.61	41.76	37.61	37.18	40.54	39.30	40.54	42.16	41.35	34.99	43.73
Run 18	43.73	40.13	38.46	38.88	39.72	41.76	41.76	36.75	37.61	37.61	35.87	38.88	35.43	42.95	41.76	43.73	37.18	37.61
Run 19	36.31	38.88	45.27	45.64	33.64	38.88	41.35	36.75	38.46	34.99	39.30	40.13	40.54	39.72	39.30	40.54	37.61	46.02
Run 20	41.35	35.43	38.88	39.30	36.31	41.35	35.43	32.72	41.76	42.95	36.75	38.88	35.43	44.12	37.61	38.88	46.39	39.72
Run 21	38.46	44.50	39.30	42.55	40.54	42.16	38.04	38.46	42.16	43.73	40.54	42.16	44.12	36.31	35.43	44.50	42.95	46.39
Run 22	36.75	41.35	38.04	42.55	38.46	39.30	42.95	40.95	36.75	37.61	38.46	46.39	42.16	40.13	42.55	44.89	44.12	42.55
Run 23	42.55	39.72	37.61	37.18	37.61	46.76	36.75	46.02	41.76	38.04	35.43	45.64	43.34	44.50	44.12	41.35	39.30	39.72
Run 24	36.31	44.12	42.55	42.55	40.54	40.13	42.16	43.34	38.46	38.88	39.30	42.16	38.46	42.16	42.55	36.75	43.34	41.76
Run 25	42.55	41.76	40.95	36.31	41.35	40.13	37.18	41.76	40.95	38.46	40.54	40.13	41.76	40.13	42.16	39.30	42.16	46.76
Run 26	39.72	39.30	40.13	40.13	39.30	41.35	42.16	37.61	41.35	39.30	38.04	38.88	46.76	39.30	40.13	46.02	46.39	40.95
Run 27	40.54	42.95	43.34	36.75	40.95	44.50	36.31	42.55	39.30	37.18	40.13	39.72	40.95	42.55	45.64	37.61	42.55	38.88
Run 28	39.72	43.73	34.09	46.02	41.35	38.04	41.35	38.04	40.95	41.76	37.61	38.46	40.54	41.35	42.55	42.95	37.61	44.50
Run 29	40.95	36.31	36.31	38.46	46.76	41.35	36.75	44.50	40.54	40.54	37.18	43.34	42.55	40.95	40.13	36.75	40.13	44.50
Run 30	38.46	39.30	42.16	43.34	37.18	41.35	46.02	39.72	38.88	37.61	42.55	44.12	38.04	36.75	39.72	41.35	44.12	37.18
Mean	39.47	40.99	40.07	40.38	40.86	40.84	40.26	39.53	40.39	39.33	38.51	41.50	40.72	40.55	41.45	40.79	41.58	41.97
Std dev	0.0273	0.0249	0.0301	0.0318	0.0341	0.0283	0.0275	0.0347	0.0257	0.0235	0.0285	0.0259	0.0314	0.0321	0.0283	0.0291	0.0361	0.0276

Table B-31: Detailed results for the 30 runs of the cross-validation experiments using the video processing system for optimising the peripheral block threshold on the MNI_by_5 template in %. The threshold value tested here was 0.1.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	36.31	38.88	40.95	40.54	41.76	35.87	38.88	37.61	40.13	35.43	45.64	40.13	41.76	40.95	39.72	44.50	42.55	43.34
Run 2	34.54	42.16	39.72	35.87	38.04	37.18	46.76	41.76	40.54	37.61	37.61	41.35	44.89	36.31	38.04	40.95	37.18	36.75
Run 3	38.88	33.64	38.04	39.72	38.04	40.95	36.31	37.61	41.35	42.16	40.54	41.35	35.87	41.76	40.54	40.13	42.55	42.95
Run 4	34.99	38.46	39.30	39.30	38.88	37.61	40.95	37.61	41.76	42.55	44.12	39.72	40.95	44.50	40.13	38.04	38.88	39.72
Run 5	38.46	33.18	38.04	40.95	35.43	41.35	40.13	38.04	41.35	44.89	41.76	42.55	41.76	38.46	39.72	37.18	38.46	40.95
Run 6	35.43	43.34	41.76	38.46	37.61	31.32	40.54	38.88	37.18	42.55	43.73	42.55	39.30	42.55	40.13	42.55	35.43	42.95
Run 7	39.30	35.87	40.13	34.54	37.61	41.76	38.04	38.88	40.13	42.95	41.35	44.89	43.34	44.50	33.18	41.35	39.72	40.95
Run 8	38.46	34.09	37.61	46.02	39.72	38.04	38.04	40.95	38.04	44.89	34.09	50.35	43.34	34.54	39.72	39.72	35.43	40.13
Run 9	32.72	35.43	33.64	38.04	35.43	42.16	40.13	44.89	38.04	46.76	40.95	44.12	39.72	42.16	35.87	43.34	42.16	44.12
Run 10	33.64	39.30	34.09	37.61	36.75	42.55	39.72	40.95	42.16	42.55	35.43	39.30	38.46	35.87	42.16	43.34	40.95	44.50
Run 11	37.61	41.35	34.99	40.54	41.35	38.46	39.72	43.34	40.13	46.02	34.09	42.95	47.50	42.95	43.34	42.16	35.87	43.34
Run 12	39.30	34.09	34.99	38.46	42.95	43.34	36.31	37.18	40.54	40.54	38.88	39.72	40.13	39.30	35.87	42.55	44.50	40.13
Run 13	41.35	29.41	42.16	30.85	31.79	38.04	38.04	40.13	39.30	35.87	40.13	40.13	42.55	35.87	37.18	42.55	42.16	45.27
Run 14	38.46	42.55	36.75	37.61	38.88	42.55	37.18	42.16	45.27	38.88	39.30	33.18	38.46	40.95	41.35	46.76	36.31	40.95
Run 15	36.31	42.95	36.75	34.99	40.54	35.87	37.18	40.95	42.16	40.13	34.09	42.55	39.72	42.95	41.76	38.46	40.95	38.04
Run 16	40.54	40.95	39.30	40.13	40.95	41.76	36.31	41.76	40.54	46.02	37.18	42.95	46.02	40.95	45.27	40.54	44.12	43.73
Run 17	34.99	34.09	37.61	41.35	41.76	41.35	39.30	44.50	43.73	36.31	38.46	38.04	34.09	34.99	38.88	44.50	44.12	42.16
Run 18	32.26	42.55	36.31	38.88	35.87	34.54	41.35	35.87	41.76	39.72	43.34	44.12	42.55	40.54	38.04	40.13	38.88	43.73
Run 19	33.64	36.31	43.34	40.13	46.02	38.46	42.95	40.13	42.95	37.18	35.87	40.54	42.95	40.13	43.34	41.76	37.61	40.54
Run 20	38.88	36.31	40.95	39.30	38.04	39.72	35.87	34.09	38.46	44.89	40.13	40.95	38.04	40.95	41.35	40.54	41.35	38.04
Run 21	31.32	38.46	34.54	42.16	38.46	38.04	39.30	48.22	39.72	40.13	45.64	41.76	44.50	42.16	41.35	45.64	39.30	38.88
Run 22	39.30	42.55	39.30	38.88	34.99	38.46	38.04	47.50	38.88	35.43	40.54	41.76	42.55	44.89	34.54	40.54	40.54	41.76
Run 23	40.13	38.46	33.18	37.18	40.13	35.87	36.31	37.61	37.18	37.61	40.54	39.72	38.46	40.54	38.46	39.72	45.64	38.46
Run 24	33.64	39.72	34.09	36.31	39.72	38.88	40.13	38.88	42.16	41.35	41.35	41.35	38.88	41.76	37.61	38.88	44.89	44.89
Run 25	43.73	31.79	38.04	37.18	34.09	40.54	38.46	40.54	41.76	38.88	36.75	34.99	38.46	40.54	40.95	40.95	42.95	40.13
Run 26	38.46	35.87	38.46	34.54	37.61	39.30	38.46	42.95	43.34	42.16	34.09	38.46	33.64	44.89	39.30	42.16	37.18	38.04
Run 27	37.61	33.64	35.43	40.13	35.87	37.61	36.75	45.64	40.95	39.30	38.04	35.43	38.88	36.31	38.46	41.76	34.09	39.30
Run 28	33.64	39.30	38.46	40.54	38.46	38.88	42.55	31.79	37.61	34.99	39.72	40.54	35.87	39.30	40.54	38.04	35.87	36.75
Run 29	32.26	40.95	44.12	32.72	31.32	43.34	36.75	41.35	40.13	43.73	42.55	37.61	41.76	43.73	38.04	34.54	41.76	42.55
Run 30	38.88	32.26	34.54	33.18	35.43	39.30	41.76	34.99	40.13	39.30	43.34	44.12	42.16	40.13	34.99	40.13	34.99	42.55
Mean	36.84	37.60	37.89	38.20	38.12	39.10	39.07	40.23	40.58	40.69	39.64	40.91	40.55	40.51	39.33	41.11	39.88	41.19
Std dev	0.0305	0.0382	0.0292	0.0309	0.0313	0.0274	0.0240	0.0373	0.0195	0.0341	0.0336	0.0323	0.0326	0.0293	0.0270	0.0253	0.0329	0.0241

Table B-32: Detailed results for the 30 runs of the cross-validation experiments using the video processing system for optimising the peripheral block threshold on the MNI_by_5 template in %. The threshold value tested here was 0.2.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	36.75	38.04	39.72	43.34	36.31	37.18	42.16	44.50	38.46	35.43	37.61	42.16	40.95	41.35	34.99	39.72	34.54	43.34
Run 2	39.30	45.27	41.76	41.76	40.13	44.89	42.95	44.50	42.16	45.64	40.54	41.76	44.12	45.27	46.39	43.34	43.34	41.35
Run 3	38.88	46.76	42.55	41.76	46.02	41.76	34.09	42.16	44.12	39.72	41.76	39.72	42.55	35.87	42.95	42.55	40.95	42.95
Run 4	40.54	43.73	39.72	40.95	40.13	42.55	40.54	40.95	41.35	38.88	40.54	43.34	48.94	44.12	37.18	39.30	46.02	38.88
Run 5	37.18	39.72	39.30	46.76	38.46	41.35	48.94	38.04	44.12	38.88	41.76	45.27	44.12	39.72	44.12	40.54	35.43	41.76
Run 6	39.30	40.13	40.54	43.34	33.64	46.02	42.16	37.18	44.89	36.31	39.72	44.50	40.13	42.55	42.16	40.95	40.95	41.76
Run 7	49.30	40.54	41.35	42.95	42.95	38.88	47.50	42.55	45.27	36.31	44.12	45.64	42.16	37.61	42.95	38.46	42.55	40.54
Run 8	41.35	40.54	44.89	41.76	41.76	38.88	40.13	42.55	39.30	42.16	44.50	42.95	49.65	46.39	39.72	41.76	40.13	37.18
Run 9	42.55	35.43	39.30	39.72	38.04	40.54	41.35	39.30	39.72	44.89	41.76	44.50	47.50	41.76	40.95	35.43	42.16	36.75
Run 10	42.16	42.16	40.54	42.55	38.88	42.55	48.58	40.95	41.35	43.34	37.61	42.95	41.76	43.73	44.89	39.72	42.95	34.54
Run 11	35.87	34.99	43.34	38.88	45.27	42.55	42.95	41.35	43.73	38.88	38.46	44.89	40.13	43.73	46.02	37.18	36.75	35.87
Run 12	41.76	40.13	39.30	40.54	37.61	37.61	36.31	39.30	40.54	39.30	44.89	46.39	46.02	36.75	41.76	44.12	40.54	38.04
Run 13	44.50	41.35	31.32	42.95	43.34	38.88	38.88	37.61	35.43	36.31	38.88	36.75	42.16	41.76	40.54	45.64	40.95	46.76
Run 14	40.54	40.13	42.95	38.88	42.95	39.72	47.13	37.61	36.31	36.31	40.13	44.12	40.95	37.61	40.54	42.95	42.95	42.55
Run 15	44.50	42.95	38.04	46.39	39.72	39.30	39.30	42.95	42.95	42.55	34.99	41.76	46.76	38.04	45.27	44.50	41.76	42.16
Run 16	35.43	41.35	39.30	37.18	39.30	41.35	39.30	43.34	38.88	39.72	34.54	46.02	42.16	42.16	44.50	40.54	45.64	39.72
Run 17	45.27	39.30	42.16	34.99	37.61	38.04	43.73	44.89	37.18	47.13	48.22	40.95	36.31	39.30	41.35	40.95	37.18	36.31
Run 18	34.99	39.72	47.13	43.73	38.46	42.55	38.46	38.88	45.27	40.95	38.04	45.27	38.88	38.88	40.13	41.35	40.54	39.30
Run 19	39.30	41.35	41.76	39.72	46.39	38.46	39.72	40.54	35.43	45.64	37.61	37.61	43.73	34.99	36.75	34.54	46.76	40.13
Run 20	38.04	42.55	38.88	41.76	37.61	37.61	42.95	34.54	43.73	41.76	44.50	40.13	38.88	38.88	43.34	42.55	44.50	40.13
Run 21	45.27	40.54	44.89	44.50	42.16	40.54	50.35	40.95	42.95	40.13	42.55	41.76	38.88	37.61	46.39	40.54	38.04	35.43
Run 22	40.95	43.34	39.30	42.16	35.87	42.16	39.30	38.88	41.76	40.95	38.04	48.94	37.18	36.75	38.04	46.39	45.64	39.30
Run 23	38.04	41.35	40.95	44.50	44.89	42.16	46.76	41.76	38.04	42.55	40.95	38.88	42.16	37.18	40.95	42.16	41.76	42.95
Run 24	38.46	42.55	37.61	43.34	38.04	38.04	42.16	36.75	40.95	42.95	38.46	36.31	41.35	44.12	44.50	40.13	42.16	36.75
Run 25	39.72	40.13	42.55	40.95	43.73	43.73	38.88	46.02	42.16	43.73	39.30	43.34	43.73	45.27	42.55	42.95	44.89	38.04
Run 26	36.31	41.35	41.76	47.50	43.73	48.58	41.76	42.16	46.76	44.12	36.75	41.76	40.13	44.12	44.89	42.16	42.95	38.46
Run 27	42.16	40.13	40.95	40.95	47.50	42.95	42.95	45.64	40.54	39.72	40.54	35.87	36.75	38.04	40.13	36.31	40.54	37.61
Run 28	32.26	44.12	38.88	39.72	44.89	38.46	44.50	40.54	37.18	42.55	39.30	48.22	42.55	41.35	41.35	45.27	43.34	39.72
Run 29	34.54	37.61	40.95	38.88	38.88	41.35	42.55	40.13	35.87	42.16	38.04	40.54	37.18	41.76	42.55	38.46	47.50	37.18
Run 30	34.99	39.72	36.31	45.64	41.76	40.95	46.39	41.35	37.61	39.30	40.13	45.64	39.30	43.73	42.16	40.95	42.55	39.72
Mean	39.67	40.90	40.60	41.93	40.87	40.99	42.42	40.93	40.80	40.94	40.14	42.60	41.90	40.68	42.00	41.05	41.86	39.51
Std dev	0.0376	0.0248	0.0285	0.0278	0.0345	0.0264	0.0378	0.0274	0.0319	0.0304	0.0298	0.0331	0.0340	0.0315	0.0282	0.0284	0.0314	0.0273

TABLE B-33: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM FOR OPTIMISING THE PERIPHERAL BLOCK THRESHOLD ON THE MNI_BY_5 TEMPLATE IN %. THE THRESHOLD VALUE TESTED HERE WAS 0.3.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	39.30	45.64	41.35	44.50	44.89	44.12	37.61	47.13	44.89	43.73	40.54	38.88	48.94	45.64	44.50	39.30	40.95	45.27
Run 2	48.58	42.16	43.34	44.50	44.50	38.46	42.16	40.13	46.02	40.95	39.72	43.73	44.12	47.13	44.50	44.89	38.88	44.12
Run 3	42.16	44.12	38.88	41.76	40.13	42.95	44.12	41.76	45.27	35.43	42.16	44.89	45.64	42.16	43.34	40.13	46.76	47.50
Run 4	40.54	44.12	42.95	42.95	50.00	38.04	40.95	40.54	42.16	37.61	42.55	44.12	44.12	42.95	44.50	48.58	49.65	50.35
Run 5	44.50	43.34	38.46	37.18	44.12	44.12	46.02	37.18	39.72	46.39	44.89	36.31	40.13	43.34	44.12	42.16	43.73	41.35
Run 6	43.34	42.95	45.64	40.13	40.95	44.89	44.12	43.73	44.12	43.34	47.13	45.27	45.27	43.73	40.13	45.27	45.64	48.58
Run 7	38.46	42.95	42.55	44.50	48.94	43.34	45.27	40.95	42.16	51.04	38.88	45.27	40.54	40.54	46.02	42.95	45.64	42.16
Run 8	47.50	46.39	39.72	43.34	46.39	41.35	38.46	42.16	42.95	40.13	45.27	46.02	44.89	50.35	41.35	41.76	42.55	43.73
Run 9	34.09	44.89	41.35	45.27	44.89	44.12	42.95	38.04	36.31	42.95	37.18	43.34	45.64	38.46	44.50	38.88	41.76	42.55
Run 10	40.54	44.89	44.89	46.02	39.30	43.34	40.95	42.95	42.95	42.55	40.13	43.73	41.76	44.12	45.27	44.89	43.73	48.22
Run 11	40.13	38.88	38.46	44.50	43.34	41.35	42.55	42.55	41.35	41.35	42.16	44.89	44.12	46.39	46.02	39.30	44.89	45.27
Run 12	42.95	41.35	42.55	40.54	39.72	42.55	49.30	47.13	42.16	42.55	41.35	44.50	43.73	47.86	42.95	44.12	45.64	43.34
Run 13	41.76	41.76	39.30	42.55	44.50	44.12	38.46	43.34	47.50	42.16	45.27	39.72	38.04	43.34	45.64	46.39	40.95	48.94
Run 14	40.54	42.95	47.13	40.54	44.50	44.12	48.58	43.34	40.13	41.35	46.39	41.35	47.50	46.39	46.02	43.34	38.88	39.72
Run 15	39.30	44.89	42.95	37.18	40.13	44.12	46.02	41.35	44.89	41.76	39.30	48.58	42.55	47.13	45.27	45.27	41.35	46.02
Run 16	40.95	43.73	47.13	41.35	42.55	40.95	42.95	47.13	40.54	47.13	45.64	45.64	44.89	39.72	46.02	43.34	44.12	47.86
Run 17	42.16	46.76	41.76	37.61	40.54	44.89	40.95	42.16	37.61	42.95	44.50	44.50	39.72	37.18	38.04	41.35	46.39	48.58
Run 18	42.55	39.72	42.55	43.34	42.16	37.61	38.04	40.54	45.27	39.72	47.13	44.50	43.73	44.50	41.35	44.50	42.55	40.95
Run 19	40.54	40.13	41.76	44.12	42.55	44.89	41.76	41.76	40.54	39.30	38.88	41.76	42.95	46.02	40.95	39.72	39.72	41.76
Run 20	44.50	46.02	35.43	45.27	46.02	41.76	38.04	43.34	40.13	45.27	39.30	40.13	44.89	40.95	41.76	44.89	42.95	43.34
Run 21	39.30	40.95	40.54	34.99	40.13	43.73	40.54	43.73	43.34	45.64	44.89	39.30	43.73	44.12	46.76	48.58	47.50	42.55
Run 22	43.73	38.88	44.12	41.35	42.16	39.30	45.27	39.72	40.13	44.89	40.13	42.55	47.50	42.55	42.16	46.39	43.34	48.58
Run 23	45.27	38.46	40.95	42.95	47.86	43.34	45.64	46.39	47.13	40.54	44.89	46.02	44.12	44.12	38.46	40.95	40.13	47.50
Run 24	40.95	42.16	42.55	40.54	40.13	40.95	46.76	42.55	43.73	41.35	45.64	44.12	38.46	44.89	42.55	46.39	42.55	41.35
Run 25	44.50	41.35	45.64	46.76	42.16	42.95	45.27	45.64	44.89	42.55	46.39	44.89	40.13	47.13	43.34	43.34	42.95	42.95
Run 26	42.95	42.16	39.72	44.12	46.76	46.02	41.76	43.34	40.13	40.54	41.76	41.76	47.50	42.55	45.27	41.76	39.72	43.73
Run 27	44.50	40.13	40.54	40.13	46.76	41.76	43.73	42.16	38.46	41.35	41.76	44.89	43.73	40.95	39.72	44.12	47.50	48.58
Run 28	44.89	37.61	42.95	47.86	46.76	36.75	47.86	46.02	42.16	37.61	43.34	42.55	47.86	46.02	48.94	44.12	42.95	43.34
Run 29	39.72	40.54	38.88	43.73	44.50	40.95	41.35	43.73	39.30	43.34	43.73	46.39	46.02	47.50	49.65	42.55	42.16	46.02
Run 30	41.76	45.64	41.35	43.34	46.39	40.54	42.95	40.13	39.72	44.12	45.27	47.50	40.95	44.50	45.27	40.54	41.35	38.88
Mean	42.06	42.52	41.85	42.43	43.79	42.24	43.01	42.69	42.19	42.32	42.87	43.57	43.77	44.07	43.81	43.32	43.23	44.77
Std dev	0.0283	0.0249	0.0264	0.0296	0.0290	0.0235	0.0315	0.0251	0.0277	0.0303	0.0280	0.0267	0.0278	0.0292	0.0274	0.0259	0.0268	0.0307

TABLE B-34: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM FOR OPTIMISING THE PERIPHERAL BLOCK THRESHOLD ON THE MNI_BY_5 TEMPLATE IN %. THE THRESHOLD VALUE TESTED HERE WAS 0.4.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	39.30	39.72	42.16	44.89	38.88	37.61	38.46	36.31	35.43	44.12	42.16	40.95	43.34	34.54	38.04	40.54	38.88	34.99
Run 2	35.43	36.31	40.95	37.61	40.95	37.18	44.50	39.72	41.35	37.18	39.72	40.95	35.87	40.95	40.95	44.12	41.35	34.99
Run 3	42.16	47.50	37.61	47.50	43.73	42.95	46.76	37.61	46.02	41.76	36.31	37.61	38.88	41.76	38.88	34.99	37.61	36.31
Run 4	41.35	42.55	44.50	40.95	43.73	44.89	41.76	42.95	35.87	40.13	32.26	43.34	40.13	40.13	44.50	34.54	43.73	41.76
Run 5	39.72	43.73	46.39	38.46	40.54	38.46	40.13	38.88	40.54	40.13	48.22	38.04	35.43	45.64	43.73	36.75	40.13	38.88
Run 6	41.76	38.04	38.88	43.34	40.13	42.16	42.16	40.13	41.35	41.76	43.34	40.54	35.43	33.64	38.04	42.16	38.88	38.46
Run 7	38.46	40.54	41.35	39.72	40.95	46.02	40.95	39.30	44.50	42.95	37.18	45.27	41.35	38.04	40.54	42.55	35.43	38.04
Run 8	38.88	41.35	39.72	41.76	38.04	43.34	40.95	41.35	41.35	36.31	35.87	37.18	44.12	37.61	31.79	36.75	38.04	40.13
Run 9	41.35	43.73	40.13	38.88	42.16	45.27	38.88	49.30	39.72	46.76	36.31	40.54	42.16	41.35	39.72	38.88	40.13	35.87
Run 10	40.13	38.46	46.39	44.12	42.55	38.04	43.34	42.16	42.95	40.95	40.95	39.30	43.34	40.95	39.72	35.87	39.72	38.04
Run 11	45.64	40.54	42.55	40.13	43.73	43.73	39.72	44.89	40.54	40.54	42.55	36.31	39.30	44.50	37.61	38.46	46.02	39.30
Run 12	37.61	39.72	35.87	45.64	45.64	44.89	42.95	43.34	42.55	46.02	40.13	38.04	38.88	41.35	41.35	38.04	39.30	33.64
Run 13	43.73	39.30	43.73	39.72	40.95	39.30	36.31	39.72	39.72	43.73	39.72	45.64	37.18	38.46	40.54	37.61	44.50	38.46
Run 14	42.55	42.95	47.50	47.86	45.27	37.18	43.73	44.12	40.13	37.18	40.54	40.95	35.87	43.73	36.75	37.61	35.87	35.43
Run 15	40.54	36.75	37.18	37.18	46.02	42.95	37.61	44.50	38.88	45.64	39.72	35.43	37.61	44.12	39.30	44.12	38.88	39.30
Run 16	47.50	36.75	44.12	44.50	41.76	46.02	42.16	45.27	41.35	43.34	38.04	38.88	40.54	42.16	37.18	40.54	42.16	36.75
Run 17	42.16	41.35	33.64	41.35	45.64	42.16	46.39	38.04	39.30	35.87	40.95	31.79	39.72	39.72	40.95	34.09	39.72	41.76
Run 18	38.04	34.99	43.34	45.64	37.61	40.95	38.04	35.43	38.88	42.16	40.54	33.18	38.04	37.18	34.09	40.95	40.95	35.87
Run 19	38.04	39.72	35.43	43.73	44.50	43.73	43.73	42.95	40.95	43.34	42.95	44.12	35.87	38.46	35.87	42.55	40.54	35.43
Run 20	34.54	36.75	40.13	37.61	44.89	39.30	40.54	41.76	38.46	34.54	40.95	37.18	44.89	36.31	35.43	31.79	38.46	44.50
Run 21	35.87	43.73	41.76	41.76	46.02	41.35	40.13	39.30	41.76	32.72	43.34	42.95	40.95	48.58	42.55	43.34	39.30	39.30
Run 22	35.43	36.31	42.55	42.16	42.95	42.55	40.13	43.34	35.43	38.88	38.88	38.88	38.46	40.95	42.95	33.18	33.64	35.87
Run 23	40.95	42.95	44.89	38.46	41.35	47.50	45.64	38.88	41.76	42.55	46.76	43.34	38.88	40.13	34.99	41.76	40.95	35.43
Run 24	44.12	44.12	40.54	40.95	39.30	38.46	35.43	37.18	38.46	39.72	51.04	43.73	39.30	41.76	41.35	44.50	39.72	42.95
Run 25	35.87	38.88	44.50	42.95	45.64	42.16	48.58	38.46	39.72	37.61	39.72	42.95	40.13	38.46	39.72	41.76	38.04	36.31
Run 26	42.55	43.34	41.35	41.35	39.72	42.16	39.30	36.75	37.61	39.30	36.75	38.04	44.50	34.54	40.13	41.76	42.16	37.61
Run 27	41.35	41.76	40.13	41.76	34.54	43.73	36.31	43.34	37.18	44.12	40.54	40.95	35.43	42.16	31.79	42.16	38.88	42.55
Run 28	39.30	41.35	42.55	39.72	43.73	39.30	45.27	38.46	38.46	43.73	38.88	38.88	41.76	39.72	44.12	40.95	42.16	45.27
Run 29	41.76	40.13	43.34	44.89	47.50	38.88	42.55	46.76	37.61	41.35	44.50	37.61	37.18	34.99	44.12	37.61	40.54	40.13
Run 30	38.04	43.73	39.30	39.72	40.54	43.34	39.72	38.04	46.02	38.46	34.09	39.72	39.72	36.31	41.76	42.55	36.31	40.95
Mean	40.14	40.57	41.42	41.81	42.30	41.85	41.40	40.94	40.13	40.76	40.43	39.74	39.48	39.94	39.28	39.42	39.73	38.48
Std dev	0.0306	0.0292	0.0327	0.0287	0.0296	0.0285	0.0322	0.0331	0.0265	0.0341	0.0392	0.0329	0.0278	0.0344	0.0338	0.0347	0.0258	0.0296

TABLE B-35: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM FOR OPTIMISING THE PERIPHERAL BLOCK THRESHOLD ON THE MNI_BY_5 TEMPLATE IN %. THE THRESHOLD VALUE TESTED HERE WAS 0.5.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	34.54	40.95	42.16	47.13	37.61	38.88	42.16	42.55	40.13	40.54	44.12	46.02	41.76	45.64	45.64	43.73	45.64	42.55
Run 2	38.88	30.37	44.89	44.12	44.50	42.16	40.13	42.95	38.88	36.75	45.27	48.58	44.89	40.95	48.22	38.46	41.35	42.55
Run 3	41.76	41.35	36.75	37.61	40.95	37.61	42.95	34.54	40.13	44.12	44.89	40.95	47.86	47.13	40.95	40.95	43.73	38.46
Run 4	33.64	43.73	40.95	43.34	43.34	39.72	36.31	40.13	42.16	45.27	44.12	39.30	48.94	41.35	47.50	46.76	44.89	42.55
Run 5	37.61	41.35	40.13	42.16	39.72	42.16	36.75	42.95	43.34	38.88	38.88	45.27	43.34	44.89	43.73	47.50	39.30	41.76
Run 6	45.64	43.73	34.99	39.30	37.61	41.76	44.89	41.35	43.34	42.55	46.39	43.73	43.34	47.86	39.72	47.13	43.73	42.55
Run 7	40.54	38.46	38.04	38.88	38.46	40.13	42.95	40.95	39.72	39.72	40.54	42.55	38.46	44.50	48.22	42.16	44.89	45.27
Run 8	44.89	34.99	40.13	38.46	38.46	39.30	42.95	41.76	44.12	39.30	41.76	44.89	40.95	40.54	40.54	42.95	46.02	45.64
Run 9	44.89	41.76	46.39	42.95	44.50	39.30	38.88	40.13	40.54	38.04	42.95	45.64	44.50	47.86	42.55	43.34	45.27	43.73
Run 10	43.73	40.95	43.34	45.27	36.75	42.55	40.54	39.72	44.12	44.89	42.16	46.39	43.34	44.50	46.39	42.55	46.02	44.12
Run 11	38.46	35.43	39.72	45.27	41.35	39.72	45.64	44.89	45.64	40.13	41.35	43.73	45.64	39.72	44.89	46.02	46.02	48.22
Run 12	40.13	42.16	42.55	43.73	38.88	46.02	43.73	39.30	44.12	45.27	40.54	40.95	41.76	43.34	44.12	43.34	46.02	44.50
Run 13	42.95	40.13	39.72	38.46	38.46	38.88	45.64	46.39	43.34	43.34	45.64	45.27	46.02	41.76	44.89	41.35	44.50	46.02
Run 14	44.89	44.89	36.31	39.72	35.87	41.76	39.72	41.35	44.50	42.95	42.55	36.31	47.50	44.89	42.55	44.89	40.95	40.13
Run 15	41.35	38.88	37.61	43.73	39.72	37.18	42.55	42.55	39.30	38.46	43.73	45.64	38.88	44.12	45.64	40.54	38.04	42.16
Run 16	37.61	40.13	32.72	39.72	45.27	38.88	40.95	39.72	44.50	43.73	43.73	46.02	40.95	36.31	44.89	44.50	45.64	47.13
Run 17	38.88	45.27	38.88	44.12	40.13	41.76	44.89	41.35	42.95	44.12	40.54	44.89	47.50	44.89	40.13	46.02	44.12	42.95
Run 18	41.76	43.34	39.72	43.73	39.72	42.95	41.76	38.88	40.13	35.87	39.72	46.02	43.34	44.12	46.02	44.89	42.55	48.22
Run 19	42.55	42.95	43.34	39.72	40.13	47.13	41.76	42.16	43.34	40.95	42.55	44.50	41.76	44.89	38.46	45.27	43.34	38.04
Run 20	44.12	37.18	33.18	41.35	40.95	42.55	44.50	48.22	36.75	38.04	46.02	46.76	44.89	44.89	38.46	46.76	44.12	43.73
Run 21	31.79	42.55	36.31	40.13	38.46	41.76	37.61	39.30	42.55	45.64	39.30	45.64	42.55	48.22	43.34	44.50	42.55	42.55
Run 22	42.55	41.76	38.46	43.73	42.55	46.39	40.95	41.35	43.73	47.50	41.35	47.50	44.50	42.16	41.76	47.13	43.34	40.54
Run 23	40.13	40.54	40.13	42.16	38.46	39.72	40.95	44.50	40.54	42.16	36.75	45.64	45.27	49.65	52.40	48.22	46.39	44.50
Run 24	34.99	35.87	40.54	40.13	42.55	42.55	39.72	41.76	40.13	40.13	45.64	49.30	44.12	50.35	47.86	45.27	40.13	44.50
Run 25	37.18	42.55	42.95	38.04	38.88	40.54	44.89	42.55	41.76	46.76	48.94	44.50	50.35	47.13	41.76	43.73	42.16	45.64
Run 26	39.30	40.13	44.89	44.12	44.12	42.16	40.95	42.55	40.95	41.76	41.76	42.95	42.95	46.76	47.13	46.39	41.76	43.73
Run 27	40.13	38.04	41.35	38.88	45.27	43.73	45.27	42.16	48.22	40.13	42.55	39.72	47.86	41.35	44.50	42.16	48.58	49.30
Run 28	40.95	44.50	40.54	43.34	44.12	43.34	45.27	46.39	46.76	48.58	45.27	42.95	41.76	44.89	41.35	44.89	47.50	43.34
Run 29	34.99	40.54	36.75	43.34	44.89	38.46	38.88	43.73	41.76	46.02	41.76	42.95	48.94	41.76	40.95	45.27	42.16	46.02
Run 30	38.88	35.87	37.61	38.46	38.88	42.16	45.64	42.16	40.95	47.50	41.35	44.12	48.58	45.27	44.50	45.64	45.64	47.86
Mean	39.99	40.35	39.70	41.70	40.69	41.37	41.99	41.94	42.28	42.30	42.74	44.29	44.42	44.39	43.97	44.41	43.88	43.94
Std dev	0.0355	0.0334	0.0328	0.0258	0.0272	0.0243	0.0269	0.0257	0.0245	0.0338	0.0257	0.0273	0.0300	0.0305	0.0321	0.0227	0.0240	0.0270

Table B-36: Detailed results for the 30 runs of the cross-validation experiments using the video processing system for optimising the peripheral block threshold on the MNI_by_5 template in %. The threshold value tested here was 0.6.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	46.39	43.73	42.55	46.39	39.72	34.54	38.04	42.55	38.88	40.95	44.50	41.35	49.30	40.13	42.55	44.50	42.16	47.13
Run 2	39.30	44.12	37.61	44.89	44.50	39.30	39.30	36.75	47.86	45.64	44.50	39.30	43.73	41.35	44.89	36.75	42.55	43.73
Run 3	41.35	45.27	46.39	44.89	38.46	39.30	43.73	40.95	40.95	44.50	46.39	41.35	46.02	43.73	43.34	38.88	41.35	41.76
Run 4	40.13	42.16	44.12	41.35	39.30	39.72	43.73	44.12	48.58	39.30	48.58	44.50	45.64	42.55	42.16	41.35	42.16	43.73
Run 5	32.72	42.55	40.13	42.55	43.73	41.35	44.50	42.16	42.95	41.35	43.34	41.76	34.54	40.13	40.54	38.04	47.13	43.34
Run 6	38.04	34.99	35.87	47.86	44.89	46.39	41.35	47.13	44.89	42.55	44.50	40.95	44.50	45.64	45.27	44.50	34.99	43.34
Run 7	46.39	48.58	44.50	42.55	40.95	46.76	38.04	40.95	41.76	39.72	43.34	47.50	46.76	48.94	47.50	45.27	45.64	43.73
Run 8	43.73	42.16	43.34	39.72	40.95	40.54	36.75	40.95	43.34	42.55	42.55	44.50	45.64	44.12	42.16	43.73	42.16	38.88
Run 9	36.75	42.16	46.39	41.35	38.46	41.76	41.76	44.12	38.46	48.94	40.13	42.16	41.35	46.02	44.12	42.95	43.34	43.34
Run 10	44.50	39.30	39.30	46.02	43.73	37.61	40.13	43.73	39.72	41.76	43.73	40.13	39.72	38.88	45.27	41.76	40.54	44.12
Run 11	40.95	40.95	40.13	42.95	42.55	36.75	37.18	39.72	39.72	40.13	43.34	42.95	42.95	47.86	44.50	44.89	48.22	43.34
Run 12	33.64	40.54	46.02	44.89	44.89	40.95	42.55	46.76	41.76	40.95	37.18	46.39	42.55	44.50	43.73	44.89	44.89	47.86
Run 13	41.76	37.61	44.50	45.27	40.95	46.39	37.61	43.73	42.16	41.35	46.02	43.73	42.95	44.50	41.76	47.50	35.87	40.13
Run 14	45.27	37.61	36.75	43.73	42.16	38.04	44.12	45.27	46.02	45.27	43.73	48.94	41.35	41.35	40.54	47.13	46.02	46.76
Run 15	40.13	36.75	48.94	39.30	45.27	37.18	39.30	40.95	42.95	42.95	44.89	46.02	44.89	44.12	46.76	41.76	46.39	43.73
Run 16	45.27	40.95	36.31	43.34	42.55	38.04	38.46	40.95	40.54	40.13	44.12	40.95	46.39	41.35	34.99	39.72	42.16	39.30
Run 17	45.27	37.61	48.58	41.76	46.02	45.64	36.75	41.35	39.72	42.95	42.55	47.86	43.34	48.22	38.88	48.22	44.50	40.95
Run 18	36.31	40.95	42.95	41.76	40.13	42.55	42.55	48.94	45.27	41.76	43.34	38.04	42.16	47.50	48.58	46.39	41.35	47.13
Run 19	39.30	43.34	44.89	46.02	38.46	38.46	38.46	41.35	37.61	41.76	44.12	52.06	42.95	38.88	42.16	41.35	42.55	42.95
Run 20	42.55	41.76	42.95	39.30	40.13	41.76	39.72	45.27	40.54	42.16	42.55	34.99	44.50	39.30	37.61	43.73	45.64	36.75
Run 21	40.95	38.88	37.18	38.88	41.35	40.13	44.50	43.73	42.95	45.27	46.02	38.46	46.02	43.73	48.22	41.35	44.89	38.88
Run 22	39.72	39.30	42.95	38.88	42.95	40.13	42.55	41.35	50.00	43.34	41.35	42.95	48.58	44.12	42.55	38.88	43.34	41.76
Run 23	40.13	43.34	40.95	43.73	46.02	40.54	36.75	44.89	47.13	43.34	48.22	45.27	43.73	46.02	39.72	44.50	49.30	37.61
Run 24	40.13	40.95	40.54	44.12	42.55	38.46	44.50	47.13	42.55	42.16	46.02	43.34	45.64	43.73	46.02	44.50	39.30	41.76
Run 25	39.30	41.35	48.22	41.76	43.73	42.55	39.72	44.12	40.13	43.73	41.76	41.76	44.50	43.34	42.55	44.12	44.50	42.55
Run 26	40.95	40.95	41.35	41.76	40.95	33.64	37.61	33.64	34.54	37.61	43.34	43.73	48.22	46.76	43.73	45.27	43.34	45.27
Run 27	43.73	42.55	44.50	39.72	39.72	40.54	44.12	39.30	43.73	49.30	42.95	45.27	44.12	46.02	42.95	45.64	41.35	47.13
Run 28	39.72	39.30	39.72	47.50	41.76	43.73	46.02	44.89	40.13	37.18	38.46	43.34	44.89	47.13	46.39	40.95	41.35	38.46
Run 29	44.89	40.95	44.12	39.72	37.61	43.73	43.73	42.95	46.39	44.12	46.76	41.76	44.50	43.34	42.95	43.73	45.27	42.95
Run 30	41.35	41.76	42.95	36.75	42.55	41.76	40.95	44.89	41.35	47.50	46.02	44.89	42.95	40.54	38.88	43.34	42.55	42.16
Mean	41.02	41.08	42.49	42.62	41.90	40.61	40.82	42.82	42.42	42.67	43.81	43.21	44.15	43.79	43.04	43.19	43.16	42.68
Std dev	0.0339	0.0268	0.0357	0.0278	0.0233	0.0321	0.0284	0.0311	0.0342	0.0281	0.0247	0.0342	0.0277	0.0284	0.0309	0.0280	0.0308	0.0288

Table B-37: Detailed results for the 30 runs of the cross-validation experiments using the video processing system for optimising the peripheral block threshold on the MNI_by_5 template in %. The threshold value tested here was 0.7.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	38.46	44.89	38.88	45.27	40.13	41.76	44.89	46.76	42.16	45.64	40.13	44.50	44.89	41.35	46.02	45.27	41.35	48.58
Run 2	44.89	42.95	43.73	36.75	44.12	44.12	41.35	40.95	37.61	43.73	36.75	42.95	45.64	47.50	44.50	46.02	40.95	46.76
Run 3	37.61	46.02	39.72	36.75	41.76	40.54	39.30	38.04	45.27	38.88	42.55	41.35	38.04	44.89	39.72	45.27	45.27	44.50
Run 4	43.73	45.64	39.72	42.95	43.34	39.72	42.55	41.76	43.34	38.46	38.46	42.95	42.55	47.13	47.13	42.55	45.64	42.16
Run 5	37.18	45.64	44.89	39.30	40.13	42.55	37.61	38.46	37.61	43.73	40.54	44.50	40.95	41.76	47.13	42.55	39.30	44.12
Run 6	42.55	39.30	46.02	45.27	45.27	35.43	46.76	35.87	43.34	38.04	38.88	45.64	42.16	44.12	40.54	46.02	41.35	42.16
Run 7	33.64	40.54	42.55	49.65	38.46	43.34	37.61	38.04	44.89	42.95	38.88	39.30	44.12	45.64	47.86	45.64	45.27	42.55
Run 8	40.95	38.88	38.04	42.55	41.35	43.73	36.31	47.86	42.55	38.46	43.34	38.88	41.76	44.12	44.89	48.58	48.22	43.34
Run 9	39.30	45.64	38.46	39.30	39.30	40.54	46.39	45.27	37.61	39.72	42.55	39.72	41.76	42.55	40.13	44.89	46.39	43.73
Run 10	43.73	35.43	42.95	39.30	41.76	41.76	39.30	38.46	35.87	37.61	43.34	44.50	41.76	45.27	42.55	44.12	48.22	46.39
Run 11	34.54	42.55	42.95	40.54	38.04	42.55	38.88	40.95	35.43	35.43	40.54	45.27	45.64	40.95	47.50	45.64	45.27	46.02
Run 12	37.61	40.13	44.89	42.55	38.46	42.16	36.75	39.30	38.04	39.30	37.18	40.54	46.02	44.50	45.64	41.76	48.58	44.89
Run 13	42.95	39.72	37.61	41.35	37.18	43.34	43.34	38.88	40.54	38.04	40.54	44.50	40.13	40.13	44.12	49.30	43.34	44.50
Run 14	35.87	38.04	41.35	42.16	43.73	44.50	39.72	38.04	42.16	47.13	31.79	46.76	39.72	43.34	46.02	44.12	43.73	44.50
Run 15	40.54	42.16	43.73	37.61	38.88	37.18	37.61	38.88	40.95	41.35	40.13	44.89	47.50	43.34	43.34	46.39	45.64	50.00
Run 16	37.18	47.50	46.02	46.02	41.76	33.18	33.64	38.46	41.35	42.55	40.54	42.95	46.76	46.39	43.73	45.27	40.95	48.22
Run 17	34.54	40.13	36.75	40.54	37.61	38.46	37.61	42.55	41.76	37.61	35.43	39.72	41.76	47.86	46.76	40.54	42.95	44.50
Run 18	37.18	37.18	38.04	43.73	36.31	34.54	39.72	43.34	41.35	38.46	43.73	41.76	41.35	41.35	43.34	43.73	44.12	47.50
Run 19	44.12	38.46	37.18	46.76	44.12	41.35	42.16	39.30	44.12	43.34	46.02	46.76	42.95	47.13	47.13	50.00	47.86	46.76
Run 20	36.31	42.95	40.54	42.95	35.87	37.18	36.75	45.27	45.27	47.86	40.95	40.13	43.34	44.12	44.50	46.39	44.89	45.27
Run 21	40.13	39.30	42.55	39.72	40.95	31.79	43.73	42.55	40.54	41.76	41.76	42.95	43.34	46.02	45.27	38.88	37.18	46.02
Run 22	31.79	38.04	46.39	39.72	37.18	38.46	42.16	42.16	43.34	40.13	37.61	48.22	42.16	42.16	40.13	47.50	42.55	41.35
Run 23	36.75	39.72	41.76	50.35	39.72	34.54	44.89	43.34	44.12	43.34	44.89	50.70	46.02	44.50	40.95	44.50	47.13	45.27
Run 24	41.35	38.04	42.16	44.50	37.61	42.16	42.16	30.37	42.16	43.34	38.46	46.76	43.73	45.64	42.55	35.43	43.34	45.64
Run 25	38.88	43.73	37.61	47.50	39.72	39.30	38.04	38.88	42.55	46.76	42.16	40.54	45.27	40.54	42.95	48.94	46.39	44.89
Run 26	43.73	39.30	40.54	40.13	43.73	38.04	41.76	42.55	34.99	42.16	34.99	47.86	42.16	41.76	47.13	45.64	47.13	46.39
Run 27	36.31	36.75	47.13	44.50	42.16	44.50	34.09	38.88	37.61	44.12	39.30	40.95	46.76	47.50	52.06	41.76	32.26	42.16
Run 28	37.18	38.46	39.30	38.46	38.46	45.27	34.99	35.87	40.13	40.13	40.95	35.87	42.55	43.73	46.39	42.16	39.30	46.39
Run 29	31.79	39.72	42.55	41.76	42.16	35.43	37.61	35.43	44.12	44.12	38.88	45.64	37.61	38.88	39.30	51.72	40.95	44.89
Run 30	46.39	38.04	35.87	38.88	47.86	35.87	36.31	38.46	38.46	36.31	36.75	46.76	34.99	47.13	47.86	47.50	44.12	47.50
Mean	38.91	40.83	41.33	42.23	40.57	39.78	39.80	40.17	40.97	41.35	39.93	43.46	42.78	44.04	44.57	44.94	43.65	45.23
Std dev	0.0384	0.0314	0.0312	0.0352	0.0286	0.0369	0.0350	0.0360	0.0296	0.0326	0.0304	0.0329	0.0284	0.0245	0.0295	0.0332	0.0358	0.0203

TABLE B-38: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM FOR OPTIMISING THE PERIPHERAL BLOCK THRESHOLD ON THE MNI_BY_5 TEMPLATE IN %. THE THRESHOLD VALUE TESTED HERE WAS 0.8.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	41.35	32.72	41.76	42.16	40.13	38.46	37.61	46.02	40.54	40.95	44.89	45.27	45.64	40.54	44.12	37.61	52.40	46.76
Run 2	32.26	38.04	43.34	44.89	39.72	42.16	44.12	34.09	40.95	41.35	39.30	45.27	39.72	45.27	47.13	41.76	40.95	40.54
Run 3	40.13	38.04	40.13	35.43	42.55	39.72	44.12	40.13	34.99	40.54	45.64	44.89	46.39	41.35	40.13	47.13	48.94	46.76
Run 4	34.99	40.54	39.72	44.50	43.73	41.35	37.18	39.30	43.34	38.46	47.50	40.95	37.18	48.58	42.55	44.89	42.95	40.95
Run 5	42.95	41.35	43.73	42.95	42.16	41.35	37.61	37.61	40.95	40.95	44.89	45.64	46.02	45.64	45.27	45.64	45.27	42.55
Run 6	40.95	38.46	46.76	43.73	38.46	35.87	38.88	48.94	41.35	40.95	43.73	42.55	44.89	48.22	48.94	44.89	42.55	40.13
Run 7	37.61	38.88	37.61	42.55	42.16	40.13	43.34	38.46	42.55	35.43	42.55	43.34	40.95	47.50	48.22	46.02	48.94	43.73
Run 8	39.72	42.16	40.95	42.16	39.72	35.43	36.75	38.46	33.64	48.22	42.16	46.02	47.50	46.39	46.39	48.94	41.35	42.55
Run 9	39.30	36.31	48.22	39.72	36.31	46.76	39.30	42.95	39.72	37.61	38.04	43.34	37.61	47.86	47.13	51.04	47.86	46.39
Run 10	30.37	39.72	46.02	37.61	39.72	40.95	40.95	40.54	38.04	38.46	44.50	43.34	44.50	48.94	46.39	48.94	43.34	41.35
Run 11	38.46	41.35	44.50	40.95	40.54	40.54	40.13	40.13	38.88	44.12	46.76	47.13	43.73	41.35	44.50	45.27	48.22	45.64
Run 12	40.54	37.61	36.75	47.86	39.30	35.87	43.34	40.95	45.27	41.35	41.35	43.34	46.39	47.50	44.89	40.95	43.34	39.72
Run 13	36.31	39.72	43.34	41.35	40.13	43.34	39.30	48.94	42.95	44.12	44.12	44.50	47.13	47.13	39.72	43.73	42.95	43.73
Run 14	40.54	36.75	39.72	40.13	38.88	47.13	33.18	29.89	42.95	38.88	45.64	42.55	48.22	43.34	46.39	43.73	48.22	48.58
Run 15	39.30	47.13	44.50	37.61	41.76	34.54	41.76	38.88	41.35	42.16	44.50	42.16	48.58	42.55	39.72	48.94	43.73	39.72
Run 16	35.87	40.13	42.16	38.88	45.27	36.75	39.72	41.35	34.99	36.75	38.88	42.55	46.39	46.76	46.02	42.16	44.89	45.27
Run 17	34.09	42.55	38.46	44.89	45.27	46.76	45.27	41.35	38.46	42.16	44.89	45.64	43.73	43.73	45.64	45.64	48.22	41.76
Run 18	43.34	42.55	42.16	42.55	41.35	42.16	35.43	36.31	37.18	41.76	40.13	39.30	43.73	47.86	41.76	46.39	47.13	42.16
Run 19	40.13	42.95	45.64	41.35	39.30	40.54	42.16	42.55	40.95	42.16	43.34	43.34	40.95	39.30	48.58	45.27	39.72	44.12
Run 20	43.34	46.39	42.55	46.39	36.31	43.73	34.54	41.35	38.88	48.22	41.76	50.00	45.64	48.22	42.55	44.89	44.50	48.22
Run 21	38.46	38.04	45.27	41.35	37.61	39.72	32.72	40.54	38.04	35.87	40.95	38.46	51.04	48.58	47.86	46.76	42.95	44.89
Run 22	35.43	42.55	41.35	43.73	36.75	42.16	44.89	40.95	43.34	36.75	37.18	45.27	45.27	46.76	46.39	39.30	40.13	39.30
Run 23	38.04	42.55	43.34	36.31	40.54	40.13	37.18	40.54	36.75	38.88	42.95	42.95	48.22	38.46	50.70	42.16	50.00	41.35
Run 24	40.13	47.50	37.18	44.12	40.13	41.35	34.54	45.27	38.88	39.72	38.04	43.34	48.58	44.89	44.89	47.13	47.86	42.16
Run 25	35.87	38.88	44.12	44.50	35.87	34.54	38.88	39.30	36.75	42.55	39.30	41.76	42.95	42.95	44.12	44.50	42.55	44.89
Run 26	34.09	36.75	38.88	40.54	46.76	40.95	41.35	38.04	44.50	43.34	38.46	42.55	43.73	39.30	46.39	47.86	44.12	38.46
Run 27	42.55	42.16	46.76	38.88	39.72	40.95	36.75	39.72	37.18	40.54	38.04	42.55	42.16	41.35	42.55	45.64	48.22	47.13
Run 28	42.95	43.73	39.72	39.72	40.95	47.50	44.89	43.73	39.30	40.13	43.34	45.27	47.86	45.64	43.73	38.46	42.55	44.50
Run 29	35.43	42.95	37.18	36.75	42.55	36.31	39.72	39.30	40.13	40.54	41.76	46.76	43.73	45.64	40.13	50.00	38.88	48.94
Run 30	45.64	39.72	40.13	40.13	43.34	39.72	40.95	38.04	39.30	38.46	40.54	43.34	44.89	49.65	44.12	37.61	45.27	44.12
Mean	38.67	40.61	42.06	41.46	40.57	40.56	39.55	40.45	39.74	40.71	42.17	43.78	44.78	45.04	44.90	44.77	44.93	43.55
Std dev	0.0356	0.0325	0.0311	0.0299	0.0265	0.0355	0.0351	0.0377	0.0284	0.0300	0.0284	0.0227	0.0319	0.0321	0.0280	0.0348	0.0332	0.0289

Table B-39: Detailed results for the 30 runs of the cross-validation experiments using the video processing system for optimising the peripheral block threshold on the MNI_by_5 template in %. The threshold value tested here was 0.9.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	39.30	41.76	46.39	41.76	42.16	41.76	42.55	39.30	35.43	38.88	38.46	44.89	46.76	47.50	47.13	41.76	41.76	37.18
Run 2	41.35	35.87	35.87	36.75	30.85	42.16	38.46	44.50	37.18	41.35	38.88	42.16	43.73	36.31	37.18	44.89	40.54	41.76
Run 3	43.34	41.76	36.75	45.64	38.46	37.61	42.95	46.76	39.30	36.31	40.95	40.95	39.72	38.04	44.12	48.94	42.16	39.72
Run 4	40.54	40.95	40.13	38.46	40.95	37.61	40.54	43.34	36.75	41.35	44.89	41.35	41.76	40.95	38.04	42.55	44.89	39.72
Run 5	38.46	41.35	34.99	40.13	43.34	40.95	44.50	39.72	41.35	36.75	40.54	44.12	38.46	39.72	45.64	39.30	38.88	40.95
Run 6	37.18	41.35	39.72	37.18	43.34	41.35	40.54	38.46	42.55	35.43	40.95	39.72	42.16	43.34	38.04	40.13	44.89	40.54
Run 7	37.61	39.30	35.43	41.76	40.95	41.76	40.95	42.55	38.04	38.88	44.89	38.04	37.61	42.55	38.88	40.54	45.27	41.76
Run 8	42.95	44.89	40.13	40.54	42.16	43.73	45.27	44.12	40.54	38.88	40.54	42.55	40.54	38.46	42.55	38.46	43.34	42.16
Run 9	34.54	46.76	41.35	43.73	36.31	42.16	34.09	44.12	39.30	37.61	41.35	43.34	45.27	53.07	44.50	40.95	44.12	37.61
Run 10	46.02	43.73	42.55	41.76	36.75	39.72	35.87	37.18	36.31	38.88	44.12	42.55	44.50	42.55	41.76	46.39	44.12	42.55
Run 11	34.54	35.87	42.55	43.34	42.55	37.18	38.46	42.95	35.87	44.50	38.46	42.16	41.76	47.13	41.76	41.76	45.64	45.64
Run 12	43.34	42.16	37.18	46.02	40.13	36.31	41.35	43.73	42.95	40.54	44.50	47.50	39.30	42.55	42.16	44.89	41.76	42.16
Run 13	34.99	37.61	39.72	40.54	41.76	41.76	34.09	44.89	41.76	42.55	38.46	42.95	46.02	37.61	45.27	49.65	38.04	39.72
Run 14	36.75	40.95	39.72	40.13	40.13	37.18	39.30	41.76	37.61	36.31	41.35	44.12	40.13	41.35	48.22	44.89	42.55	39.72
Run 15	38.04	40.95	41.76	38.04	38.04	40.95	38.04	38.46	39.30	39.30	33.64	38.46	39.30	42.55	41.35	38.88	44.12	39.72
Run 16	38.46	41.76	40.95	40.54	39.30	40.54	37.18	38.46	28.93	40.13	41.35	46.02	42.55	37.18	44.50	38.46	40.95	39.72
Run 17	36.31	40.95	45.64	40.95	45.27	40.54	39.30	41.76	43.34	40.13	40.54	42.55	46.39	50.00	43.34	41.76	39.30	41.76
Run 18	40.54	41.35	42.16	35.87	44.89	42.16	40.95	36.75	38.46	40.54	38.04	38.46	42.55	41.35	38.46	39.30	42.16	40.54
Run 19	41.76	45.27	35.87	41.35	33.64	42.16	37.18	38.46	36.31	36.31	42.95	44.50	40.13	42.95	42.95	40.13	46.02	38.88
Run 20	40.54	39.30	41.35	39.30	38.46	40.13	40.13	38.46	37.61	34.54	36.75	41.35	42.95	36.75	42.95	32.72	40.54	38.88
Run 21	35.43	42.95	38.04	40.95	34.09	40.95	42.16	40.54	40.95	39.30	38.46	46.76	40.95	39.30	43.73	32.72	44.50	44.89
Run 22	33.64	33.64	40.13	39.30	41.35	36.75	38.88	43.34	41.35	39.30	40.95	41.35	39.30	44.50	43.34	48.22	39.30	40.95
Run 23	36.31	41.35	43.34	36.31	44.89	42.16	35.43	40.13	36.31	41.76	40.54	42.55	42.16	44.12	39.30	38.88	43.34	38.46
Run 24	34.54	38.88	46.39	41.76	43.34	42.55	45.64	40.54	34.54	44.12	42.95	38.88	42.55	52.73	44.12	39.72	46.76	38.88
Run 25	41.35	38.88	39.72	37.61	36.31	38.88	39.72	44.50	36.31	35.87	40.54	41.76	39.72	42.16	40.95	40.95	41.76	41.76
Run 26	39.30	42.95	39.72	36.75	33.18	45.64	40.95	45.64	42.16	36.31	38.04	45.64	41.35	44.50	41.35	42.16	44.50	43.34
Run 27	37.18	40.13	41.76	37.18	45.64	40.95	35.43	40.13	40.13	41.76	32.72	43.34	46.39	46.76	40.95	44.12	43.34	38.46
Run 28	41.76	38.46	34.54	46.02	36.75	35.43	42.55	39.30	42.95	38.46	48.22	49.30	41.76	47.86	46.76	42.16	40.95	39.72
Run 29	38.88	41.76	46.02	44.50	43.73	38.46	41.35	43.73	39.30	42.95	42.55	40.95	44.50	49.30	41.35	41.35	42.95	37.61
Run 30	38.46	47.13	38.46	42.55	39.30	32.72	38.46	35.87	44.50	33.64	42.55	39.72	38.04	43.34	37.61	42.55	38.04	40.95
Mean	38.78	41.00	40.28	40.56	39.93	40.07	39.74	41.32	38.91	39.09	40.64	42.60	41.94	43.22	42.28	41.64	42.55	40.52
Std dev	0.0306	0.0296	0.0328	0.0285	0.0383	0.0271	0.0299	0.0286	0.0324	0.0274	0.0318	0.0270	0.0256	0.0444	0.0287	0.0383	0.0235	0.0199

TABLE B-40: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM FOR OPTIMISING THE FOVEAL BLOCK THRESHOLD ON THE MNI_BY_5 TEMPLATE IN %. THE THRESHOLD VALUE TESTED HERE WAS 0.0.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	42.55	44.89	44.12	42.95	39.30	39.72	46.02	45.64	46.02	40.13	45.27	43.34	46.39	44.50	43.73	40.54	43.34	46.02
Run 2	46.02	34.99	39.72	34.54	42.95	38.46	39.30	43.73	42.55	46.02	46.76	35.87	42.55	44.12	44.89	40.95	46.39	41.76
Run 3	40.13	41.76	44.12	44.12	47.86	37.61	46.76	49.30	39.72	42.95	43.34	45.27	45.64	46.76	48.22	41.76	42.16	48.94
Run 4	34.09	38.88	47.13	45.64	40.54	42.95	40.95	48.22	39.30	39.72	40.13	45.64	49.30	46.02	42.16	46.76	43.34	36.75
Run 5	40.54	42.55	40.54	45.64	42.55	40.95	42.95	41.35	38.88	42.55	44.50	40.13	43.73	39.72	42.55	44.89	44.12	40.54
Run 6	36.75	37.18	40.54	39.72	39.72	40.95	48.58	46.39	42.16	41.76	39.30	41.76	44.50	42.16	45.64	44.50	43.34	44.12
Run 7	43.34	42.55	39.30	42.16	40.54	43.73	40.54	39.30	43.73	46.76	47.86	43.34	42.16	40.13	40.54	37.61	44.12	37.18
Run 8	42.95	39.72	45.64	42.95	46.76	41.76	46.39	40.54	43.73	46.39	45.27	49.30	37.18	39.72	43.73	42.95	44.89	40.13
Run 9	44.50	40.13	34.99	40.54	36.75	40.95	43.73	40.13	40.13	45.27	46.76	41.35	47.50	44.89	39.30	39.72	39.30	36.75
Run 10	39.30	39.72	39.72	38.88	39.72	40.54	43.73	42.55	45.64	40.95	42.95	43.73	46.02	45.27	34.99	44.12	47.13	42.55
Run 11	47.13	39.72	44.50	36.75	39.30	39.30	45.64	42.55	41.35	42.16	46.02	45.64	43.73	48.58	43.34	48.58	42.95	42.55
Run 12	34.54	38.04	43.34	43.34	41.76	37.18	43.34	35.87	42.95	47.86	40.54	44.89	41.76	40.95	41.76	42.95	43.73	44.89
Run 13	46.02	43.34	46.02	44.89	44.50	42.95	42.95	46.76	44.12	39.72	46.02	44.89	50.70	42.95	48.22	41.76	44.50	47.50
Run 14	40.54	39.30	42.95	41.35	35.43	46.39	38.88	42.95	43.34	38.04	37.18	42.95	43.73	44.50	43.34	43.73	41.76	46.02
Run 15	38.04	44.12	38.46	50.35	42.95	39.30	41.35	49.30	41.76	38.46	42.95	44.50	49.30	42.55	43.34	46.02	37.18	46.02
Run 16	50.70	41.76	43.34	41.35	38.46	45.27	45.27	32.72	42.55	42.55	40.95	44.12	38.46	43.34	42.16	41.35	42.55	41.76
Run 17	42.95	42.16	43.73	44.89	41.76	42.55	38.88	37.18	41.76	50.00	38.88	40.95	42.55	40.95	43.34	42.16	37.61	46.02
Run 18	38.04	48.22	39.30	40.95	46.02	44.50	45.27	38.88	42.16	42.55	43.34	42.55	42.16	45.27	44.50	46.76	44.89	42.16
Run 19	40.54	43.34	42.95	49.30	42.95	43.73	46.02	44.12	47.50	42.95	44.50	49.30	40.95	40.13	46.02	44.50	42.55	47.50
Run 20	39.30	39.72	38.46	42.55	43.34	34.99	47.13	36.75	46.76	42.55	43.34	44.50	42.95	44.50	43.34	49.65	44.50	45.64
Run 21	36.31	41.35	42.55	42.55	42.55	44.12	39.72	45.64	39.30	40.54	42.55	36.75	43.34	46.39	42.16	38.88	49.30	39.72
Run 22	41.76	40.13	44.50	42.95	50.35	42.55	42.95	41.76	40.95	36.31	41.35	42.16	43.73	50.35	38.46	38.46	47.13	46.39
Run 23	38.04	40.95	42.55	44.12	44.50	39.30	43.34	38.88	44.50	41.35	48.58	45.27	41.76	43.73	48.22	46.02	47.86	44.50
Run 24	44.89	39.72	41.76	46.02	41.76	45.27	42.55	47.50	39.30	38.46	45.27	46.76	46.76	42.55	44.12	41.76	42.16	42.55
Run 25	40.95	38.88	40.95	45.27	39.72	41.76	38.88	38.46	43.73	42.16	43.34	40.54	50.00	46.39	42.55	44.50	38.46	46.02
Run 26	32.26	44.12	41.35	45.64	45.64	37.18	44.12	41.76	46.39	44.12	44.12	46.39	42.16	40.13	41.76	44.89	42.16	42.95
Run 27	38.46	40.13	41.76	45.64	47.50	43.73	48.22	44.12	40.54	40.13	39.30	41.76	40.54	41.76	41.35	42.95	46.02	43.73
Run 28	37.18	47.50	43.73	40.54	40.13	43.34	41.35	42.55	46.02	42.55	45.64	42.55	42.16	44.89	48.94	39.30	47.13	50.00
Run 29	40.95	42.55	40.54	46.76	41.35	43.73	42.55	40.13	44.89	42.55	45.64	46.02	46.02	44.12	46.39	41.35	45.64	47.50
Run 30	41.76	43.34	39.72	42.55	37.18	41.35	42.55	45.27	41.76	40.95	40.54	44.89	44.89	44.89	42.16	40.54	50.00	47.13
Mean	40.68	41.36	41.94	43.16	42.13	41.54	43.33	42.34	42.78	42.28	43.41	43.57	44.09	43.74	43.37	43.00	43.87	43.84
Std dev	0.0401	0.0279	0.0259	0.0327	0.0341	0.0272	0.0275	0.0403	0.0241	0.0299	0.0281	0.0295	0.0317	0.0260	0.0296	0.0292	0.0308	0.0343

Table B-41: Detailed results for the 30 runs of the cross-validation experiments using the video processing system for optimising the foveal block threshold on the MNI_by_5 template in %. The threshold value tested here was 0.1.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	42.95	41.35	40.13	42.55	46.02	42.55	44.89	48.22	42.55	42.95	46.02	48.58	46.76	43.73	45.64	44.89	50.00	47.50
Run 2	39.72	45.64	38.88	41.76	39.72	45.27	43.34	44.50	40.95	44.12	44.12	38.88	46.02	39.72	43.73	42.95	39.30	46.76
Run 3	47.13	38.04	39.30	43.73	47.50	43.73	46.76	47.86	46.39	50.35	45.27	47.86	43.34	46.39	42.95	46.02	38.88	44.50
Run 4	39.72	40.54	43.73	37.18	37.61	42.95	41.76	47.13	40.95	43.34	48.58	43.34	46.39	34.09	47.13	45.27	43.73	43.34
Run 5	43.73	47.13	42.95	44.89	40.54	45.27	37.61	40.54	45.64	50.00	48.22	44.89	46.39	47.50	46.02	50.70	46.02	44.50
Run 6	38.88	39.30	38.88	45.64	40.95	46.39	48.94	42.55	44.12	44.50	42.55	40.13	46.39	50.00	51.04	48.58	46.39	43.34
Run 7	38.88	36.31	39.30	41.35	45.27	47.86	46.02	44.50	39.30	47.50	42.95	46.02	42.55	53.40	46.02	45.27	44.89	42.55
Run 8	41.35	38.04	42.16	40.13	42.16	45.64	40.13	38.04	40.54	40.54	44.12	41.76	40.13	48.58	47.13	45.64	48.22	44.50
Run 9	42.95	40.95	44.50	43.34	42.55	41.35	39.30	43.73	42.55	50.00	47.13	44.50	46.76	44.12	44.12	44.89	52.73	47.13
Run 10	44.50	40.54	36.75	37.61	41.76	43.34	43.34	40.13	44.50	47.86	50.00	52.06	47.13	45.64	45.64	42.95	43.34	44.12
Run 11	43.34	38.88	40.13	40.13	42.16	46.39	43.34	51.38	42.95	49.30	47.50	46.39	47.13	47.50	43.73	42.16	44.50	45.27
Run 12	39.72	40.95	39.72	39.72	39.72	40.13	47.50	40.95	45.64	44.50	45.64	44.12	45.64	40.54	44.50	44.89	46.76	50.00
Run 13	46.76	40.95	41.76	44.12	44.50	39.30	44.12	47.13	46.39	46.39	44.12	43.73	50.70	45.27	38.04	45.64	46.76	40.54
Run 14	44.89	39.30	46.39	39.72	39.30	45.27	46.76	47.86	45.27	44.50	45.27	52.73	47.50	46.39	44.50	47.13	48.22	44.89
Run 15	46.02	40.54	46.02	38.46	38.04	36.75	40.95	40.13	44.50	47.86	46.76	38.04	44.50	50.35	50.70	46.02	48.58	49.30
Run 16	45.27	41.35	42.55	41.35	43.73	47.86	42.95	42.16	47.50	46.76	44.50	41.35	43.34	50.35	46.39	40.13	46.76	46.02
Run 17	39.30	44.50	46.39	40.13	38.04	44.89	44.12	42.16	40.54	49.65	38.46	48.22	45.27	49.65	45.64	42.95	42.95	42.16
Run 18	38.04	44.89	40.54	44.12	41.35	37.61	39.72	42.55	46.02	42.95	53.07	44.12	45.27	50.70	42.55	40.54	48.58	42.55
Run 19	43.34	42.55	41.35	40.54	44.12	44.50	47.13	41.76	44.12	48.58	40.54	43.34	46.76	49.30	48.22	46.76	43.34	45.64
Run 20	39.72	40.13	38.88	43.34	42.16	47.13	40.95	48.58	41.35	48.58	49.65	46.02	37.61	42.16	48.22	43.34	48.22	48.58
Run 21	41.76	38.88	47.13	44.89	44.12	45.27	40.54	44.50	46.02	46.02	44.12	39.72	45.64	50.35	45.64	46.02	44.50	44.89
Run 22	40.13	39.30	41.76	41.35	45.64	37.61	39.72	40.95	39.30	46.02	44.50	47.86	47.50	44.12	48.94	44.50	44.50	46.02
Run 23	38.04	43.73	42.16	44.12	44.50	44.89	42.55	44.12	45.27	40.13	47.86	42.16	50.70	41.76	47.86	42.55	45.64	50.70
Run 24	37.61	43.34	37.61	40.13	44.12	42.55	44.12	41.35	42.16	41.35	45.64	46.76	41.76	44.50	48.58	44.12	46.02	43.34
Run 25	39.72	34.99	40.13	41.76	42.55	38.04	43.73	42.95	40.95	44.89	44.12	47.50	46.39	47.13	46.39	41.76	46.39	42.16
Run 26	46.39	39.30	41.76	34.99	42.16	44.50	47.86	40.13	41.76	43.73	43.73	42.95	46.76	48.58	44.12	45.27	47.50	48.22
Run 27	37.18	45.64	37.61	46.02	42.16	43.34	46.76	47.13	46.02	47.13	44.12	48.58	46.39	46.76	47.13	37.18	48.58	44.12
Run 28	40.13	42.95	42.95	42.55	38.04	44.89	45.64	43.73	43.73	45.27	47.86	47.86	44.89	42.55	43.34	44.12	43.34	44.50
Run 29	38.88	38.88	32.72	41.76	39.30	42.16	44.50	42.95	40.54	40.54	45.64	46.76	47.86	44.12	47.13	40.54	52.06	46.39
Run 30	41.76	42.95	36.31	45.64	41.35	44.89	41.76	40.13	41.76	44.12	44.12	44.50	50.70	40.95	47.13	45.64	42.16	43.34
Mean	41.59	41.06	41.02	41.77	42.04	43.41	43.56	43.66	43.31	45.65	45.54	45.02	45.81	45.87	45.94	44.28	45.96	45.23
Std dev	0.0293	0.0279	0.0321	0.0265	0.0255	0.0304	0.0286	0.0317	0.0234	0.0293	0.0282	0.0351	0.0278	0.0405	0.0256	0.0264	0.0312	0.0243

Table B-42: Detailed results for the 30 runs of the cross-validation experiments using the video processing system for optimising the foveal block threshold on the MNI_by_5 template in %. The threshold value tested here was 0.2.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	39.72	40.54	38.04	40.54	44.12	46.02	47.50	40.95	43.73	40.54	47.86	43.73	45.64	43.34	43.73	43.73	45.27	40.95
Run 2	39.72	42.16	38.88	47.50	47.86	39.72	45.64	42.95	46.39	47.86	42.95	42.16	43.73	38.88	44.89	40.13	39.72	40.54
Run 3	40.54	40.54	43.73	39.72	46.02	46.02	41.35	46.02	46.76	46.02	46.02	42.55	47.50	42.95	39.72	39.30	36.75	40.95
Run 4	41.76	41.76	38.88	44.50	41.35	46.76	47.86	43.34	41.76	48.94	45.27	44.89	50.00	43.73	46.76	43.73	37.61	40.95
Run 5	40.13	42.55	43.34	43.73	42.55	44.50	44.89	45.64	47.13	39.30	47.86	41.76	38.46	44.89	46.76	40.95	41.35	39.72
Run 6	39.30	42.95	46.02	39.30	40.54	43.34	42.95	44.12	45.27	44.12	43.34	48.58	44.12	40.54	44.12	44.12	42.55	42.95
Run 7	40.54	43.73	46.39	40.95	37.18	42.55	39.72	47.86	44.12	47.86	40.13	44.50	47.13	42.55	42.55	43.34	42.95	39.30
Run 8	45.27	46.76	44.89	41.76	47.50	47.50	40.13	40.54	41.76	41.76	37.18	49.65	45.64	42.55	46.76	44.12	44.89	45.27
Run 9	41.35	38.88	39.30	43.73	47.86	45.64	43.34	46.76	42.95	40.13	46.02	46.02	42.16	41.35	41.76	43.73	40.13	39.72
Run 10	43.34	46.39	46.76	42.16	48.94	49.65	40.13	44.50	42.55	44.89	42.95	47.13	46.39	50.00	42.95	43.34	44.12	42.95
Run 11	39.30	46.76	40.13	42.95	46.02	39.30	47.50	44.12	42.95	40.95	42.95	44.12	44.50	45.64	42.55	46.39	45.64	44.50
Run 12	42.55	46.02	45.27	47.86	44.50	47.13	50.35	41.76	41.35	42.55	42.95	50.35	41.35	42.16	39.30	43.34	41.35	45.27
Run 13	35.43	44.50	39.30	39.72	41.76	48.58	45.64	40.13	48.22	43.34	45.64	43.34	44.12	42.95	40.95	38.46	46.76	40.95
Run 14	46.76	45.27	40.13	44.89	42.95	38.88	42.16	44.89	42.16	44.50	46.76	47.86	42.95	44.89	45.64	36.31	41.35	40.54
Run 15	38.46	38.04	35.87	44.50	47.86	43.34	47.50	44.12	48.22	46.39	44.12	38.04	45.27	45.64	46.02	45.27	41.35	46.39
Run 16	41.35	44.12	44.50	43.34	46.76	49.65	46.39	50.00	46.02	40.95	43.34	44.89	40.95	45.64	37.61	43.73	44.50	44.12
Run 17	39.72	45.27	42.55	50.70	44.89	45.27	41.76	48.94	41.76	44.50	50.00	42.95	42.16	42.16	43.73	41.35	41.76	39.72
Run 18	45.64	40.95	43.34	45.27	44.50	42.55	43.73	49.65	44.12	42.95	43.34	38.04	42.95	44.12	39.30	40.95	42.95	35.87
Run 19	42.16	47.13	44.89	40.54	47.86	47.86	47.13	44.12	45.64	43.34	43.73	42.95	49.30	44.89	44.50	43.73	40.95	43.73
Run 20	34.54	40.54	41.76	41.35	45.27	38.04	42.16	44.12	43.34	38.88	47.13	45.64	40.54	45.27	42.16	37.61	36.75	40.95
Run 21	38.04	42.95	45.27	41.35	40.95	45.64	50.00	47.86	41.76	46.76	42.55	44.12	38.88	43.73	46.76	33.18	40.54	49.30
Run 22	43.34	44.12	42.16	39.72	47.13	38.88	45.64	44.50	40.95	42.55	42.55	37.61	40.13	47.50	47.86	50.00	42.95	40.13
Run 23	38.46	38.46	44.50	38.88	43.34	44.89	47.86	42.16	42.16	46.76	39.30	44.12	42.55	45.64	40.13	41.35	38.88	46.02
Run 24	42.55	39.30	48.22	38.46	45.64	47.50	44.89	42.55	47.50	38.46	37.61	45.27	39.72	43.73	45.27	49.30	45.64	43.34
Run 25	44.89	44.12	39.30	42.95	46.02	44.89	45.64	41.35	44.89	46.02	45.27	42.95	40.54	42.95	40.95	40.95	36.75	42.16
Run 26	38.88	47.13	39.72	43.34	35.43	42.16	46.76	50.70	38.46	44.89	42.16	46.39	44.89	42.55	45.64	36.31	45.27	43.34
Run 27	35.87	41.76	46.02	44.50	51.38	39.72	49.30	41.76	43.34	44.89	39.30	40.54	42.95	50.00	38.88	50.35	39.30	36.75
Run 28	36.31	43.34	44.89	45.64	43.73	42.95	42.95	46.76	42.55	47.86	41.76	44.12	41.35	40.95	47.50	44.89	47.86	41.35
Run 29	42.16	44.12	42.55	46.39	47.50	44.12	46.02	42.55	46.76	44.89	45.64	35.43	45.27	47.13	44.50	45.64	47.86	40.54
Run 30	43.73	42.55	46.76	37.18	49.30	38.88	44.89	45.27	43.73	44.12	37.61	43.73	43.34	42.55	43.73	44.50	45.27	40.13
Mean	40.73	43.09	42.78	42.78	44.89	44.06	45.06	44.67	43.94	43.90	43.44	43.78	43.48	44.03	43.43	42.67	42.30	41.95
Std dev	0.0301	0.0260	0.0311	0.0305	0.0346	0.0337	0.0287	0.0284	0.0237	0.0281	0.0314	0.0339	0.0285	0.0247	0.0281	0.0387	0.0317	0.0281

Table B-43: Detailed results for the 30 runs of the cross-validation experiments using the video processing system for optimising the foveal block threshold on the MNI_by_5 template in %. The threshold value tested here was 0.3.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	43.73	45.64	43.73	48.94	47.50	47.50	47.86	41.76	42.55	43.73	47.50	49.65	52.40	46.02	41.76	47.86	46.39	51.04
Run 2	44.50	50.70	49.30	48.22	47.50	53.40	43.73	55.34	51.38	47.50	49.65	42.16	48.58	50.00	45.27	44.89	45.27	47.13
Run 3	46.39	41.76	46.76	45.64	52.40	48.58	51.04	46.02	47.50	45.64	46.02	49.65	42.95	44.12	56.29	44.50	47.13	45.27
Run 4	43.34	45.27	47.13	45.27	48.22	44.12	47.50	42.16	48.94	44.50	44.12	52.40	50.70	45.64	46.39	44.50	45.27	46.76
Run 5	39.72	43.73	40.54	46.39	47.13	45.27	46.39	43.73	40.13	43.34	49.65	46.39	49.30	51.38	47.86	48.94	41.35	46.76
Run 6	40.54	43.73	45.64	48.22	47.50	48.58	44.89	45.27	47.13	44.50	45.27	41.76	44.50	43.73	41.76	45.27	44.89	45.64
Run 7	45.64	43.34	48.58	54.38	46.02	46.39	46.02	47.86	47.86	48.58	46.02	50.35	43.73	46.02	50.35	49.65	47.50	46.02
Run 8	39.72	43.73	48.22	43.34	46.76	42.95	51.38	46.02	52.06	44.50	41.35	46.76	46.02	44.89	52.06	42.16	48.94	43.73
Run 9	46.02	54.05	48.58	47.13	46.39	46.39	50.35	41.76	46.76	44.89	50.00	44.89	46.39	47.13	45.64	44.12	45.64	48.58
Run 10	48.22	48.94	50.70	50.35	46.76	46.76	43.73	46.39	45.64	42.95	44.50	42.55	49.30	52.73	50.35	47.50	46.76	46.02
Run 11	47.50	45.64	50.35	45.64	48.58	40.95	46.02	46.39	44.12	43.73	46.76	48.58	46.39	49.65	38.46	51.04	42.95	46.02
Run 12	44.50	41.35	46.76	43.34	46.76	48.22	48.22	47.13	45.64	46.39	46.76	47.86	49.30	47.13	50.35	48.58	50.70	43.73
Run 13	40.54	50.00	45.64	48.22	42.95	50.70	47.50	42.55	41.35	47.13	51.38	46.02	49.30	50.35	50.35	43.73	42.95	45.64
Run 14	42.16	47.50	49.30	40.95	50.35	42.95	44.50	45.27	44.89	44.50	48.58	46.39	49.30	50.70	44.50	51.72	44.89	45.64
Run 15	39.72	50.70	47.13	48.22	47.86	47.50	46.02	46.76	44.12	49.65	51.72	38.88	48.22	46.02	50.70	48.94	47.50	47.86
Run 16	38.46	46.39	42.95	46.39	50.70	46.39	46.02	48.22	45.27	54.70	46.39	49.30	49.30	51.72	44.89	48.58	50.00	49.65
Run 17	40.95	42.95	44.50	48.58	42.95	49.65	47.50	46.76	44.12	46.02	47.86	49.30	52.40	48.58	51.04	55.34	43.73	54.05
Run 18	39.30	44.12	45.27	49.30	48.22	44.50	50.00	49.30	42.95	41.35	48.58	42.55	47.50	52.06	44.50	46.02	42.95	47.13
Run 19	45.64	42.16	46.76	41.76	40.13	48.58	46.76	46.39	46.02	40.54	47.13	52.73	46.39	48.94	49.30	51.04	41.35	42.55
Run 20	44.12	47.13	46.39	48.58	46.02	44.50	45.27	44.12	50.35	49.30	47.50	47.86	48.58	48.58	42.55	44.89	46.76	47.13
Run 21	45.64	40.54	49.65	45.27	47.13	46.02	50.70	43.34	41.35	46.39	47.13	46.76	49.65	44.50	43.34	47.13	52.06	46.76
Run 22	40.54	45.64	45.64	42.95	45.27	46.39	49.65	51.04	49.30	46.76	48.58	53.07	51.38	47.50	45.27	44.50	48.22	48.22
Run 23	41.35	40.54	47.13	41.76	47.50	44.89	46.39	44.50	43.34	43.73	45.64	49.65	47.50	53.40	49.65	47.50	48.22	48.22
Run 24	43.34	43.34	50.70	47.86	49.30	44.12	45.27	46.76	47.86	49.65	52.40	47.86	46.39	46.39	48.58	46.76	45.27	45.27
Run 25	47.50	51.72	46.02	48.22	43.34	48.58	47.50	48.22	52.06	44.89	51.38	44.12	53.73	44.12	51.38	42.55	46.76	46.76
Run 26	45.27	40.95	44.50	45.27	48.58	43.73	51.72	44.50	42.55	45.64	51.72	48.22	48.22	49.30	47.13	49.30	44.50	50.35
Run 27	47.13	45.64	46.76	48.58	44.89	45.64	46.02	50.70	45.64	43.73	47.50	49.65	47.13	41.35	48.58	47.86	49.65	48.58
Run 28	42.16	47.50	47.13	45.64	44.12	48.58	46.02	51.04	49.65	46.39	49.65	50.00	47.50	46.39	43.34	47.86	49.30	49.65
Run 29	44.50	44.89	39.30	43.34	44.50	50.00	46.39	48.94	49.30	52.40	44.50	46.76	53.73	48.58	47.50	51.72	47.13	51.72
Run 30	41.76	49.30	46.39	39.72	47.86	49.30	51.04	48.58	47.50	45.27	46.02	47.50	47.50	47.86	46.39	46.39	39.30	48.94
Mean	43.33	45.63	46.58	46.25	46.77	46.70	47.38	46.56	46.24	45.94	47.71	47.32	48.44	47.83	47.18	47.36	46.11	47.36
Std dev	0.0278	0.0348	0.0264	0.0311	0.0249	0.0264	0.0230	0.0304	0.0320	0.0298	0.0256	0.0333	0.0260	0.0291	0.0376	0.0297	0.0291	0.0244

TABLE B-44: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM FOR OPTIMISING THE FOVEAL BLOCK THRESHOLD ON THE MNI_BY_5 TEMPLATE IN %. THE THRESHOLD VALUE TESTED HERE WAS 0.4.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	44.12	48.58	45.64	49.65	46.02	44.50	47.50	49.30	46.02	50.70	48.58	49.30	47.86	48.94	47.86	48.22	50.35	47.50
Run 2	43.73	46.76	46.39	47.13	50.35	51.04	45.64	48.22	51.04	49.30	48.94	51.72	48.94	46.39	50.70	54.70	46.39	55.66
Run 3	44.89	48.58	48.94	47.50	53.40	50.35	43.73	50.35	46.02	46.39	50.00	48.58	48.94	50.00	47.13	52.40	48.58	53.40
Run 4	44.50	44.89	46.76	45.27	50.35	48.94	47.86	54.38	53.07	50.70	50.00	52.40	50.00	50.00	47.86	46.76	50.00	49.30
Run 5	41.35	46.02	45.64	47.86	47.50	48.58	48.58	51.38	52.40	48.58	48.94	46.02	48.94	49.65	49.65	51.72	45.64	46.02
Run 6	43.73	44.12	47.86	47.86	44.89	48.94	50.35	48.22	47.50	42.95	50.00	47.50	52.06	46.39	49.65	47.13	40.54	46.02
Run 7	40.95	50.00	48.58	44.12	44.50	49.30	49.65	52.06	46.02	51.72	46.76	49.30	51.04	51.04	52.73	48.94	51.72	47.50
Run 8	44.12	46.39	45.64	42.55	52.06	50.00	46.76	47.86	48.58	43.73	47.50	51.04	48.94	51.72	52.73	50.00	48.22	50.70
Run 9	44.50	50.00	42.16	47.50	45.64	49.30	46.39	50.35	49.30	47.86	39.30	50.00	48.94	50.70	47.50	48.94	46.02	44.89
Run 10	47.50	50.35	45.64	45.27	44.50	48.94	53.40	48.94	48.58	48.94	50.00	45.27	50.35	47.86	47.86	50.00	50.70	54.05
Run 11	46.76	45.64	47.50	47.86	44.89	49.30	49.30	54.70	42.16	51.72	52.06	49.30	50.70	51.72	53.07	47.86	47.13	52.06
Run 12	46.39	49.30	44.89	35.43	52.73	45.64	50.70	45.27	40.54	50.35	48.94	48.94	51.04	50.35	46.02	44.89	48.22	50.35
Run 13	51.72	49.65	48.94	44.12	46.76	44.50	54.38	47.86	52.73	44.89	47.50	51.04	48.94	48.94	47.50	51.38	47.50	51.04
Run 14	50.00	51.04	47.86	45.27	50.00	52.06	55.02	48.94	51.04	44.89	40.13	52.40	48.22	48.22	46.76	45.27	47.50	51.04
Run 15	50.70	48.22	47.13	49.65	44.89	46.76	50.70	49.30	46.76	49.30	50.70	46.76	49.30	49.30	48.22	46.02	44.89	50.35
Run 16	46.76	48.22	46.02	38.88	47.86	48.58	51.04	54.05	50.00	46.02	52.06	50.00	47.13	47.13	52.40	49.65	45.64	46.76
Run 17	45.27	50.00	44.89	50.70	46.39	48.22	43.34	49.30	53.40	49.30	47.50	47.13	47.13	44.50	49.65	40.95	55.02	49.65
Run 18	42.95	46.02	47.13	43.34	53.73	50.35	47.86	47.86	52.40	44.50	49.30	46.02	46.76	48.94	49.30	50.00	46.39	50.00
Run 19	45.64	47.86	44.12	41.35	49.30	44.12	49.65	52.73	52.06	53.07	48.94	48.58	48.22	50.00	48.22	48.94	46.39	49.65
Run 20	40.54	42.95	46.02	45.27	49.30	51.38	47.86	48.22	47.50	42.16	53.73	48.58	49.30	48.22	51.04	48.58	48.22	52.40
Run 21	48.22	42.55	52.40	45.64	52.73	50.70	49.65	47.86	50.35	44.12	51.04	43.73	52.40	51.04	48.22	51.72	44.50	51.38
Run 22	49.65	46.39	42.55	44.50	48.58	45.64	45.27	48.94	52.06	49.30	49.65	48.58	50.35	52.73	46.02	51.04	50.35	40.54
Run 23	49.30	41.76	51.04	55.66	47.13	49.65	53.73	52.40	51.72	51.04	55.34	54.05	49.30	50.70	47.13	48.94	47.86	48.58
Run 24	46.76	44.50	46.76	44.12	46.39	52.06	42.55	47.13	51.04	45.27	51.04	46.39	46.02	45.64	49.65	48.58	46.76	48.22
Run 25	47.50	46.76	45.64	46.76	47.50	52.06	48.22	48.94	51.04	43.34	49.65	50.70	49.30	51.04	45.64	51.04	45.27	46.39
Run 26	47.86	43.34	48.22	45.64	53.07	51.72	43.73	50.00	46.76	50.35	48.94	44.89	51.72	52.40	51.72	49.30	48.22	46.76
Run 27	44.89	46.02	50.70	49.65	48.58	49.30	46.02	53.40	50.70	45.64	51.72	44.50	52.06	48.58	49.65	53.07	47.86	48.22
Run 28	48.22	42.55	50.35	45.64	49.65	53.07	56.91	51.04	50.70	48.94	48.22	52.06	44.12	48.58	50.70	44.89	48.22	50.70
Run 29	42.55	47.86	47.86	44.89	47.13	48.94	45.64	49.30	47.13	44.12	47.86	53.73	48.58	44.89	47.86	46.39	51.04	50.00
Run 30	47.86	48.22	47.86	50.00	44.50	46.02	46.02	50.35	44.12	46.02	52.40	53.40	45.27	48.94	47.50	47.50	48.22	44.50
Mean	45.96	46.82	47.04	45.97	48.34	49.00	48.58	49.96	49.09	47.51	49.22	49.06	49.06	49.15	49.00	48.83	47.78	49.12
Std dev	0.0284	0.0257	0.0229	0.0369	0.0289	0.0239	0.0356	0.0227	0.0322	0.0303	0.0317	0.0280	0.0196	0.0209	0.0208	0.0279	0.0261	0.0308

Table B-45: Detailed results for the 30 runs of the cross-validation experiments using the video processing system for optimising the foveal block threshold on the MNI_by_5 template in %. The threshold value tested here was 0.5.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	44.50	48.94	50.00	51.04	51.38	46.39	50.00	49.65	44.89	49.30	51.72	51.04	50.70	50.00	53.07	50.00	47.50	48.58
Run 2	43.73	47.13	48.94	45.64	52.06	48.94	51.04	52.73	53.40	50.00	47.50	49.30	56.29	54.05	55.34	48.22	51.72	53.73
Run 3	45.64	49.30	50.70	51.04	45.27	49.30	52.06	53.07	47.13	54.70	49.65	50.35	44.89	50.35	48.58	52.73	50.35	48.94
Run 4	53.40	53.40	47.86	44.89	47.13	53.40	45.64	50.00	46.76	47.86	49.30	50.35	48.58	49.30	53.40	47.86	43.73	42.55
Run 5	50.35	48.22	45.27	46.02	48.58	52.73	52.40	47.50	48.94	52.06	49.30	52.73	48.58	49.30	48.94	47.13	50.70	52.40
Run 6	48.94	48.94	43.34	46.39	47.50	48.58	48.58	50.70	49.30	50.70	48.58	51.04	50.00	50.70	47.13	53.40	46.02	50.00
Run 7	46.76	43.73	44.12	51.04	47.13	44.89	47.50	48.58	48.22	43.73	50.70	51.04	49.30	52.40	54.05	53.73	55.02	46.39
Run 8	45.64	44.89	48.94	50.70	48.94	48.22	47.86	52.73	46.76	51.04	46.76	50.70	51.04	54.70	51.38	50.70	50.35	52.06
Run 9	47.50	44.89	48.94	46.76	50.70	50.00	50.00	48.94	54.05	47.13	44.89	53.40	47.13	53.07	48.58	55.34	47.86	47.50
Run 10	48.22	50.70	51.72	48.94	53.40	44.50	49.65	50.00	50.00	50.35	52.40	52.06	49.65	55.02	50.00	54.05	48.94	44.89
Run 11	45.64	47.50	48.58	50.70	46.02	47.50	46.76	48.58	49.65	49.65	51.38	53.40	50.35	51.04	48.94	46.39	44.89	46.02
Run 12	46.02	46.39	53.07	50.70	52.73	45.64	46.76	48.94	48.22	49.30	55.98	53.73	46.76	55.66	46.39	46.76	51.04	50.00
Run 13	48.58	42.55	44.89	47.50	55.02	49.30	50.70	53.07	48.94	52.40	42.16	52.40	49.30	51.72	46.39	47.86	46.39	53.73
Run 14	47.13	46.76	47.86	47.86	44.50	45.27	51.04	51.04	49.65	50.35	48.58	49.65	56.29	48.22	51.04	53.40	51.72	53.73
Run 15	47.50	46.39	45.64	49.30	42.95	47.13	48.94	50.00	47.86	45.64	50.00	46.39	49.65	47.50	46.39	50.00	47.86	53.07
Run 16	51.04	49.30	44.50	45.64	49.65	51.38	48.22	52.06	47.13	47.86	47.13	57.22	51.04	47.13	50.00	48.94	49.65	49.30
Run 17	47.86	49.65	47.13	45.27	47.50	50.70	52.06	52.06	43.34	48.58	49.65	52.40	48.22	51.04	50.35	44.12	50.70	51.04
Run 18	48.94	45.64	48.94	47.13	48.94	48.58	48.22	46.39	42.55	52.40	43.73	46.39	50.35	48.58	48.94	55.34	52.73	47.86
Run 19	47.13	46.76	47.13	48.94	53.73	50.00	50.00	49.30	45.64	53.07	49.65	49.30	48.22	49.30	48.22	55.02	48.58	48.94
Run 20	45.64	48.22	43.73	43.73	47.13	50.00	44.50	48.22	53.40	47.13	49.30	50.35	48.58	53.73	49.30	44.12	53.07	52.40
Run 21	46.76	51.04	48.58	49.30	46.02	46.02	50.00	47.86	43.73	43.34	40.13	50.00	46.39	53.73	48.94	43.73	46.39	48.94
Run 22	42.55	46.76	47.50	48.94	51.04	54.70	48.58	52.40	46.02	44.89	46.76	48.22	48.22	51.72	52.06	50.35	47.13	51.38
Run 23	47.13	48.94	50.35	48.22	51.04	46.76	43.73	47.86	48.58	57.22	42.16	50.70	51.04	48.58	52.40	46.76	52.40	54.05
Run 24	47.13	49.30	47.50	44.89	46.76	40.54	51.04	52.40	51.38	43.73	48.58	46.02	52.73	53.73	47.86	53.40	48.58	47.50
Run 25	49.65	45.27	50.70	46.76	46.76	46.02	44.12	44.50	46.76	49.30	49.30	46.76	48.22	51.72	45.27	44.12	46.02	47.86
Run 26	49.65	46.76	51.04	42.16	49.30	53.07	44.50	52.06	53.40	46.02	48.22	47.86	49.30	52.06	46.02	51.38	48.22	47.86
Run 27	44.50	44.89	50.00	46.39	48.94	40.95	52.06	48.22	47.86	47.86	52.40	49.30	46.02	52.40	45.27	50.70	51.04	50.00
Run 28	46.02	52.06	38.46	50.00	47.13	51.04	48.94	47.50	47.86	53.73	50.00	52.06	48.58	51.72	48.58	45.64	46.76	51.38
Run 29	48.58	51.38	46.76	46.76	50.00	47.13	46.76	45.27	53.40	47.86	45.64	47.50	54.38	50.00	54.70	48.58	53.73	48.94
Run 30	50.00	49.65	45.27	47.50	48.22	54.70	43.34	50.35	50.35	51.38	46.02	48.58	51.72	50.70	53.40	49.65	52.73	51.72
Mean	47.41	47.84	47.58	47.67	48.85	48.45	48.50	49.73	48.51	49.29	48.25	50.34	49.72	51.31	49.70	49.65	49.39	49.76
Std dev	0.0227	0.0251	0.0299	0.0232	0.0281	0.0346	0.0263	0.0230	0.0300	0.0329	0.0331	0.0250	0.0262	0.0223	0.0281	0.0353	0.0279	0.0279

TABLE B-46: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM FOR OPTIMISING THE FOVEAL BLOCK THRESHOLD ON THE MNI_BY_5 TEMPLATE IN %. THE THRESHOLD VALUE TESTED HERE WAS 0.6.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	48.94	51.72	50.70	50.35	43.34	49.65	50.70	42.95	47.86	48.22	57.53	50.00	51.04	48.22	50.00	47.50	44.89	48.22
Run 2	43.73	45.64	51.38	51.38	48.22	48.22	48.58	47.13	51.72	52.73	52.40	49.30	48.94	43.34	52.06	54.70	51.38	54.05
Run 3	43.73	48.22	46.39	49.65	46.39	47.13	44.50	47.86	53.40	50.35	54.05	49.65	51.72	49.65	47.86	55.34	41.35	50.70
Run 4	41.35	47.13	47.86	45.64	44.89	47.50	52.73	49.30	48.58	51.04	51.72	55.02	47.50	53.07	44.12	48.94	40.95	46.76
Run 5	47.86	46.39	47.86	47.86	50.35	48.58	45.27	50.70	47.13	51.72	50.70	55.02	48.22	53.07	53.40	44.50	48.94	48.58
Run 6	44.89	45.64	47.13	48.58	43.73	48.94	47.13	44.50	51.04	47.86	49.30	48.22	46.39	48.58	54.70	49.65	42.95	44.12
Run 7	44.50	50.35	53.07	52.73	47.13	50.35	49.65	43.34	48.22	50.00	53.40	53.40	46.76	48.94	48.94	46.76	51.38	47.13
Run 8	50.00	47.13	48.94	47.13	54.38	45.64	51.72	44.50	47.13	46.76	50.00	50.35	48.94	45.27	45.64	46.39	51.72	44.12
Run 9	51.04	46.76	48.22	42.16	48.94	50.35	47.86	50.00	43.73	46.02	54.70	48.22	53.07	45.64	51.04	47.13	49.65	53.73
Run 10	50.00	43.34	47.13	46.76	47.86	48.94	51.72	52.40	49.65	48.94	46.76	47.50	50.70	46.76	48.94	51.72	46.76	51.04
Run 11	45.64	48.22	45.64	46.76	46.76	49.30	48.22	47.86	52.06	49.65	51.04	44.12	48.22	54.05	47.86	51.38	52.06	51.38
Run 12	43.34	46.02	47.50	48.22	42.95	48.58	51.04	41.35	53.73	53.40	52.73	47.86	49.30	54.05	50.35	51.38	47.86	42.95
Run 13	47.13	41.35	41.35	46.76	44.12	51.04	47.86	51.04	51.04	50.70	45.64	56.60	49.65	48.94	43.34	48.58	46.39	55.34
Run 14	47.86	44.50	48.22	47.13	51.38	47.86	44.50	46.39	45.64	45.64	50.70	47.86	49.65	48.58	50.00	53.40	46.02	46.02
Run 15	52.73	47.13	48.58	48.58	50.70	55.34	50.35	50.35	47.13	50.35	47.13	50.35	49.30	52.06	42.55	47.13	49.65	49.30
Run 16	47.13	44.12	47.50	46.39	50.70	50.00	45.64	51.38	50.70	50.35	50.35	51.72	50.35	47.86	47.50	52.73	47.13	44.50
Run 17	52.06	46.02	46.76	42.95	46.02	46.02	46.39	51.04	47.50	49.30	48.58	49.30	50.35	52.40	49.30	48.58	46.02	42.95
Run 18	42.16	43.34	44.12	43.34	47.50	45.64	53.73	48.22	42.95	48.58	51.72	53.07	44.50	41.76	39.30	47.86	51.04	50.00
Run 19	44.12	44.50	46.76	49.65	42.16	55.02	49.30	44.12	48.94	48.58	43.34	52.06	53.73	47.86	48.22	52.06	46.02	47.13
Run 20	46.39	51.04	47.50	48.94	52.73	48.58	47.13	48.22	50.00	52.73	53.40	48.58	49.65	49.65	48.22	51.72	46.76	48.22
Run 21	50.00	46.76	46.39	49.30	50.70	46.76	47.50	52.40	48.58	51.04	52.06	51.72	47.50	51.72	52.06	50.70	51.38	50.35
Run 22	42.16	48.58	48.58	47.13	48.22	46.76	52.40	46.76	47.50	45.27	46.02	56.60	48.58	54.05	45.27	54.38	56.29	48.58
Run 23	46.02	46.76	43.34	50.35	42.16	46.02	47.86	49.30	54.38	53.40	49.30	55.34	44.89	46.76	49.30	51.72	52.06	38.88
Run 24	46.39	51.04	42.55	43.34	54.38	49.65	53.07	44.12	49.65	51.72	51.72	53.40	47.13	42.16	49.30	49.30	46.76	52.06
Run 25	50.00	41.76	49.65	47.86	46.39	50.00	53.40	48.22	49.65	47.50	50.70	49.65	46.76	53.40	48.58	50.00	46.02	50.00
Run 26	48.22	50.00	46.02	46.39	49.30	48.22	48.58	46.02	48.94	50.00	47.86	54.38	51.72	48.94	53.07	49.30	40.95	46.02
Run 27	46.76	48.58	50.35	48.22	44.50	45.64	48.94	45.64	51.38	51.04	50.00	50.70	47.50	44.89	54.05	54.70	46.76	52.06
Run 28	39.30	46.76	48.58	54.05	49.65	47.86	45.64	46.02	42.95	49.65	51.04	47.50	52.73	46.76	52.06	43.73	49.30	53.40
Run 29	46.02	46.39	45.27	51.38	47.13	43.34	46.02	42.55	45.27	45.64	49.30	48.22	45.64	45.64	48.58	47.50	46.39	43.34
Run 30	44.12	46.02	45.27	39.72	49.65	52.73	48.58	51.04	48.22	52.40	51.72	47.13	46.02	46.76	50.35	49.65	52.40	51.04
Mean	46.45	46.71	47.29	47.62	47.74	48.66	48.87	47.49	48.82	49.69	50.50	50.76	48.88	48.70	48.87	49.95	47.91	48.40
Std dev	0.0323	0.0255	0.0248	0.0309	0.0333	0.0259	0.0267	0.0308	0.0289	0.0229	0.0289	0.0310	0.0235	0.0348	0.0344	0.0295	0.0362	0.0384

Table B-47: Detailed results for the 30 runs of the cross-validation experiments using the video processing system for optimising the foveal block threshold on the MNI_by_5 template in %. The threshold value tested here was 0.7.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	46.02	44.89	46.02	48.94	49.30	44.12	47.86	43.73	49.30	51.38	47.50	51.38	48.22	40.13	48.94	46.76	47.13	42.95
Run 2	43.73	46.02	40.95	47.86	52.73	47.13	51.72	51.04	48.94	47.86	46.76	52.40	43.73	42.55	42.95	49.30	48.58	43.73
Run 3	43.73	44.89	50.35	49.65	46.76	48.58	51.04	46.76	52.40	46.76	53.73	46.02	49.30	46.39	45.64	45.27	47.50	41.35
Run 4	40.13	44.50	48.94	45.27	50.70	51.38	46.02	46.76	47.50	46.39	42.95	49.65	47.50	48.58	47.86	46.02	47.50	47.86
Run 5	48.22	49.65	44.12	46.76	47.50	48.22	48.58	48.22	48.58	52.73	49.30	48.94	47.50	46.39	45.27	42.95	41.35	47.50
Run 6	51.72	44.89	45.64	42.16	52.06	44.50	53.40	47.13	49.65	51.72	49.30	49.30	51.04	39.72	42.95	54.38	50.70	46.76
Run 7	44.50	48.58	45.64	51.04	46.39	48.22	48.58	45.27	47.50	43.34	45.27	46.76	50.70	47.13	51.72	40.95	52.40	43.34
Run 8	50.70	39.72	43.34	48.58	42.55	48.58	49.65	45.27	49.65	47.50	50.70	45.27	46.39	49.30	48.58	48.94	42.16	47.50
Run 9	51.04	46.39	50.00	47.86	46.76	48.58	48.58	46.39	42.95	47.86	43.73	47.86	52.40	45.64	48.22	49.65	46.76	46.76
Run 10	44.12	46.39	44.12	49.30	48.94	48.22	50.70	52.40	45.64	48.58	48.94	48.22	45.64	42.55	43.34	42.55	46.39	50.00
Run 11	44.50	44.50	47.86	51.38	48.22	46.02	44.50	44.89	48.22	46.39	47.50	46.39	43.34	50.00	49.65	42.95	46.02	45.64
Run 12	49.30	50.35	47.86	43.73	43.73	50.70	51.38	54.38	50.70	47.50	48.58	45.64	46.76	48.58	47.13	47.13	42.16	43.34
Run 13	50.00	43.34	46.39	52.06	45.27	50.00	49.65	46.02	44.50	52.06	53.07	47.13	53.73	47.50	43.73	46.39	45.27	46.02
Run 14	48.22	45.27	48.22	47.50	47.86	48.22	52.73	46.76	38.88	49.30	48.22	49.30	46.76	45.27	49.65	46.02	46.39	43.73
Run 15	49.30	44.89	46.76	54.05	50.70	53.40	51.72	49.65	50.35	44.50	48.94	50.35	47.86	48.94	51.04	51.04	49.30	44.12
Run 16	53.73	45.27	49.30	51.38	41.76	53.40	47.86	50.00	49.30	52.73	50.35	44.50	46.02	49.30	51.04	50.00	49.65	44.89
Run 17	50.35	42.95	46.02	49.65	43.73	45.27	47.50	46.02	50.00	47.86	43.34	43.34	48.58	48.22	42.55	41.76	51.72	44.50
Run 18	53.07	46.76	44.89	43.73	52.73	45.27	45.64	51.72	52.40	46.39	47.13	52.73	46.02	44.89	47.86	46.02	46.76	45.27
Run 19	43.73	47.50	48.22	44.89	45.27	50.00	47.50	46.76	47.50	47.13	47.86	47.50	50.00	46.76	42.16	41.76	47.86	44.12
Run 20	44.50	47.13	42.16	48.22	52.40	48.94	50.70	49.30	47.13	44.50	46.39	50.00	43.34	47.50	50.70	46.39	44.12	45.27
Run 21	46.02	50.35	51.72	48.94	48.22	46.02	40.13	49.30	54.05	46.02	46.39	48.58	48.94	50.70	43.73	49.65	47.13	45.27
Run 22	44.89	48.94	51.04	51.38	50.35	50.35	38.88	43.34	51.72	47.86	47.13	51.04	50.00	52.06	48.94	42.16	45.27	42.16
Run 23	45.64	51.04	49.65	50.70	46.39	47.50	50.00	51.72	51.04	50.35	48.58	47.50	41.35	42.55	42.95	48.58	48.58	41.76
Run 24	44.12	46.76	51.04	46.02	50.35	46.39	49.65	45.64	49.65	44.50	44.89	47.50	51.04	46.76	47.13	52.06	45.27	45.64
Run 25	52.73	46.39	47.13	46.02	52.40	49.30	48.58	46.39	50.00	46.76	45.64	42.95	48.58	49.30	48.94	47.13	45.27	47.13
Run 26	48.94	45.27	47.50	51.04	51.04	54.70	48.58	51.04	48.22	50.00	48.58	50.00	50.00	43.34	49.30	47.13	46.39	45.27
Run 27	46.76	44.89	45.64	49.30	47.50	48.22	41.76	52.73	49.65	48.94	44.50	48.22	46.02	49.65	46.76	42.16	43.73	46.76
Run 28	43.34	47.86	51.72	45.64	45.64	48.58	54.70	43.73	49.65	51.04	47.86	45.27	51.04	53.07	47.86	48.58	51.38	46.39
Run 29	45.64	48.22	46.02	47.13	39.72	50.35	47.13	47.50	47.50	48.58	47.13	50.00	49.30	45.64	45.64	49.30	46.39	43.73
Run 30	47.86	44.89	49.65	42.95	50.00	46.39	46.76	52.73	43.34	44.50	50.35	46.76	48.22	51.38	45.64	47.50	39.30	47.50
Mean	47.22	46.28	47.26	48.10	47.90	48.55	48.38	48.09	48.53	48.04	47.69	48.02	47.98	46.99	46.93	46.68	46.61	45.21
Std dev	0.0338	0.0240	0.0275	0.0291	0.0339	0.0253	0.0355	0.0300	0.0305	0.0255	0.0252	0.0244	0.0279	0.0328	0.0284	0.0334	0.0300	0.0197

TABLE B-48: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM FOR OPTIMISING THE FOVEAL BLOCK THRESHOLD ON THE MNI_BY_5 TEMPLATE IN %. THE THRESHOLD VALUE TESTED HERE WAS 0.8.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	47.50	46.76	45.27	43.73	45.64	43.34	44.89	44.89	42.55	44.50	45.27	45.27	48.22	45.27	48.58	53.73	45.64	46.76
Run 2	46.76	47.13	44.50	50.70	45.64	50.70	48.58	44.89	44.89	42.95	44.12	45.27	46.76	45.27	46.76	47.86	44.50	46.76
Run 3	46.76	51.38	45.27	50.35	44.12	45.27	42.95	47.86	46.76	44.89	42.95	45.64	44.12	42.55	45.27	47.50	42.55	45.64
Run 4	54.05	46.76	48.94	45.64	46.02	44.50	45.27	50.00	45.27	44.89	44.50	43.73	48.58	49.65	51.72	41.35	46.76	48.22
Run 5	46.02	46.39	44.12	52.73	44.50	45.64	43.34	43.34	47.86	44.12	48.58	46.02	44.12	46.39	41.35	51.38	49.30	52.40
Run 6	41.76	45.27	47.50	48.22	50.35	48.94	44.12	46.76	45.64	42.95	46.39	46.39	47.13	47.50	48.94	43.73	47.13	48.22
Run 7	48.58	52.06	47.50	54.38	51.38	40.95	40.95	46.02	48.94	47.13	43.73	47.86	40.95	44.50	46.02	44.12	47.13	44.12
Run 8	43.73	46.76	40.54	44.50	48.22	50.35	39.72	52.73	46.02	44.89	43.73	50.70	49.30	51.38	44.89	46.02	47.50	47.86
Run 9	55.66	51.38	43.34	50.70	48.58	47.50	42.16	47.86	44.89	48.94	51.38	46.02	45.27	50.00	48.58	48.94	48.94	47.86
Run 10	41.76	47.86	46.39	48.58	43.34	46.76	45.64	47.86	43.34	45.64	46.02	47.50	40.95	43.73	42.95	45.64	48.94	45.64
Run 11	41.35	48.94	51.38	38.46	48.94	40.54	47.50	48.58	49.30	48.22	49.65	48.58	45.64	47.86	49.30	43.34	47.50	45.27
Run 12	50.70	43.73	46.76	47.50	48.58	46.76	43.73	43.73	51.38	47.86	41.35	47.13	44.89	43.73	41.76	48.94	42.95	42.95
Run 13	48.22	51.04	44.50	48.58	47.86	43.34	47.86	42.95	45.64	47.50	40.95	49.30	49.30	52.06	47.50	52.40	46.76	50.35
Run 14	42.95	44.50	45.27	44.50	47.50	46.02	40.13	46.39	51.38	47.13	45.27	44.50	47.50	43.73	47.13	49.30	47.50	48.58
Run 15	42.16	44.89	49.30	40.95	45.64	50.35	47.86	47.50	47.50	49.65	50.35	44.50	49.65	46.39	53.40	53.07	41.76	49.30
Run 16	47.13	47.86	45.64	45.64	51.04	46.76	44.89	40.13	47.86	45.64	41.76	44.50	47.13	45.27	50.35	43.34	48.22	47.86
Run 17	47.13	45.64	47.13	46.76	47.86	48.58	46.76	46.02	46.76	47.50	44.89	47.50	45.64	43.73	46.02	44.89	44.12	52.06
Run 18	40.54	43.34	42.16	47.50	45.64	46.02	46.02	47.13	44.12	45.27	43.73	44.89	46.76	42.16	42.95	47.50	42.16	48.22
Run 19	48.22	42.55	43.73	47.13	44.89	43.34	44.12	42.95	46.02	45.27	44.12	46.39	49.65	50.35	48.22	46.76	44.89	49.30
Run 20	44.89	51.38	46.02	44.89	50.35	50.00	44.12	49.30	39.72	46.76	47.50	47.50	44.50	43.73	48.94	46.39	48.94	43.34
Run 21	47.86	47.50	42.16	40.95	47.50	44.89	42.95	44.50	48.22	48.58	41.76	49.30	49.30	47.86	43.73	50.35	47.50	47.50
Run 22	47.13	42.16	42.55	44.89	47.86	47.50	51.38	44.12	43.34	45.27	42.95	41.76	48.58	44.50	46.39	46.76	45.27	45.64
Run 23	41.76	47.13	50.70	46.39	47.50	44.50	46.39	42.55	47.50	43.73	46.39	46.02	47.13	42.95	47.86	48.58	46.76	46.76
Run 24	48.58	41.76	47.50	48.22	44.12	47.13	47.50	44.89	44.12	47.50	45.27	47.50	47.13	51.04	47.13	45.27	44.50	49.30
Run 25	46.39	48.58	47.50	54.38	51.38	44.12	42.95	45.27	47.50	50.35	47.13	46.76	46.76	47.13	51.38	53.07	45.64	46.02
Run 26	47.86	48.22	47.50	47.86	50.00	48.94	46.39	43.73	42.55	49.30	48.22	46.76	49.30	45.27	50.00	44.12	51.04	46.02
Run 27	52.06	44.12	42.95	49.30	47.50	49.30	45.27	43.34	52.06	44.89	45.64	49.30	46.76	52.06	46.76	47.86	51.04	44.12
Run 28	52.06	53.07	48.94	44.89	40.95	44.12	51.38	44.12	46.39	47.50	49.65	47.50	41.76	45.27	51.04	51.38	45.64	47.50
Run 29	46.39	46.39	50.70	47.50	48.22	40.54	48.22	46.02	39.30	49.30	46.76	43.34	47.13	44.89	45.64	45.27	47.13	44.12
Run 30	47.50	50.00	41.35	45.64	42.16	43.73	47.13	46.39	44.12	45.27	44.50	55.02	45.64	52.73	48.22	48.22	44.12	46.02
Mean	46.78	47.15	45.90	47.05	47.11	46.01	45.34	45.73	46.03	46.44	45.48	46.75	46.52	46.63	47.29	47.57	46.39	47.12
Std dev	0.0370	0.0302	0.0282	0.0358	0.0265	0.0285	0.0282	0.0255	0.0300	0.0201	0.0267	0.0247	0.0239	0.0311	0.0290	0.0320	0.0238	0.0229

Table B-49: Detailed results for the 30 runs of the cross-validation experiments using the video processing system for optimising the foveal block threshold on the MNI_by_5 template in %. The threshold value tested here was 0.9.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	43.34	46.39	47.86	50.70	46.02	48.22	52.06	47.50	49.65	45.27	45.27	53.07	50.70	50.70	50.35	54.05	48.22	51.38
Run 2	41.35	45.27	47.13	45.27	43.73	47.50	47.86	50.70	50.00	47.13	50.70	50.00	50.00	52.40	45.64	50.00	46.76	52.40
Run 3	42.55	45.64	50.35	46.76	45.64	48.58	47.50	42.16	50.35	48.22	49.65	55.02	53.73	51.72	55.98	51.04	53.07	50.00
Run 4	42.55	40.54	50.70	46.76	51.38	47.50	47.50	52.06	47.50	50.35	51.04	49.65	50.70	47.50	49.65	47.50	51.38	54.38
Run 5	47.50	42.55	47.50	47.50	46.02	50.00	48.22	49.30	51.38	53.07	53.73	49.65	47.50	50.00	52.06	45.27	53.07	54.70
Run 6	51.04	43.73	47.86	50.35	44.50	50.35	48.58	49.30	48.22	43.73	52.73	48.94	49.30	53.07	54.38	50.35	48.22	51.38
Run 7	48.22	45.64	46.02	51.72	50.35	48.94	47.86	47.86	52.06	47.13	49.30	47.13	48.58	55.02	51.72	44.89	52.06	46.39
Run 8	42.95	51.38	44.50	43.73	49.30	47.50	50.35	46.76	50.00	53.40	50.35	52.73	49.30	53.40	49.65	47.86	50.70	55.34
Run 9	47.13	45.64	49.65	44.12	49.30	50.70	45.64	46.39	51.38	51.04	52.73	52.73	55.98	51.72	52.73	46.39	45.27	51.38
Run 10	42.16	44.12	47.86	50.00	52.06	51.72	54.70	49.65	47.50	51.04	48.22	52.73	48.58	49.30	52.73	53.73	49.65	51.04
Run 11	44.89	43.34	48.94	47.86	49.30	46.02	48.94	48.22	53.40	48.58	53.40	46.39	52.73	45.27	51.38	50.35	49.30	50.70
Run 12	43.34	50.70	46.76	48.22	50.70	54.70	43.34	53.40	50.00	50.70	48.94	50.35	50.00	48.22	52.06	52.40	56.60	49.65
Run 13	44.50	48.94	47.50	50.70	44.50	50.35	49.30	48.58	43.73	52.73	50.35	47.50	53.07	50.00	55.02	52.40	52.06	50.00
Run 14	42.16	46.39	43.34	46.76	48.22	52.06	48.58	50.35	51.72	47.13	45.27	47.50	48.22	53.07	50.35	48.22	48.22	51.72
Run 15	42.95	42.55	47.86	48.58	47.50	51.38	44.50	50.35	49.65	47.50	54.05	50.35	48.94	53.40	50.35	48.94	50.00	51.04
Run 16	44.12	45.27	47.50	48.22	49.30	47.86	42.55	46.39	53.73	48.58	49.30	44.12	50.00	50.00	44.50	50.00	55.66	52.06
Run 17	46.39	47.50	46.39	43.73	50.70	49.65	42.16	51.72	51.38	48.58	45.64	47.50	46.39	50.70	52.73	52.73	47.13	50.00
Run 18	45.27	44.50	46.02	48.94	50.00	49.30	47.13	43.73	46.02	52.40	52.06	51.72	48.58	51.72	48.94	50.00	51.38	46.02
Run 19	45.64	46.76	48.22	41.76	49.65	48.58	43.34	47.50	50.70	48.94	48.94	51.38	50.70	53.40	50.35	47.13	49.65	50.35
Run 20	39.72	48.94	45.27	47.50	48.22	48.58	47.86	42.95	47.86	47.86	53.40	52.40	44.89	49.30	55.34	49.30	48.58	50.35
Run 21	46.02	48.22	44.89	50.00	47.86	51.72	48.94	48.22	51.04	46.76	46.02	47.86	47.86	49.30	51.72	51.04	51.38	52.40
Run 22	44.50	49.65	46.02	46.76	50.35	50.35	49.65	45.27	46.39	44.12	47.86	50.70	52.06	46.39	53.40	52.40	52.06	51.04
Run 23	51.72	45.64	44.89	46.39	47.50	46.76	47.13	51.38	52.73	49.65	54.38	50.70	47.86	50.35	54.05	50.00	52.06	51.04
Run 24	42.95	45.27	45.64	42.95	46.02	47.86	44.50	52.06	49.30	47.86	44.89	53.40	55.02	52.40	48.22	49.30	54.38	50.35
Run 25	52.06	43.34	46.39	42.55	47.86	47.13	48.58	44.50	46.02	49.30	51.72	57.53	54.38	46.39	48.22	52.06	45.27	52.40
Run 26	44.12	40.13	50.70	54.38	48.94	49.65	49.30	51.38	49.65	45.64	50.00	45.64	52.73	49.65	54.70	55.02	53.40	51.72
Run 27	44.89	46.76	46.76	45.27	48.58	49.65	51.04	50.00	47.86	46.02	50.00	46.39	52.40	53.40	52.06	52.73	53.07	54.05
Run 28	46.76	42.95	44.12	50.70	48.58	51.38	49.65	47.13	46.39	46.02	50.35	48.22	51.04	52.73	53.73	52.06	51.38	52.73
Run 29	44.50	46.02	48.58	53.07	55.34	50.70	50.70	47.50	51.72	46.02	49.65	46.02	50.70	55.98	54.05	47.50	50.00	46.39
Run 30	46.02	46.02	48.58	50.70	52.73	47.86	52.73	47.13	52.40	48.94	51.04	51.38	52.06	47.86	47.50	52.73	54.70	51.04
Mean	45.04	45.66	47.13	47.73	48.67	49.42	48.01	48.31	49.66	48.46	50.03	49.96	50.47	50.81	51.45	50.25	50.82	51.12
Std dev	0.0289	0.0265	0.0187	0.0312	0.0253	0.0188	0.0293	0.0278	0.0242	0.0252	0.0268	0.0300	0.0255	0.0258	0.0278	0.0256	0.0281	0.0214

Table B-50: Detailed results for the 30 runs of the cross-validation experiments using the video processing system without the foveal colour encoding on the MNI_by_5 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	43.34	44.50	44.12	46.39	47.13	50.00	46.02	41.76	49.30	50.70	52.73	52.73	49.30	47.86	49.65	48.94	50.70	54.38
Run 2	39.72	40.95	51.04	45.64	44.50	47.13	49.65	47.86	54.38	50.35	49.65	51.04	55.66	43.34	48.58	42.95	47.50	48.94
Run 3	46.76	47.86	44.12	47.13	47.50	44.50	51.72	44.89	47.86	51.72	54.70	51.04	53.07	48.94	47.86	54.70	54.38	49.65
Run 4	42.95	44.12	50.70	50.35	46.76	47.86	45.27	48.22	49.65	50.70	52.40	52.40	48.58	49.30	51.72	53.07	51.72	45.27
Run 5	42.95	49.30	48.94	48.22	48.58	40.95	48.22	53.07	50.70	52.06	49.30	55.02	50.70	50.00	51.38	50.70	50.70	46.02
Run 6	48.94	46.39	43.34	42.16	47.13	46.39	49.30	53.73	48.94	53.07	50.70	45.64	51.38	49.65	52.40	53.73	53.40	49.65
Run 7	51.04	44.50	48.22	46.76	46.76	53.73	46.39	50.35	46.76	49.30	45.27	48.94	47.86	49.30	55.98	40.95	50.35	55.66
Run 8	46.39	51.04	47.50	43.73	50.35	46.02	44.50	50.00	53.73	53.07	48.22	46.76	50.00	53.73	51.38	46.39	48.58	48.94
Run 9	43.73	41.76	46.02	47.50	51.04	48.94	46.76	47.86	49.65	51.38	51.72	54.05	50.70	50.35	48.58	49.65	52.06	50.35
Run 10	47.86	47.86	44.12	43.73	42.16	47.86	47.86	50.00	44.89	53.07	56.91	54.70	51.04	47.50	51.04	47.13	50.35	50.00
Run 11	43.34	48.22	48.22	46.76	51.04	42.95	43.34	50.35	47.50	52.40	52.40	53.07	52.40	50.35	48.94	45.27	44.12	44.50
Run 12	43.34	49.65	51.04	51.04	44.50	46.39	44.12	56.91	39.30	50.35	47.50	47.50	42.55	53.07	46.76	52.73	50.70	47.86
Run 13	40.13	43.73	51.72	45.64	49.65	50.00	46.02	46.39	51.38	43.34	51.04	52.06	53.07	51.72	48.94	51.72	51.72	51.72
Run 14	44.89	48.58	49.30	48.58	52.40	40.95	44.50	47.50	53.07	49.30	50.70	51.72	50.00	50.00	49.65	54.38	49.65	48.94
Run 15	41.76	42.95	48.58	46.39	49.30	47.86	50.00	50.70	48.58	54.70	56.29	52.06	50.00	53.73	50.00	50.70	54.70	51.04
Run 16	42.55	48.58	50.00	49.65	48.22	47.50	48.58	45.27	54.05	53.07	51.72	55.34	52.06	43.34	48.22	47.86	49.30	48.58
Run 17	43.73	48.94	48.58	47.13	45.27	50.00	50.35	50.35	45.64	50.70	45.27	43.73	51.38	48.58	50.70	49.65	49.30	53.07
Run 18	48.22	46.39	49.65	47.50	48.22	50.00	55.34	52.73	46.76	52.06	53.73	50.70	52.06	46.02	51.72	46.02	55.34	46.02
Run 19	43.34	44.12	44.89	48.94	47.13	51.38	48.94	50.35	52.73	49.65	44.50	49.65	46.76	46.02	49.30	50.35	49.65	48.94
Run 20	39.72	44.89	51.04	43.34	47.50	44.89	47.86	46.39	44.89	50.00	54.38	51.72	50.70	52.73	49.65	51.72	46.39	48.22
Run 21	42.16	51.04	49.30	48.94	52.40	46.02	48.58	48.58	51.04	45.64	48.94	49.30	51.72	51.72	50.00	53.07	44.89	50.70
Run 22	45.64	48.58	41.76	51.72	46.39	49.30	47.50	44.89	49.30	51.38	53.07	50.00	52.40	51.04	49.65	53.07	48.58	51.72
Run 23	43.34	42.55	44.12	44.50	45.64	54.05	49.65	49.65	41.35	53.40	52.40	53.07	48.58	49.30	47.86	53.07	47.86	47.50
Run 24	48.94	51.38	52.40	45.27	45.64	48.22	51.38	47.86	49.30	52.40	45.27	47.86	48.94	48.94	50.35	45.27	55.66	44.12
Run 25	44.89	44.50	49.30	49.30	49.30	48.58	46.76	54.38	51.72	49.65	56.29	55.34	56.60	46.39	53.07	46.76	50.00	48.94
Run 26	47.86	44.50	46.02	44.50	43.73	44.89	46.39	46.02	52.73	59.04	52.73	44.50	53.07	45.64	50.00	47.86	44.12	51.04
Run 27	47.86	49.65	50.00	47.13	50.70	51.72	54.38	46.39	51.04	46.02	51.04	49.30	51.38	47.13	50.00	48.94	52.40	50.35
Run 28	43.34	38.04	43.34	44.12	52.73	45.64	51.04	48.22	51.38	48.58	49.30	52.73	47.86	46.76	48.58	48.58	51.04	47.50
Run 29	44.12	38.88	47.13	47.50	47.13	49.30	46.02	49.30	49.65	50.35	47.50	53.73	45.27	48.22	50.35	49.30	51.38	54.38
Run 30	45.64	44.12	45.64	42.95	47.86	50.35	54.05	47.50	52.73	45.64	47.50	50.00	48.94	51.04	44.50	55.02	50.35	49.65
Mean	44.62	45.92	47.67	46.75	47.89	47.78	48.35	48.92	49.33	50.77	50.77	50.86	50.47	49.06	49.89	49.65	50.23	49.45
Std dev	0.0281	0.0348	0.0288	0.0242	0.0260	0.0314	0.0300	0.0314	0.0353	0.0297	0.0335	0.0306	0.0279	0.0269	0.0204	0.0350	0.0293	0.0275

Table B-51: Detailed results for the 30 runs of the cross-validation experiments using the video processing system and Izhikevich Type F neurons in %.

Template	MNI			Talairach					
	by 3	by 4	by 5	by 5	by 6	by 7	by 8	by 9	by 10
Run 1	36.75	44.50	53.07	38.88	44.89	42.55	51.38	41.35	42.16
Run 2	36.75	41.35	46.02	39.30	36.31	43.34	45.64	46.02	47.13
Run 3	43.34	46.02	42.55	44.12	36.31	45.64	46.02	43.34	52.06
Run 4	35.43	42.95	48.94	37.61	40.54	38.04	42.55	42.55	49.65
Run 5	40.54	45.27	44.12	42.16	35.43	50.70	42.16	46.02	45.64
Run 6	40.95	45.27	51.04	43.34	40.13	44.12	51.04	42.55	46.02
Run 7	42.16	52.06	47.13	43.34	48.58	45.64	48.22	41.76	49.30
Run 8	41.35	44.12	49.30	44.50	43.73	38.88	45.64	46.39	46.02
Run 9	42.55	43.34	44.89	40.13	41.35	42.95	43.73	45.27	47.50
Run 10	44.12	45.27	51.72	43.73	41.76	42.16	46.76	42.55	43.73
Run 11	40.13	39.72	46.39	40.13	34.99	39.72	45.64	41.35	40.13
Run 12	38.46	38.04	45.64	44.50	43.34	40.54	45.64	40.95	47.13
Run 13	40.95	40.54	47.50	47.86	41.35	45.64	43.34	41.76	45.64
Run 14	37.61	40.95	45.64	43.73	42.16	47.86	42.95	42.95	43.34
Run 15	42.16	43.34	42.95	43.73	47.13	43.34	48.94	40.54	48.58
Run 16	42.55	48.22	51.04	44.89	38.46	43.34	43.34	42.55	50.70
Run 17	40.54	44.50	47.13	44.12	42.16	44.12	45.64	43.34	44.12
Run 18	44.89	42.55	52.06	43.73	40.13	41.35	44.12	46.02	45.27
Run 19	39.30	43.34	46.76	39.72	38.04	44.89	46.39	37.18	43.34
Run 20	46.39	40.95	45.64	39.30	42.16	44.89	46.76	46.39	46.02
Run 21	40.13	50.70	45.27	42.16	38.04	42.95	47.13	43.34	41.35
Run 22	42.95	43.34	48.94	40.95	33.64	44.50	48.22	47.13	43.73
Run 23	45.27	45.27	43.34	49.30	39.72	42.55	46.76	46.76	44.12
Run 24	41.76	47.86	42.55	40.13	40.13	40.54	47.86	38.88	51.72
Run 25	40.54	37.18	48.22	39.72	41.76	37.61	46.76	41.35	45.64
Run 26	35.43	38.46	51.72	41.35	42.95	41.76	42.95	40.95	49.65
Run 27	42.95	43.34	48.58	39.72	45.64	44.50	45.64	42.95	45.64
Run 28	38.04	38.04	44.12	40.54	42.16	44.89	47.13	44.50	42.55
Run 29	38.04	40.95	46.02	38.46	46.76	40.13	38.46	45.27	51.38
Run 30	44.50	45.27	42.55	44.50	42.55	41.76	49.65	46.02	51.04
Mean	40.88	43.42	47.03	42.19	41.08	43.03	45.88	43.27	46.34
Std dev	0.0285	0.0348	0.0304	0.0273	0.0358	0.0277	0.0271	0.0244	0.0320

TABLE B-52: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM WITH OPTIMISED RETINAL ENCODING THRESHOLDS ON THE MNI_BY_3 TEMPLATE IN %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	42.55	38.46	45.64	42.95	45.64	41.76	47.50	47.50	41.35	40.54	45.64	47.86	43.73	40.54	45.27	46.39	44.12	45.27
Run 2	38.04	47.86	44.50	49.30	44.89	44.12	41.35	43.34	45.64	49.30	41.35	41.76	40.54	49.65	47.13	45.64	47.13	44.89
Run 3	37.18	47.86	46.02	42.55	51.04	50.00	49.65	46.76	47.86	44.12	39.72	44.12	50.00	42.95	50.35	43.73	44.50	42.95
Run 4	42.95	43.73	45.27	42.16	46.02	43.34	50.00	45.27	44.89	44.89	47.86	44.12	46.02	39.72	44.50	42.95	47.13	43.34
Run 5	44.50	41.76	42.55	34.54	42.95	39.72	44.50	47.50	46.02	49.65	47.86	47.86	44.50	47.13	44.50	49.65	44.12	42.16
Run 6	38.88	48.58	45.64	47.50	47.13	50.35	42.55	43.73	47.86	48.22	49.65	43.34	46.76	40.54	43.73	45.64	52.40	45.64
Run 7	45.27	48.94	45.64	43.73	44.50	47.86	47.50	46.39	45.27	46.39	44.12	48.22	45.27	45.27	47.13	39.30	47.13	43.34
Run 8	37.61	43.34	42.95	44.12	47.13	50.35	43.34	50.70	43.34	42.55	44.12	44.12	41.76	40.54	42.95	39.30	42.95	43.34
Run 9	43.73	46.39	46.39	46.39	41.76	44.89	45.64	45.64	47.50	46.39	46.76	42.16	44.89	44.89	40.13	45.27	44.12	47.13
Run 10	43.73	46.39	42.95	37.61	44.89	42.55	44.50	44.12	40.54	47.13	47.13	42.55	47.13	43.34	42.55	44.50	43.34	43.73
Run 11	43.34	40.95	47.13	48.58	48.22	46.39	50.70	50.35	43.34	39.72	41.35	42.16	47.86	42.95	41.76	45.64	43.73	45.27
Run 12	41.35	48.58	47.13	46.39	42.55	46.39	51.38	45.27	49.65	47.13	42.55	44.50	46.02	41.35	45.27	39.30	40.54	44.89
Run 13	45.27	48.22	51.38	44.12	49.30	47.50	40.13	40.54	46.76	47.50	44.89	48.94	47.86	42.95	43.34	40.95	37.61	45.27
Run 14	42.16	47.50	46.02	46.76	42.16	47.50	42.16	45.27	42.55	46.76	46.39	44.89	38.04	43.34	42.95	50.70	41.35	46.39
Run 15	44.89	42.95	46.39	46.02	48.94	44.50	46.39	45.64	47.50	42.16	44.50	40.95	43.34	44.89	42.16	41.76	43.73	51.04
Run 16	42.95	43.73	46.02	40.95	40.13	40.54	47.13	44.50	46.02	52.40	44.12	43.73	42.55	42.16	46.39	44.89	46.76	42.16
Run 17	42.16	46.39	47.50	44.89	50.35	49.65	46.76	43.73	44.50	47.13	49.30	46.39	45.64	45.64	46.02	46.39	43.73	41.76
Run 18	43.73	48.58	46.39	46.39	46.76	47.13	48.94	44.89	40.54	45.27	44.50	43.34	40.13	49.65	46.76	46.76	46.02	46.02
Run 19	44.12	48.58	43.34	47.13	45.64	46.39	44.50	45.64	46.76	49.30	41.76	44.50	43.73	47.50	39.30	46.02	41.35	45.64
Run 20	38.88	40.13	43.73	46.39	43.34	36.75	48.22	43.34	49.30	38.46	42.55	44.89	42.55	44.50	43.73	47.13	44.12	45.64
Run 21	37.18	45.27	46.76	46.39	43.73	51.72	47.50	48.22	49.30	44.50	46.02	43.73	46.02	44.12	49.30	43.34	42.55	40.95
Run 22	39.72	46.39	46.76	47.50	49.30	45.64	46.02	51.04	49.65	51.72	47.13	49.30	43.73	41.76	42.95	47.50	43.34	46.02
Run 23	41.35	44.50	50.35	43.73	45.64	35.87	46.02	48.22	47.13	44.50	41.76	43.34	39.72	44.12	44.12	46.76	47.86	40.13
Run 24	36.75	47.86	44.89	38.88	47.50	51.72	44.89	47.13	44.50	44.89	39.30	46.39	40.95	42.55	46.76	42.55	42.16	42.16
Run 25	43.73	41.35	46.39	46.02	44.50	50.35	47.50	41.76	45.64	46.76	47.50	48.94	52.73	47.50	47.86	40.13	43.73	43.73
Run 26	42.55	44.12	47.86	45.27	50.70	44.12	44.12	43.34	44.89	48.94	45.27	37.18	45.64	45.27	42.55	45.64	40.95	43.34
Run 27	38.88	42.16	43.34	46.76	46.02	46.02	47.50	46.02	50.35	50.35	40.95	41.76	44.12	44.89	44.50	46.39	44.12	45.27
Run 28	43.73	44.12	43.73	47.13	42.16	46.02	45.27	52.73	46.02	52.40	40.54	46.76	47.50	46.76	47.50	44.89	46.39	47.13
Run 29	38.04	46.02	49.65	47.13	48.94	43.34	48.94	45.64	42.95	45.27	43.73	45.64	47.86	42.55	49.30	45.64	47.86	44.89
Run 30	40.54	49.30	44.12	43.34	46.39	40.95	42.95	50.35	46.39	47.86	45.27	43.34	43.34	39.30	45.27	43.34	45.27	40.13
Mean	41.53	45.33	45.88	44.69	45.94	45.45	46.12	46.15	45.80	46.41	44.45	44.56	44.66	43.94	44.87	44.60	44.34	44.32
Std dev	0.0268	0.0294	0.0211	0.0325	0.0282	0.0404	0.0275	0.0280	0.0261	0.0343	0.0277	0.0268	0.0313	0.0267	0.0261	0.0283	0.0277	0.0227

Table B-53: Detailed results for the 30 runs of the cross-validation experiments using the video processing system with optimised retinal encoding thresholds on the MNI_by_4 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	41.35	46.76	45.27	48.94	47.50	49.65	48.58	49.65	46.02	43.73	42.95	43.34	48.58	48.58	50.70	54.05	45.27	52.73
Run 2	46.76	44.89	42.16	46.76	45.64	46.02	53.07	52.40	49.65	45.27	40.54	40.95	47.86	44.89	44.50	49.65	48.94	48.94
Run 3	43.34	44.89	49.30	46.76	48.22	48.22	50.00	40.95	46.39	47.50	47.50	45.64	45.64	46.39	46.76	48.94	46.02	46.39
Run 4	42.95	41.76	46.76	46.02	48.22	47.86	42.95	46.39	47.13	43.73	47.86	48.58	47.50	46.02	49.65	48.22	46.76	46.39
Run 5	48.58	48.22	48.22	51.38	48.22	47.50	47.13	53.40	51.04	50.00	51.38	48.94	48.94	51.38	48.22	45.64	49.65	50.00
Run 6	42.95	51.38	45.64	47.13	45.64	51.38	49.30	47.13	49.65	46.76	48.94	49.65	46.76	45.27	52.73	48.94	47.50	48.94
Run 7	42.16	50.00	45.64	43.34	47.13	48.22	50.00	48.22	47.50	47.13	48.58	48.58	45.27	53.73	48.58	50.35	50.70	49.65
Run 8	41.76	46.02	44.89	49.65	43.34	46.39	49.30	48.94	50.00	44.12	50.00	49.30	43.73	49.30	43.73	46.02	48.58	43.34
Run 9	43.34	42.16	46.39	48.58	42.95	44.89	53.07	46.39	47.50	45.64	46.76	50.35	44.89	41.35	47.86	48.58	48.58	52.40
Run 10	42.95	46.76	40.54	49.30	47.13	46.39	43.73	44.89	47.50	47.50	42.95	49.65	48.22	45.27	47.86	49.65	44.50	46.76
Run 11	41.35	49.65	47.13	47.86	52.40	46.02	49.30	43.34	47.13	44.50	42.95	47.86	47.50	44.89	52.73	45.27	50.35	48.58
Run 12	47.13	45.27	40.13	45.64	44.50	49.30	51.72	40.54	51.38	46.76	44.50	48.58	47.86	50.00	50.70	47.13	55.34	50.35
Run 13	45.27	44.50	47.13	50.35	49.65	52.40	51.04	47.13	46.02	48.22	46.76	50.00	50.70	51.38	49.30	50.00	48.22	47.50
Run 14	46.02	38.88	44.50	44.89	51.72	45.64	47.50	44.12	48.22	45.64	44.50	49.30	51.04	49.65	44.50	44.50	47.86	46.39
Run 15	35.87	45.27	47.86	51.04	48.58	53.07	50.35	43.34	46.39	46.39	46.39	51.72	48.58	52.40	46.39	49.30	51.38	48.58
Run 16	36.31	42.55	45.64	44.50	47.86	46.39	47.50	47.13	47.50	46.39	47.86	52.73	50.35	45.64	48.22	45.27	48.22	48.22
Run 17	43.34	41.76	45.64	49.65	49.65	48.58	47.13	46.02	51.04	54.05	44.50	45.64	46.76	45.27	44.50	48.94	50.00	50.70
Run 18	43.73	46.02	45.64	46.76	53.40	49.65	52.06	47.50	45.27	48.58	48.22	51.38	52.40	53.40	46.76	44.89	46.02	48.94
Run 19	46.02	48.58	48.94	42.55	46.02	44.89	50.70	45.64	49.65	48.22	46.39	48.58	52.06	47.86	47.86	46.39	48.22	50.35
Run 20	43.34	40.54	45.27	45.64	47.86	48.22	39.30	44.50	55.02	49.30	47.50	46.76	46.76	49.30	46.76	46.76	50.70	46.39
Run 21	44.50	48.22	46.39	49.65	44.50	46.02	47.50	44.89	45.27	47.86	46.02	46.02	44.12	47.13	49.30	47.86	50.00	52.06
Run 22	45.27	44.12	47.13	50.00	48.58	49.30	46.76	51.38	51.72	49.30	47.13	46.39	44.12	43.73	51.38	47.50	44.12	46.76
Run 23	45.64	47.86	46.76	40.95	44.50	50.70	48.22	47.50	43.73	52.40	51.04	48.22	48.22	45.64	46.76	44.89	49.30	46.02
Run 24	45.27	44.50	49.30	56.29	47.86	44.89	51.04	54.05	48.94	52.40	49.30	47.86	48.94	47.50	52.40	45.27	47.86	50.00
Run 25	43.73	43.34	46.02	49.65	48.58	50.70	51.04	46.39	49.65	48.94	50.70	46.02	49.30	49.65	47.86	46.39	44.50	49.65
Run 26	43.73	35.87	46.76	49.65	51.04	53.40	50.00	48.94	54.70	48.22	52.73	46.02	44.12	49.30	44.89	51.72	51.38	53.40
Run 27	42.95	40.54	44.89	45.64	46.02	48.58	53.07	42.55	53.07	49.65	46.76	48.58	44.50	50.00	45.64	49.65	44.50	49.30
Run 28	46.02	41.76	49.65	47.50	44.50	53.07	46.02	49.30	50.35	49.30	47.13	50.70	48.58	48.22	48.58	47.50	49.30	49.30
Run 29	43.73	46.39	40.13	38.46	49.30	47.50	41.35	46.76	54.38	48.22	51.38	45.64	44.12	49.65	47.13	48.94	48.22	43.73
Run 30	40.54	45.64	47.13	48.58	48.58	48.58	46.02	48.94	45.64	49.65	45.27	47.13	47.50	45.64	48.22	48.94	47.50	52.06
Mean	43.53	44.80	45.90	47.44	47.64	48.45	48.49	46.94	48.91	47.85	47.15	48.00	47.50	47.95	48.02	47.90	48.32	48.79
Std dev	0.0269	0.0341	0.0244	0.0340	0.0254	0.0248	0.0333	0.0325	0.0292	0.0247	0.0282	0.0245	0.0241	0.0291	0.0243	0.0221	0.0244	0.0246

Table B-54: Detailed results for the 30 runs of the cross-validation experiments using the video processing system with optimised retinal encoding thresholds on the MNI_by_5 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	41.35	51.72	48.22	48.22	46.02	48.94	44.50	47.86	46.39	48.94	52.40	50.70	50.70	50.70	52.73	49.65	46.76	53.73
Run 2	48.22	48.22	47.13	52.06	46.02	47.50	52.40	48.94	48.22	50.00	47.50	49.30	53.07	50.35	48.22	47.86	42.55	50.70
Run 3	51.38	47.13	42.16	43.73	45.64	46.39	47.13	47.50	49.30	46.76	50.00	43.73	52.73	52.73	42.16	45.27	51.04	47.86
Run 4	47.86	51.04	42.16	49.30	51.38	46.76	50.00	46.39	46.76	49.65	50.35	48.58	51.72	50.70	52.06	49.30	51.38	45.64
Run 5	44.50	45.27	50.00	46.76	46.76	51.04	47.50	51.04	47.86	49.65	46.02	52.40	48.22	46.02	48.94	48.58	52.40	48.22
Run 6	44.50	43.34	44.50	44.89	41.76	49.65	50.70	46.76	50.00	47.86	47.50	50.00	49.30	50.35	55.98	53.07	53.40	53.40
Run 7	44.89	47.86	49.30	50.00	44.12	45.64	49.30	50.00	48.94	49.65	53.40	51.38	53.07	51.04	48.58	51.72	47.50	50.70
Run 8	47.50	48.58	43.34	42.16	46.76	52.06	53.07	48.22	49.30	47.50	52.06	50.70	54.05	53.40	45.64	53.07	45.27	46.02
Run 9	42.16	43.34	48.22	46.76	47.13	47.13	46.39	47.50	47.86	52.06	52.40	44.89	48.22	40.54	50.35	47.50	52.06	47.50
Run 10	49.30	44.50	45.64	48.94	47.50	51.04	50.35	49.30	50.00	48.94	50.35	47.86	50.35	46.76	48.94	55.66	50.00	48.58
Run 11	48.94	50.00	49.65	48.22	46.76	49.30	49.30	50.35	46.76	41.35	48.94	54.05	53.73	49.65	48.94	48.22	50.35	45.27
Run 12	44.50	49.30	47.50	46.76	51.72	50.00	43.34	50.35	50.35	46.02	48.22	48.94	50.70	48.58	48.58	51.38	43.73	55.02
Run 13	49.65	47.13	46.39	39.30	44.89	54.05	39.72	47.50	48.94	47.86	54.38	50.35	51.72	49.30	50.35	47.86	48.58	48.94
Run 14	42.95	45.27	50.35	49.65	49.65	51.38	47.86	52.06	53.07	49.65	49.65	53.40	54.38	48.58	52.06	47.13	51.72	48.94
Run 15	48.22	45.64	48.22	46.02	50.00	50.00	47.86	44.89	46.76	51.38	50.00	53.40	55.02	45.64	51.38	52.06	46.02	44.89
Run 16	45.27	47.13	48.58	46.02	46.39	51.72	50.00	50.00	48.58	46.02	52.73	49.65	54.70	50.00	54.05	50.00	50.00	49.30
Run 17	47.50	46.02	46.39	51.38	47.86	50.70	48.22	47.86	48.94	43.73	47.50	51.72	48.58	47.50	50.35	55.98	48.22	47.50
Run 18	48.58	42.95	44.50	51.38	47.86	44.50	44.12	48.22	53.07	44.89	51.04	50.70	49.65	49.65	47.86	44.50	47.86	48.94
Run 19	41.35	43.73	46.39	47.13	52.06	50.00	50.35	46.76	51.72	52.73	48.58	50.70	45.27	48.94	53.07	47.86	52.06	44.50
Run 20	54.70	46.39	48.22	46.02	47.13	50.70	48.94	50.35	50.35	45.64	56.60	52.73	49.65	52.73	50.70	47.50	43.73	45.27
Run 21	51.72	47.86	45.64	45.27	47.86	50.35	44.12	47.86	52.40	50.70	51.72	50.70	46.76	52.06	54.70	46.39	47.13	54.70
Run 22	50.35	43.34	45.27	47.86	46.76	46.39	44.50	51.38	46.39	52.40	48.58	48.58	48.58	47.86	51.72	48.58	53.07	51.38
Run 23	45.27	46.02	49.30	47.13	49.65	45.64	48.58	49.65	46.39	49.30	47.50	53.40	43.34	52.40	52.73	48.58	51.72	54.38
Run 24	45.27	46.39	47.13	48.58	51.72	54.70	52.06	48.22	47.86	51.38	44.50	53.40	54.05	51.72	48.58	45.27	45.27	47.13
Run 25	52.73	42.55	46.02	46.39	43.34	48.58	42.55	46.39	49.30	50.70	48.94	51.72	50.00	54.05	47.50	51.04	49.30	49.30
Run 26	46.02	43.34	44.12	50.35	49.65	46.39	48.58	48.22	47.13	50.00	46.76	50.00	46.39	51.04	50.70	49.30	48.94	54.38
Run 27	50.70	49.65	51.72	52.06	51.72	46.39	47.13	48.22	48.22	48.58	46.76	47.50	51.38	51.04	46.39	51.38	47.86	50.70
Run 28	46.02	48.94	48.22	47.13	45.64	47.50	46.02	49.30	52.06	51.38	47.13	46.76	47.13	54.70	50.70	47.86	50.70	53.40
Run 29	51.72	44.12	46.39	45.64	51.72	47.13	54.05	49.65	50.00	44.50	46.02	47.50	48.22	50.00	48.58	48.22	48.58	48.94
Run 30	45.27	42.95	49.30	48.22	42.95	43.73	47.50	47.50	51.72	49.30	51.72	50.00	47.86	50.00	54.05	51.38	45.64	48.94
Mean	47.28	46.32	47.00	47.45	47.61	48.84	47.87	48.61	49.16	48.62	49.64	50.16	50.29	49.93	50.22	49.41	48.76	49.47
Std dev	0.0338	0.0257	0.0235	0.0279	0.0278	0.0267	0.0323	0.0163	0.0199	0.0271	0.0275	0.0248	0.0296	0.0279	0.0287	0.0277	0.0292	0.0311

Table B-55: Detailed results for the 30 runs of the cross-validation experiments using the video processing system with optimised retinal encoding thresholds on the TAL_by_5 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	42.16	41.76	41.76	40.54	38.46	44.50	47.86	44.50	45.64	50.00	47.86	48.58	52.73	47.13	47.13	44.89	44.50	47.50
Run 2	38.04	42.55	42.95	48.22	46.76	41.76	48.94	46.02	40.95	44.12	48.94	45.64	48.58	48.58	49.30	51.72	44.89	44.50
Run 3	38.04	37.18	36.75	38.46	46.02	43.34	40.54	46.39	47.50	51.04	50.35	49.65	47.13	45.64	48.22	50.35	51.04	46.02
Run 4	34.09	43.73	44.12	44.12	44.89	47.13	47.13	48.58	45.64	48.94	48.58	42.95	48.94	44.50	48.58	49.30	51.04	48.22
Run 5	47.13	38.46	36.31	43.73	47.13	43.73	44.50	45.64	46.76	50.35	47.50	49.30	42.16	48.22	44.89	41.76	47.13	42.95
Run 6	41.35	40.13	40.13	40.95	44.50	46.76	40.95	42.16	48.94	44.50	39.30	50.70	46.39	44.12	51.04	51.38	46.39	49.65
Run 7	36.75	38.46	49.30	46.02	44.50	47.86	42.95	51.38	46.02	43.34	50.00	48.58	51.72	45.27	44.50	43.73	43.34	43.34
Run 8	39.30	40.13	40.54	51.38	45.27	42.55	42.16	52.73	50.00	46.02	49.65	49.65	49.30	47.13	47.50	45.27	46.02	46.76
Run 9	40.95	39.72	44.50	46.76	46.76	47.50	37.61	43.73	41.76	51.04	52.06	46.02	43.34	46.39	46.39	47.50	48.58	46.39
Run 10	40.13	39.72	38.04	44.50	44.89	42.55	44.50	50.00	49.65	46.39	46.39	45.27	44.89	51.38	48.94	52.40	48.22	45.27
Run 11	40.95	38.88	44.12	44.50	42.95	43.73	46.02	40.54	42.16	48.58	51.38	45.64	47.50	45.64	52.06	43.34	46.02	48.94
Run 12	38.88	36.31	40.13	44.89	43.34	54.05	46.39	41.76	45.64	44.89	44.50	47.13	50.00	43.73	52.40	47.86	42.55	47.86
Run 13	37.61	39.30	44.50	44.12	46.39	52.40	47.86	38.88	41.35	44.89	51.04	48.94	46.02	45.27	51.38	42.95	47.13	45.27
Run 14	41.76	39.30	43.73	42.95	46.76	40.95	51.04	45.27	46.02	46.39	45.27	44.12	48.22	40.54	42.95	46.76	48.22	50.35
Run 15	40.13	38.04	40.95	46.02	48.22	47.13	46.76	45.27	51.38	48.22	53.07	49.30	45.27	48.58	45.64	49.65	47.50	44.89
Run 16	35.43	38.04	42.55	49.65	41.76	42.55	48.94	51.38	44.89	48.22	44.50	41.76	47.50	41.76	53.40	48.58	46.39	48.22
Run 17	38.88	38.46	45.64	45.27	47.50	42.55	46.39	40.13	44.12	45.64	51.72	40.95	48.94	48.22	48.58	42.55	48.22	46.02
Run 18	37.61	40.13	38.46	43.34	45.27	48.58	44.50	44.12	51.38	51.38	49.30	47.13	46.02	50.00	44.89	47.86	45.64	48.94
Run 19	38.04	43.34	47.50	45.64	42.55	47.13	53.07	47.86	44.89	54.70	47.86	45.27	48.22	49.30	50.00	51.04	40.95	47.50
Run 20	41.35	42.55	39.30	46.02	46.02	40.13	51.04	43.34	44.50	48.58	48.58	46.39	49.65	46.02	46.02	48.94	44.50	45.64
Run 21	40.13	38.04	39.30	48.58	51.04	49.30	47.13	47.13	48.94	50.70	48.58	52.06	50.35	45.27	50.00	47.13	43.73	48.94
Run 22	34.99	42.55	43.34	41.35	47.86	42.95	44.89	44.50	46.02	48.94	45.64	47.86	51.38	43.34	50.00	46.76	46.76	48.94
Run 23	39.30	43.73	39.72	40.95	48.22	42.95	47.50	47.86	42.16	42.95	47.50	47.50	47.86	46.02	50.00	47.13	40.54	49.30
Run 24	46.02	38.04	42.16	47.13	49.65	44.89	48.22	48.22	46.39	47.13	46.76	46.76	46.02	48.94	46.76	48.94	48.94	49.30
Run 25	44.50	43.73	41.35	47.86	43.34	49.65	47.13	50.70	47.86	50.00	46.76	45.64	46.39	46.76	46.76	46.39	45.27	41.35
Run 26	38.04	37.18	44.50	39.72	46.76	47.50	47.50	50.70	50.70	46.39	50.35	49.65	44.50	51.72	48.94	47.13	46.76	42.95
Run 27	42.95	40.54	40.13	46.76	42.16	44.12	44.89	43.73	47.13	46.39	49.65	50.70	48.22	45.27	48.58	49.65	40.54	49.30
Run 28	38.88	38.04	43.73	43.34	42.55	45.64	42.55	42.55	47.13	51.04	45.27	46.39	37.18	47.86	49.30	46.76	50.35	51.72
Run 29	38.46	40.95	42.95	42.16	51.04	47.50	46.02	46.02	50.00	48.94	49.30	42.95	46.02	46.39	44.50	46.76	50.70	51.04
Run 30	41.76	43.34	43.34	44.50	38.88	52.73	44.12	48.22	46.76	47.86	48.22	42.95	48.58	46.02	43.34	46.76	49.65	46.02
Mean	39.79	40.08	42.06	44.65	45.38	45.80	45.97	45.98	46.41	47.92	48.20	46.85	47.30	46.50	48.07	47.38	46.38	47.10
Std dev	0.0292	0.0221	0.0293	0.0296	0.0299	0.0348	0.0323	0.0353	0.0289	0.0272	0.0276	0.0276	0.0302	0.0250	0.0268	0.0271	0.0290	0.0253

Table B-56: Detailed results for the 30 runs of the cross-validation experiments using the video processing system with optimised retinal encoding thresholds on the TAL_by_6 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	35.87	35.87	45.64	44.12	48.22	45.64	50.35	48.58	50.35	46.02	41.76	46.02	46.39	38.04	48.94	44.50	48.58	46.39
Run 2	42.55	41.76	50.00	45.64	48.58	47.50	46.76	46.02	46.39	42.95	42.95	47.86	42.55	47.13	45.64	47.13	44.89	46.76
Run 3	39.72	36.75	44.89	48.22	46.39	49.65	42.95	48.58	48.22	38.04	43.34	51.04	48.58	51.04	46.76	44.12	50.70	49.65
Run 4	42.95	45.27	46.02	43.34	40.13	47.50	49.30	48.94	45.27	50.70	51.04	42.55	47.13	47.50	48.22	47.13	48.94	44.50
Run 5	40.13	39.72	48.58	50.35	48.94	40.95	47.13	48.58	42.55	48.22	41.35	45.27	44.12	50.35	45.27	44.50	45.64	42.55
Run 6	42.95	40.95	45.64	43.73	47.86	53.07	46.76	45.27	41.35	44.89	45.64	45.64	46.39	47.86	49.30	43.34	45.27	45.27
Run 7	47.13	44.12	42.55	47.13	49.30	43.73	46.02	47.86	50.70	49.30	47.13	49.30	43.34	47.50	44.50	43.34	47.13	44.12
Run 8	42.95	37.18	46.39	47.13	48.94	48.94	48.58	46.76	41.76	41.76	47.86	45.27	47.86	45.27	46.39	46.02	46.39	49.65
Run 9	37.18	40.54	42.55	46.76	42.95	47.13	45.64	50.35	47.50	43.34	45.27	46.02	47.13	45.27	42.95	41.35	51.04	43.34
Run 10	37.18	46.39	45.27	44.50	50.00	48.58	41.76	47.50	48.58	47.86	47.50	42.95	46.02	43.73	50.00	42.95	47.86	47.13
Run 11	37.61	42.16	53.07	44.50	43.34	44.12	52.40	46.39	48.58	46.76	46.39	50.35	47.50	48.58	47.86	46.39	43.34	42.55
Run 12	38.88	43.73	44.50	48.22	45.27	42.95	45.64	48.22	46.39	47.50	45.27	49.30	43.34	46.39	46.76	41.76	47.50	47.86
Run 13	46.39	41.35	48.58	46.39	45.27	46.39	48.94	46.39	48.58	42.55	43.73	48.58	48.22	45.27	46.39	46.39	42.55	44.89
Run 14	38.46	41.76	42.95	39.72	47.13	49.65	48.58	44.89	46.02	45.64	50.70	38.88	51.04	52.73	45.27	49.30	47.13	45.27
Run 15	37.61	43.34	40.54	47.86	46.02	46.76	45.27	46.76	48.22	48.94	41.76	42.95	43.73	44.50	50.35	46.76	46.76	45.64
Run 16	43.34	42.55	42.55	42.95	42.16	47.86	43.34	47.50	48.94	44.50	46.76	37.18	47.50	49.30	47.13	48.58	51.72	47.50
Run 17	46.39	38.46	42.95	46.39	49.30	44.50	53.07	46.39	42.95	51.04	43.34	46.39	49.65	47.86	50.70	46.02	49.65	43.73
Run 18	40.13	41.76	37.61	48.94	45.64	51.38	50.00	51.72	44.50	44.12	47.13	46.02	47.13	50.35	46.02	47.50	46.76	50.00
Run 19	40.95	40.13	38.46	47.50	40.13	47.13	45.64	43.73	48.22	45.64	42.55	45.64	50.00	44.50	44.50	45.64	49.30	47.50
Run 20	34.54	38.04	42.55	45.27	42.95	47.50	41.76	46.76	46.76	42.95	46.02	43.34	42.16	46.76	53.07	42.55	46.39	44.12
Run 21	43.73	37.61	44.50	41.76	48.58	45.64	47.50	50.00	49.65	45.64	43.34	45.64	48.22	44.89	40.54	44.50	44.50	46.39
Run 22	47.13	39.30	47.13	49.30	45.27	48.22	42.95	42.16	46.39	53.40	44.50	48.22	50.35	47.13	47.50	46.76	45.64	47.13
Run 23	38.04	44.12	44.50	44.12	47.13	44.89	43.73	50.70	44.89	50.35	46.76	42.55	49.30	46.39	45.27	49.65	44.50	44.12
Run 24	40.95	37.61	45.64	42.55	47.86	44.89	46.76	49.30	43.34	43.34	47.86	44.89	48.22	43.73	47.86	47.13	43.34	47.50
Run 25	42.95	38.46	45.27	45.64	51.04	50.70	51.04	44.50	46.39	44.89	42.95	42.16	48.94	44.89	45.27	45.27	42.55	46.39
Run 26	41.76	42.55	38.88	40.54	48.22	45.27	49.30	48.22	50.35	45.27	45.27	46.02	42.16	43.34	41.76	49.30	47.13	46.76
Run 27	42.16	44.50	44.89	46.39	46.39	45.27	49.65	49.30	42.16	49.65	45.27	43.73	49.65	48.94	40.54	42.95	46.02	46.02
Run 28	39.30	35.43	40.13	41.35	45.64	51.38	43.34	47.86	47.50	51.38	49.30	40.95	47.50	43.73	48.94	50.00	45.64	42.16
Run 29	35.43	33.18	38.04	44.12	42.95	50.00	44.89	47.86	42.55	42.55	39.72	44.12	43.34	49.65	46.02	45.64	48.58	47.50
Run 30	37.61	45.27	41.76	44.12	46.39	46.76	46.02	44.89	51.38	42.95	38.04	47.50	45.27	46.39	42.55	47.50	44.50	46.76
Mean	40.73	40.66	44.07	45.29	46.27	47.13	46.84	47.40	46.55	46.07	45.02	45.21	46.76	46.63	46.41	45.80	46.67	45.97
Std dev	0.0343	0.0324	0.0351	0.0261	0.0278	0.0270	0.0299	0.0211	0.0283	0.0342	0.0297	0.0313	0.0256	0.0286	0.0290	0.0232	0.0239	0.0205

Table B-57: Detailed results for the 30 runs of the cross-validation experiments using the video processing system with optimised retinal encoding thresholds on the TAL_BY_7 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	34.54	40.54	44.89	46.02	47.13	46.76	45.64	50.70	48.94	42.16	46.76	49.30	52.73	49.65	50.00	47.13	41.35	44.89
Run 2	42.95	40.13	43.34	50.70	45.64	51.72	54.05	45.27	48.22	51.04	55.34	50.00	51.72	51.04	52.40	46.76	49.30	48.58
Run 3	33.18	46.39	42.16	49.65	48.58	46.39	46.02	43.73	46.39	49.30	45.27	47.50	48.58	54.05	50.00	50.70	50.35	46.02
Run 4	41.35	40.13	42.95	48.22	46.39	44.89	44.89	48.94	46.02	46.02	46.39	45.27	49.30	51.72	48.22	52.73	48.58	51.72
Run 5	34.99	40.54	46.76	42.95	46.02	42.95	45.64	44.89	47.13	52.40	50.00	47.13	42.95	49.30	55.66	47.50	50.70	50.00
Run 6	41.76	43.34	49.65	43.73	46.02	44.50	46.02	47.86	48.94	48.94	49.65	50.00	46.39	48.22	50.00	49.65	48.58	50.70
Run 7	38.04	44.12	38.04	46.39	46.76	40.95	44.89	44.89	50.00	44.12	47.13	47.50	46.76	53.40	52.73	55.66	52.73	49.65
Run 8	32.26	44.12	45.64	49.30	50.35	46.02	46.39	48.94	50.70	47.13	46.02	52.40	52.40	48.58	54.70	49.65	50.00	51.72
Run 9	39.72	40.95	43.34	40.95	44.12	43.73	45.27	48.94	48.94	53.07	50.35	51.38	46.39	51.38	53.40	51.04	47.50	51.72
Run 10	35.87	47.86	45.64	44.50	48.94	49.30	46.02	48.58	48.58	52.06	47.13	47.86	47.13	51.72	48.58	52.73	48.58	53.07
Run 11	39.72	44.89	43.34	45.64	53.07	45.27	44.50	48.22	47.13	45.64	46.39	44.12	53.40	50.35	50.35	50.00	51.72	51.72
Run 12	38.04	38.46	46.76	44.89	44.89	49.65	44.89	47.13	50.35	49.65	48.58	52.73	52.40	48.94	47.86	54.05	53.73	51.38
Run 13	39.72	44.12	47.13	49.65	50.00	48.22	50.00	46.39	47.86	48.58	47.50	55.34	48.22	42.16	50.35	52.40	50.70	50.00
Run 14	35.87	44.89	44.89	43.34	47.50	43.34	53.07	52.73	44.89	49.30	44.50	48.58	51.38	52.06	46.39	53.40	50.00	51.04
Run 15	39.72	38.88	46.02	40.54	50.35	51.38	47.50	51.04	50.00	51.38	48.22	44.89	55.98	51.04	49.65	49.65	47.50	51.72
Run 16	38.46	42.16	39.72	44.89	48.22	46.76	50.00	47.13	44.50	54.05	52.73	48.94	46.76	42.55	50.00	42.55	50.70	50.00
Run 17	41.76	40.95	42.16	44.89	46.39	50.35	47.13	50.35	44.89	48.58	46.02	47.50	52.40	46.76	52.73	48.94	42.95	49.30
Run 18	42.16	45.27	44.12	45.27	46.76	52.40	54.05	55.34	48.22	45.27	46.76	50.00	54.38	46.39	44.89	52.06	49.65	50.35
Run 19	40.54	38.88	43.34	44.50	44.12	48.94	53.40	42.95	48.58	48.58	51.72	48.94	46.76	44.50	50.00	48.22	46.02	54.05
Run 20	40.54	46.39	45.27	47.13	45.64	51.04	47.13	47.13	50.00	48.94	48.58	48.94	51.72	57.53	51.38	49.30	50.00	51.38
Run 21	40.54	35.43	45.64	48.94	42.55	45.64	46.76	47.13	46.76	45.64	45.64	46.76	48.58	46.02	51.72	50.35	51.38	50.35
Run 22	34.99	46.76	45.64	47.86	46.02	49.30	37.18	43.34	48.94	51.04	51.72	48.94	51.04	53.07	50.00	48.58	50.00	52.73
Run 23	38.88	42.16	46.39	41.35	46.39	52.40	48.94	47.86	47.50	48.58	52.40	50.35	50.00	48.58	46.39	46.76	49.65	49.30
Run 24	44.89	40.54	46.39	46.02	45.64	44.50	48.94	45.64	45.64	44.12	46.76	51.38	53.07	53.07	54.38	51.38	50.00	50.70
Run 25	46.39	36.31	38.04	44.12	46.39	40.95	51.72	52.73	45.64	45.64	50.35	45.64	49.30	48.58	50.35	45.64	54.05	48.58
Run 26	43.34	40.13	46.02	40.54	49.30	50.35	49.65	51.38	50.35	47.86	51.04	49.30	48.22	48.22	49.30	49.65	51.72	51.04
Run 27	40.95	40.54	41.35	40.54	48.94	48.22	47.50	46.02	47.86	47.13	53.07	51.04	50.35	51.72	53.07	50.35	49.30	48.22
Run 28	42.95	43.34	40.13	49.30	42.16	45.27	46.39	46.02	48.58	45.64	45.64	47.50	43.34	44.89	52.40	49.30	50.00	47.86
Run 29	42.55	37.18	44.12	46.02	47.50	47.50	55.98	51.72	43.73	47.50	46.76	44.50	46.76	52.06	51.72	48.58	48.94	49.30
Run 30	39.72	42.16	42.95	42.95	44.50	39.30	48.94	47.86	43.34	46.02	42.55	50.35	49.65	49.65	47.86	47.86	53.73	47.50
Mean	39.55	41.92	44.06	45.36	46.88	46.93	47.95	48.03	47.62	48.18	48.37	48.80	49.60	49.57	50.55	49.75	49.65	50.15
Std dev	0.0337	0.0312	0.0266	0.0292	0.0235	0.0348	0.0373	0.0295	0.0203	0.0283	0.0291	0.0253	0.0309	0.0340	0.0250	0.0265	0.0270	0.0197

Table B-58: Detailed results for the 30 runs of the cross-validation experiments using the video processing system with optimised retinal encoding thresholds on the TAL_by_8 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	34.99	41.76	44.89	47.13	49.65	46.39	47.13	50.00	44.12	51.72	47.50	44.89	47.13	48.22	52.73	49.30	43.73	48.94
Run 2	46.39	36.75	50.70	44.12	48.22	45.27	48.94	47.50	46.39	45.27	52.06	43.73	44.89	48.94	45.64	50.35	48.58	45.64
Run 3	41.35	41.76	49.30	49.65	47.50	42.55	43.34	51.38	46.02	47.50	46.02	46.39	49.30	55.02	44.12	51.04	48.58	54.38
Run 4	40.54	36.31	47.13	44.50	50.00	45.27	44.89	48.58	49.65	44.89	47.86	51.38	49.30	51.04	46.39	50.35	49.65	50.00
Run 5	46.76	40.54	40.95	43.73	44.50	48.94	48.22	50.35	54.38	49.65	47.13	51.38	45.27	47.13	51.72	44.89	48.94	50.35
Run 6	42.55	31.79	45.64	47.50	49.65	45.64	49.65	52.06	46.76	51.04	47.86	50.35	46.39	51.72	50.00	50.35	49.65	50.70
Run 7	42.16	38.88	40.54	41.35	44.12	49.30	43.73	51.04	51.72	47.50	49.65	50.35	49.65	54.05	48.58	50.35	49.30	54.70
Run 8	44.12	41.76	44.89	48.22	44.12	47.86	47.13	48.58	49.65	42.16	52.40	50.00	50.70	53.40	49.30	50.70	46.39	50.70
Run 9	42.95	38.04	47.50	46.76	50.00	46.76	51.72	48.58	48.58	50.00	44.89	49.30	48.22	52.73	49.65	47.13	53.73	50.70
Run 10	45.64	40.13	47.13	41.76	42.95	48.58	48.58	44.89	49.30	49.65	48.22	47.86	46.02	50.00	47.86	45.27	48.94	48.94
Run 11	41.76	46.02	47.86	45.64	46.02	42.55	47.50	51.72	50.70	47.86	52.40	47.86	53.07	47.13	49.65	48.94	44.89	51.04
Run 12	46.76	36.75	41.76	42.16	38.46	49.65	48.22	47.50	42.55	56.60	42.95	46.76	43.73	46.02	52.40	50.00	53.07	47.50
Run 13	44.12	47.13	43.34	45.64	45.27	45.27	50.00	48.58	51.38	50.35	52.73	54.05	49.30	47.86	51.72	50.35	48.22	53.73
Run 14	41.76	43.34	47.50	40.95	50.70	48.58	47.50	50.35	51.04	49.30	49.30	50.35	48.94	53.73	52.73	44.50	45.64	46.76
Run 15	37.61	43.34	42.95	46.76	46.02	46.39	48.94	45.27	51.72	48.94	47.50	47.50	51.04	50.35	51.72	46.76	51.04	48.22
Run 16	40.95	39.72	44.89	45.27	48.22	44.50	50.35	52.73	50.35	45.64	48.22	46.39	52.73	48.94	52.40	51.38	53.73	53.07
Run 17	36.75	44.12	44.89	47.13	44.50	47.86	52.06	47.86	52.06	50.70	49.30	46.76	50.35	53.07	55.34	44.50	48.22	48.22
Run 18	37.61	31.79	50.00	44.89	47.86	48.22	49.65	46.02	49.30	49.65	46.76	44.12	45.27	50.35	50.70	48.94	52.40	53.07
Run 19	34.09	37.61	47.50	49.30	47.86	44.50	45.27	52.06	44.12	51.04	52.40	50.00	51.38	49.30	48.94	48.58	47.50	49.30
Run 20	38.04	41.35	52.40	46.02	50.35	51.72	48.22	47.50	50.00	48.58	52.06	50.70	50.70	53.40	51.04	51.72	50.70	48.22
Run 21	38.88	43.34	44.12	47.86	46.02	44.50	50.35	48.22	48.58	47.13	48.22	47.13	50.70	49.65	51.38	51.04	44.12	46.76
Run 22	38.88	44.12	45.64	43.73	49.65	48.58	47.50	55.66	47.13	48.94	46.76	46.02	50.00	48.58	53.40	50.70	46.02	51.04
Run 23	40.13	40.54	43.34	48.58	46.76	47.13	47.13	47.86	45.64	52.73	52.06	49.65	45.27	49.30	44.89	49.30	48.58	51.72
Run 24	47.13	42.16	41.76	52.73	42.55	46.02	45.27	48.22	52.06	51.38	51.72	49.65	51.38	50.00	48.94	50.70	51.04	51.04
Run 25	42.95	39.30	47.13	46.02	48.94	38.46	52.40	47.50	43.73	50.70	52.40	50.35	50.00	50.35	47.13	50.00	49.30	44.12
Run 26	33.64	44.12	54.70	46.02	48.22	51.38	46.39	51.04	45.27	47.50	50.00	50.00	45.27	52.73	53.40	50.35	47.13	51.38
Run 27	34.09	41.76	43.34	48.58	45.64	45.64	46.02	51.72	46.02	52.40	48.22	51.04	49.65	50.00	44.89	56.29	51.38	47.50
Run 28	46.02	43.34	44.50	46.39	43.73	44.50	47.50	50.70	46.39	54.05	46.76	48.22	42.55	49.30	51.04	50.00	46.39	48.94
Run 29	43.73	40.13	46.39	47.13	44.89	47.50	52.73	47.86	48.94	51.72	55.98	53.07	51.72	50.35	52.06	48.22	43.34	53.07
Run 30	39.72	46.02	47.50	47.50	44.12	46.39	47.13	46.39	48.22	47.86	50.00	48.58	47.86	46.76	52.40	50.70	54.70	41.35
Mean	41.07	40.79	46.00	46.10	46.55	46.53	48.12	49.26	48.39	49.41	49.31	48.79	48.59	50.31	50.07	49.42	48.83	49.70
Std dev	0.0394	0.0365	0.0324	0.0257	0.0281	0.0267	0.0238	0.0240	0.0289	0.0285	0.0280	0.0245	0.0274	0.0232	0.0282	0.0242	0.0297	0.0298

Table B-59: Detailed results for the 30 runs of the cross-validation experiments using the video processing system with optimised retinal encoding thresholds on the TAL_by_9 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	39.30	35.43	40.13	43.34	44.89	42.55	42.95	56.91	48.58	50.00	43.73	47.86	50.00	46.76	48.58	49.30	51.38	46.76
Run 2	31.79	39.30	43.34	42.95	47.50	44.50	44.12	48.58	48.22	42.55	42.16	53.07	52.73	49.65	53.73	47.50	51.04	48.22
Run 3	34.99	39.72	38.88	42.95	45.64	43.73	42.95	46.39	43.34	44.50	51.72	49.65	50.00	47.50	47.86	46.39	50.00	44.12
Run 4	36.31	46.76	35.87	41.35	45.27	44.50	48.58	46.02	40.95	48.22	52.40	49.65	50.70	48.22	49.30	48.58	51.38	46.02
Run 5	38.46	38.46	41.76	43.34	46.02	45.64	43.34	46.02	50.00	47.13	50.70	43.73	51.04	46.02	47.13	52.40	48.22	50.00
Run 6	40.13	35.43	38.88	47.86	42.55	49.30	46.02	46.02	50.00	46.02	50.00	49.30	50.35	45.27	48.22	53.73	46.39	43.73
Run 7	41.35	37.18	39.72	39.72	44.89	45.27	42.95	42.95	47.13	41.76	48.22	44.50	43.73	53.07	45.27	51.72	51.04	52.40
Run 8	42.16	35.87	46.39	43.34	42.16	41.35	45.64	47.13	43.34	45.64	47.50	46.39	52.73	49.65	47.13	48.94	47.50	50.00
Run 9	37.61	33.64	47.50	42.16	48.94	55.02	47.13	43.73	40.95	47.86	47.86	50.35	45.64	46.02	46.39	51.38	47.50	50.35
Run 10	39.72	35.87	39.30	37.18	51.72	48.22	41.35	47.13	46.39	48.22	46.76	50.70	48.22	53.73	51.38	44.12	49.30	48.58
Run 11	39.30	41.35	42.16	38.46	38.46	48.22	46.02	40.95	46.39	43.73	48.94	47.86	43.34	48.58	48.58	51.04	45.64	47.86
Run 12	33.18	44.12	40.13	45.27	46.02	48.22	48.58	47.13	49.30	46.02	45.27	44.50	48.22	48.94	50.35	53.07	49.65	46.76
Run 13	35.87	42.16	38.46	38.88	40.95	40.95	50.35	44.12	48.22	42.16	46.02	46.02	51.38	51.72	51.72	50.00	50.35	46.76
Run 14	37.61	40.54	41.76	41.35	50.70	51.38	45.64	51.38	48.58	51.38	47.13	47.50	50.35	42.55	46.39	48.94	49.65	44.50
Run 15	34.54	35.43	43.34	38.46	45.64	43.73	52.40	45.64	47.13	50.00	47.86	46.02	46.76	46.02	49.65	45.27	50.35	51.38
Run 16	41.76	41.76	44.50	42.55	44.50	47.13	42.95	51.38	50.00	44.89	49.30	51.04	48.58	50.70	50.35	48.94	50.70	43.34
Run 17	38.46	40.95	35.43	39.72	42.95	43.73	42.16	43.34	46.76	42.95	50.70	44.50	50.70	51.38	53.07	52.40	47.13	48.94
Run 18	36.75	38.04	39.30	39.30	44.89	41.35	50.00	47.86	49.65	48.94	46.76	48.94	51.72	49.30	51.04	56.29	45.64	46.02
Run 19	39.30	41.76	41.35	44.89	40.95	46.02	48.22	48.58	46.76	44.12	44.89	49.65	48.22	48.94	49.65	50.00	51.72	48.94
Run 20	39.72	46.02	40.54	42.16	41.35	50.35	45.27	46.39	50.70	46.39	48.58	48.22	48.22	50.35	49.65	40.54	49.65	50.35
Run 21	41.76	41.76	42.55	41.76	47.86	49.30	44.12	51.04	48.94	44.12	45.64	53.07	50.00	46.02	50.35	47.50	43.73	52.73
Run 22	35.87	44.12	36.31	38.88	46.76	48.22	42.95	44.50	40.95	50.00	48.58	51.04	49.30	52.73	51.38	51.72	45.27	51.04
Run 23	40.54	33.64	40.54	48.94	43.73	46.39	48.22	47.86	49.30	51.72	46.39	51.72	51.38	51.04	52.06	47.13	47.13	46.76
Run 24	44.50	41.76	36.31	49.30	47.13	45.64	48.94	44.50	48.58	45.64	47.86	45.64	44.89	51.04	52.40	51.38	47.13	51.72
Run 25	40.54	41.76	40.54	48.94	44.50	47.50	46.02	47.50	47.13	52.40	46.76	48.58	51.72	48.22	52.06	48.58	45.27	49.65
Run 26	40.54	40.13	41.35	43.73	49.30	48.22	47.86	42.55	48.58	52.73	51.72	46.02	47.13	53.40	54.05	44.50	47.86	48.58
Run 27	43.73	37.61	40.13	47.86	46.02	42.95	49.30	45.64	48.58	53.07	50.70	51.04	53.73	50.70	53.73	48.58	48.58	44.89
Run 28	37.18	40.54	37.61	44.50	42.16	43.34	52.06	42.55	46.39	50.70	50.70	51.04	49.30	44.89	45.64	47.86	51.04	48.94
Run 29	37.61	35.43	38.88	44.12	48.94	43.73	47.86	48.22	48.58	48.22	52.06	50.70	46.76	41.76	53.07	54.05	47.86	44.12
Run 30	38.04	42.55	38.88	46.39	43.34	46.02	41.76	48.22	49.30	44.12	50.70	50.00	45.64	50.00	52.06	47.86	46.76	48.94
Mean	38.62	39.64	40.40	42.99	45.19	46.08	46.19	46.68	47.29	47.17	48.25	48.61	49.08	48.81	50.08	49.32	48.50	48.08
Std dev	0.0290	0.0347	0.0281	0.0334	0.0301	0.0317	0.0305	0.0319	0.0271	0.0331	0.0253	0.0259	0.0264	0.0301	0.0249	0.0325	0.0217	0.0263

TABLE B-60: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS USING THE VIDEO PROCESSING SYSTEM WITH OPTIMISED RETINAL ENCODING THRESHOLDS ON THE TAL_BY_10 TEMPLATE IN %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	39.72	46.76	47.13	49.65	50.00	46.02	50.00	46.76	51.04	46.39	48.58	46.76	42.55	48.94	54.38	50.35	50.35	47.86
Run 2	35.43	40.95	45.64	47.13	50.00	48.94	48.94	47.13	48.94	48.58	45.64	52.40	50.35	51.72	48.94	43.73	52.73	54.38
Run 3	41.76	38.88	47.86	44.50	46.02	42.55	44.12	46.76	48.22	47.86	47.86	55.34	48.22	51.04	52.06	48.22	48.58	51.04
Run 4	42.55	43.34	48.22	38.88	48.22	52.06	43.34	40.95	47.13	46.76	42.55	45.64	54.38	45.27	51.04	49.65	54.38	47.50
Run 5	47.13	46.02	44.50	46.76	47.50	49.30	49.30	44.50	49.30	47.13	48.58	48.94	51.04	46.02	57.53	45.27	51.72	46.39
Run 6	40.13	46.76	48.22	49.65	46.02	48.22	43.73	49.30	42.95	46.76	52.40	49.65	49.30	47.86	54.38	50.70	49.30	48.94
Run 7	40.13	35.43	42.16	46.76	43.34	47.13	48.22	51.72	50.00	51.04	46.76	43.73	47.50	51.38	47.86	50.00	51.72	51.38
Run 8	46.76	44.89	41.76	42.95	45.64	51.04	43.34	51.38	54.38	49.65	56.60	48.58	53.40	51.72	49.65	46.02	44.89	54.05
Run 9	48.94	45.27	48.22	43.34	46.76	45.64	49.65	49.30	43.73	46.02	48.22	50.70	44.89	50.35	46.76	46.76	49.30	47.86
Run 10	41.35	45.27	46.39	48.22	46.76	47.50	52.73	46.02	47.86	52.73	48.58	43.34	50.35	50.70	54.05	44.89	48.22	47.86
Run 11	38.04	39.72	45.27	43.34	46.76	47.13	48.22	39.72	53.40	46.76	50.00	42.16	45.64	48.58	51.38	49.65	48.22	46.39
Run 12	43.34	35.43	47.13	47.86	46.39	45.64	48.22	42.95	48.58	47.13	46.39	53.40	46.02	48.94	59.04	51.72	49.65	50.00
Run 13	49.30	41.76	45.27	47.50	40.13	48.22	52.73	46.39	44.12	50.70	42.16	46.02	47.13	53.40	51.38	50.00	51.04	47.13
Run 14	43.34	43.73	44.50	46.76	51.38	48.94	52.40	51.38	43.73	50.35	48.22	46.39	52.73	52.06	49.65	49.65	51.72	47.50
Run 15	40.95	45.27	45.64	46.76	46.76	48.94	49.65	48.58	48.22	49.65	46.76	51.04	47.86	49.65	47.86	48.94	50.35	46.76
Run 16	40.54	39.30	44.12	48.58	43.73	52.40	44.12	47.50	48.22	47.13	48.22	49.30	55.02	55.98	57.84	46.39	53.07	56.29
Run 17	46.02	42.55	50.35	42.55	46.76	47.13	44.50	47.13	44.50	51.38	46.02	47.50	53.07	57.53	51.38	56.29	49.30	52.73
Run 18	42.55	37.18	47.86	48.22	49.30	45.64	47.86	52.40	42.95	43.73	47.13	45.64	49.30	54.05	48.22	48.22	55.34	51.38
Run 19	42.16	43.34	48.22	48.22	48.22	51.04	52.40	50.70	46.02	48.94	49.65	47.86	47.50	55.98	46.76	50.00	46.02	50.70
Run 20	40.54	45.27	43.73	45.27	48.58	46.76	46.76	43.34	52.06	48.94	52.40	45.27	46.02	51.04	48.58	47.86	56.29	50.00
Run 21	39.72	45.27	42.95	46.39	45.64	45.64	47.86	47.13	46.02	44.89	48.22	53.07	47.50	44.12	54.05	53.73	48.58	48.94
Run 22	35.43	40.95	44.89	53.73	48.94	45.64	48.58	47.50	46.39	46.39	52.06	48.58	47.50	52.40	48.58	49.65	52.73	47.50
Run 23	44.89	42.55	47.50	48.22	42.95	48.58	50.35	47.86	47.50	53.07	41.35	54.38	44.89	48.94	46.02	54.70	51.72	53.40
Run 24	38.88	39.30	48.58	44.50	44.12	48.58	46.39	46.39	46.39	46.39	38.46	45.64	48.22	48.58	51.38	50.35	49.65	47.13
Run 25	40.95	40.13	45.64	47.50	48.58	51.38	43.73	48.58	50.35	50.70	47.50	48.58	49.30	46.76	55.02	47.50	50.70	50.00
Run 26	44.89	38.04	47.13	47.50	45.27	49.30	47.86	40.95	50.35	44.89	47.50	46.02	47.86	47.13	48.22	48.58	49.65	49.65
Run 27	44.89	44.12	46.39	46.39	44.89	46.76	53.07	42.16	47.13	47.13	46.76	47.86	52.40	52.73	51.04	52.73	47.13	51.72
Run 28	42.95	41.76	42.95	44.50	45.64	47.50	53.40	46.76	51.72	53.07	47.50	52.40	52.40	59.34	56.91	54.05	53.73	46.76
Run 29	44.50	47.50	44.89	49.30	48.94	46.02	46.39	51.04	49.65	45.27	48.58	50.70	48.22	49.65	48.22	52.40	46.76	48.58
Run 30	47.50	44.12	46.02	47.86	48.94	51.72	52.06	47.50	49.65	43.73	48.22	50.00	47.13	48.58	52.40	52.73	51.04	49.65
Mean	42.51	42.36	45.97	46.63	46.74	48.05	48.33	46.99	48.02	48.11	47.63	48.56	48.92	50.68	51.35	49.69	50.46	49.65
Std dev	0.0347	0.0330	0.0206	0.0272	0.0239	0.0229	0.0315	0.0329	0.0296	0.0261	0.0342	0.0328	0.0301	0.0348	0.0351	0.0295	0.0263	0.0258

Table B-61: Averaged results of the sound and video cross-validation experiments in %.

Size	**MNI**			**Talairach**					**Ø**
LIF	by 3	by 4	by 5	by 6	by 7	by 8	by 9	by 10	
0.01	80.03	79.89	83.96	76.97	78.05	78.36	76.98	78.05	79.04
0.02	87.75	86.69	85.06	79.51	79.07	78.62	76.56	75.62	81.11
0.03	88.85	87.11	84.93	77.65	77.60	78.89	77.46	80.41	81.61
0.04	89.25	87.11	85.33	79.01	77.58	79.17	75.57	82.52	81.94
0.05	88.46	87.62	85.53	78.71	81.45	81.21	82.28	78.92	83.02
0.06	88.63	86.92	85.78	84.74	85.88	84.64	85.52	77.88	85.00
0.07	89.12	87.45	85.47	86.09	85.06	86.12	85.81	78.39	85.44
0.08	88.98	88.14	85.33	86.44	85.95	86.72	85.14	79.00	85.71
0.09	88.69	88.56	85.11	87.64	86.60	87.34	82.28	80.07	85.79
0.10	88.30	88.65	84.45	87.42	86.84	86.75	81.22	79.55	85.40
0.15	**89.65**	87.79	84.17	88.25	87.60	89.11	84.36	77.22	**86.02**
0.20	**89.45**	87.31	83.67	88.33	**89.72**	88.47	83.50	76.44	85.86
0.25	**89.34**	87.19	83.51	89.33	88.55	87.36	83.45	75.37	85.51
0.30	89.10	86.70	83.41	**89.79**	87.46	87.15	83.64	75.53	85.35
0.35	88.67	87.22	83.66	89.19	87.49	87.78	83.17	76.00	85.40
0.40	89.12	86.53	83.56	88.91	87.46	87.33	83.46	75.33	85.21
0.45	88.80	86.52	83.52	88.72	87.49	87.43	83.68	75.36	85.19
0.50	88.76	86.73	83.43	88.90	87.08	87.25	83.56	75.09	85.10
Ø	**88.39**	86.90	84.44	85.31	84.83	84.98	82.09	77.60	84.32

TABLE B-62: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS ON THE SIGN LANGUAGE DATASET USING ONLY THE SOUND DATA FOR TRAINING ON THE MNI_BY_4 TEMPLATE IN %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	44.89	48.58	47.86	47.86	48.58	55.98	52.06	53.40	59.04	56.60	47.86	39.30	49.30	39.30	32.26	45.64	45.64	32.26
Run 2	40.95	50.70	51.38	49.30	55.34	53.40	52.06	53.40	50.70	56.60	57.22	47.13	47.13	40.95	38.46	39.30	50.70	46.39
Run 3	50.70	41.76	47.86	58.44	48.58	49.30	54.70	47.13	57.22	56.60	62.50	49.30	57.84	40.95	51.38	47.13	43.34	47.13
Run 4	48.58	47.86	54.05	52.06	51.38	50.00	53.40	57.22	50.00	64.15	52.73	47.13	48.58	55.34	47.13	48.58	44.89	40.13
Run 5	46.39	45.64	52.73	48.58	52.73	57.22	54.05	58.44	61.37	60.22	57.84	50.70	51.38	44.12	45.64	50.70	50.70	49.30
Run 6	47.13	45.64	52.06	55.98	56.60	50.00	51.38	55.34	54.70	55.34	51.38	54.05	58.44	45.64	44.89	40.95	41.76	54.05
Run 7	50.70	47.13	55.98	58.44	52.73	51.38	60.22	50.00	56.60	57.84	59.04	56.60	57.84	46.39	44.12	48.58	43.34	41.76
Run 8	41.76	44.89	52.73	56.60	54.05	52.73	47.86	58.44	61.37	59.04	57.84	55.98	45.64	46.39	55.34	47.86	45.64	34.09
Run 9	55.34	44.12	48.58	43.34	57.84	54.05	65.22	59.04	58.44	55.98	53.40	57.22	52.06	46.39	47.86	40.13	48.58	37.61
Run 10	46.39	51.38	42.55	52.73	49.30	51.38	64.69	55.34	52.06	62.50	52.73	53.40	44.89	48.58	53.40	41.76	45.64	40.13
Run 11	45.64	48.58	50.70	52.06	49.30	60.80	59.04	57.84	57.84	62.50	55.34	53.40	47.86	44.12	44.89	40.13	49.30	35.87
Run 12	51.38	47.86	50.00	54.05	56.60	55.98	49.30	54.70	59.04	66.78	58.44	55.98	55.98	50.00	44.89	48.58	48.58	47.86
Run 13	47.13	44.89	54.70	46.39	59.63	61.94	55.34	47.13	54.05	54.05	65.22	54.70	49.30	40.95	37.61	44.12	42.55	40.95
Run 14	46.39	54.05	47.86	50.70	40.13	54.05	52.06	53.40	55.98	56.60	46.39	47.13	46.39	55.34	48.58	50.70	45.64	47.13
Run 15	51.38	55.34	47.86	44.12	54.70	63.60	50.00	50.70	61.37	57.22	55.34	59.63	50.70	44.12	44.12	40.13	44.89	48.58
Run 16	42.55	48.58	48.58	55.98	55.34	56.60	62.50	52.06	59.04	62.50	52.06	52.06	50.70	45.64	50.00	38.46	38.46	42.55
Run 17	51.38	52.73	46.39	55.98	55.34	54.05	52.73	55.34	52.73	59.04	57.84	48.58	49.30	56.60	57.84	43.34	40.13	44.89
Run 18	40.13	55.98	48.58	47.86	54.05	58.44	57.84	60.80	55.98	47.86	57.84	54.70	44.89	48.58	40.13	42.55	38.46	46.39
Run 19	53.40	44.12	41.76	52.06	58.44	55.34	41.76	60.22	67.80	59.63	57.22	52.73	38.46	45.64	45.64	47.86	39.30	45.64
Run 20	44.89	45.64	49.30	57.22	54.70	57.22	52.73	44.89	59.63	58.44	57.84	50.00	41.76	46.39	47.86	37.61	44.12	51.38
Run 21	43.34	38.46	59.04	57.22	53.40	60.22	57.84	62.50	54.05	59.04	52.06	45.64	43.34	55.98	52.73	45.64	47.13	47.86
Run 22	40.13	48.58	49.30	47.86	45.64	54.05	52.06	55.34	55.34	51.38	60.22	44.12	42.55	59.04	37.61	40.13	51.38	42.55
Run 23	47.13	44.12	41.76	44.89	53.40	44.89	51.38	57.84	62.50	47.86	58.44	40.95	51.38	44.12	37.61	43.34	47.86	44.12
Run 24	46.39	43.34	52.73	48.58	55.98	59.04	54.05	50.70	55.34	55.34	60.22	49.30	44.89	48.58	39.30	43.34	41.76	49.30
Run 25	44.89	52.73	52.73	55.98	47.13	55.34	55.34	61.94	51.38	59.63	66.78	44.89	48.58	55.34	53.40	37.61	36.75	37.61
Run 26	54.05	46.39	47.13	46.39	53.40	50.00	54.05	52.73	54.70	52.06	57.84	53.40	44.89	50.00	47.13	38.46	43.34	40.13
Run 27	44.12	46.39	55.34	48.58	50.70	54.70	60.22	54.70	57.22	61.37	56.60	44.12	47.13	48.58	41.76	43.34	43.34	34.99
Run 28	47.13	44.89	51.38	49.30	54.05	47.86	60.22	55.34	58.44	63.06	59.63	52.73	51.38	56.60	55.34	39.30	45.64	44.89
Run 29	48.58	47.86	45.64	59.63	59.04	60.80	54.05	52.06	55.34	54.05	57.84	53.40	41.76	42.55	37.61	47.86	40.95	45.64
Run 30	40.95	54.70	50.70	51.38	56.60	53.40	46.39	56.60	61.94	59.04	52.06	49.30	52.73	53.40	43.34	34.99	35.87	42.55
Mean	46.79	47.77	49.91	51.65	53.16	54.79	54.49	54.82	57.04	57.75	56.59	50.59	48.57	48.19	45.60	43.27	44.19	43.46
Std dev	0.0409	0.0413	0.0400	0.0460	0.0422	0.0432	0.0515	0.0427	0.0389	0.0435	0.0442	0.0488	0.0489	0.0540	0.0624	0.0425	0.0405	0.0524

TABLE B-63: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS ON THE SIGN LANGUAGE DATASET USING ONLY THE VIDEO DATA FOR TRAINING ON THE MNI_BY_4 TEMPLATE IN %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	69.77	71.66	71.19	76.09	76.51	71.66	78.15	76.09	75.24	72.12	76.09	72.12	75.24	69.28	73.93	69.77	69.28	66.78
Run 2	74.80	68.79	76.51	73.48	76.51	76.09	75.24	75.66	73.48	78.55	74.37	69.77	70.72	63.60	69.77	73.48	66.78	69.77
Run 3	68.79	73.93	78.55	74.80	72.12	73.03	75.24	74.37	65.74	69.28	70.72	73.03	66.27	66.27	64.69	69.28	67.29	69.77
Run 4	70.72	70.25	77.34	79.73	72.58	75.66	73.48	73.93	77.74	76.92	68.79	71.66	66.27	75.66	63.60	71.19	70.25	67.29
Run 5	67.29	74.80	71.19	67.80	76.09	74.80	71.19	77.74	76.09	78.55	77.34	72.58	68.30	69.77	64.69	67.80	68.79	71.66
Run 6	70.25	70.72	73.48	74.37	80.12	81.26	75.66	76.09	74.37	69.77	72.58	73.93	70.25	67.80	72.58	73.93	67.80	69.28
Run 7	73.03	70.25	67.80	80.12	75.66	75.66	71.19	77.34	72.12	76.92	72.58	72.12	75.66	71.66	66.27	65.74	69.28	66.27
Run 8	76.09	72.58	71.19	71.66	75.24	76.09	73.93	79.73	76.09	74.37	70.72	68.79	69.77	68.30	68.30	66.78	66.27	68.79
Run 9	70.25	72.58	80.12	77.34	71.66	73.03	80.12	76.51	81.26	73.48	72.58	71.66	73.93	67.29	71.19	69.28	67.80	70.25
Run 10	71.19	69.77	80.50	74.37	76.09	76.92	80.12	75.66	75.66	77.74	76.92	64.69	64.69	65.22	64.69	67.80	74.80	66.78
Run 11	71.19	66.78	75.24	74.80	75.24	72.12	77.74	72.58	72.12	78.95	73.93	75.66	68.79	68.79	67.29	76.09	71.66	71.19
Run 12	69.77	71.19	74.37	72.58	73.03	73.48	77.34	72.12	75.66	78.95	71.66	69.28	66.78	63.06	69.28	66.78	71.19	73.03
Run 13	73.03	70.25	73.93	77.74	77.74	75.24	73.93	73.03	75.24	74.37	74.80	72.12	73.03	67.29	68.30	70.25	73.48	65.74
Run 14	74.37	69.28	74.80	73.03	78.95	75.66	75.24	73.93	76.09	77.74	70.25	72.58	68.30	70.72	70.25	68.30	67.29	65.22
Run 15	72.12	72.12	72.12	77.74	75.24	81.63	78.15	73.93	77.74	78.95	71.19	73.48	71.19	68.79	65.22	70.25	63.60	61.94
Run 16	71.19	69.77	73.03	77.74	76.09	72.58	74.37	74.80	76.51	74.80	71.66	69.28	68.30	68.79	68.79	70.72	69.77	67.80
Run 17	69.77	70.25	76.92	74.37	76.51	69.77	78.55	79.34	75.24	76.51	71.19	73.03	68.30	66.78	77.74	64.15	73.03	69.28
Run 18	69.28	70.72	67.80	73.93	71.66	80.50	73.93	72.12	75.24	73.48	69.77	72.58	66.78	69.77	64.69	71.66	72.12	71.66
Run 19	74.80	70.72	77.34	70.72	75.24	76.09	79.73	73.48	71.19	71.19	71.66	69.28	70.25	68.79	70.72	67.29	67.29	65.74
Run 20	70.72	68.30	73.93	74.37	73.48	74.37	76.51	73.03	79.34	70.25	74.80	70.72	73.93	70.25	66.78	66.78	66.27	72.58
Run 21	68.79	75.66	66.78	75.66	72.12	80.88	75.66	77.74	77.74	79.34	69.77	73.93	70.25	72.58	66.27	68.30	67.80	71.66
Run 22	70.25	70.25	76.92	73.93	74.37	77.34	70.72	69.28	76.51	76.92	72.58	73.03	68.79	67.80	68.79	65.74	69.77	71.66
Run 23	70.25	75.24	73.93	73.48	74.37	80.12	74.37	74.80	75.66	75.24	78.95	72.58	73.03	65.74	69.77	65.22	61.94	67.80
Run 24	67.80	65.22	70.72	74.80	73.03	79.73	81.26	74.37	77.74	74.80	67.80	67.29	74.80	69.77	69.77	70.25	70.25	64.69
Run 25	73.93	71.66	73.48	76.92	76.09	76.09	73.48	73.03	77.34	72.58	75.24	68.79	71.19	73.93	70.25	62.50	68.30	64.15
Run 26	70.25	69.77	75.24	75.66	72.12	79.34	76.92	78.55	76.92	75.24	72.58	74.80	72.58	72.12	70.72	73.48	75.24	73.03
Run 27	68.79	67.29	73.48	77.34	72.12	72.12	76.51	80.50	75.66	73.03	75.24	69.77	68.30	71.19	70.72	69.77	68.30	66.27
Run 28	73.03	72.12	73.93	71.19	74.37	72.58	74.37	78.55	77.34	76.09	73.48	69.77	70.25	64.15	68.79	69.28	67.80	61.94
Run 29	68.79	68.30	74.37	76.92	72.58	74.80	74.80	72.58	69.28	72.12	69.77	70.25	69.28	64.15	70.25	64.15	65.22	69.28
Run 30	69.77	67.29	73.93	67.29	71.66	77.74	79.34	72.12	76.92	67.80	73.03	73.93	72.58	69.77	67.29	64.15	72.12	67.29
Mean	71.00	70.58	74.00	74.67	74.62	75.88	75.91	75.10	75.44	74.87	72.73	71.42	70.26	68.64	68.71	68.67	69.03	68.29
Std dev	0.0219	0.0241	0.0326	0.0295	0.0221	0.0312	0.0270	0.0265	0.0296	0.0317	0.0260	0.0236	0.0284	0.0297	0.0303	0.0317	0.0302	0.0304

Table B-64: Detailed results for the 30 runs of the cross-validation experiments on the sign language dataset using the combined sound and video data for training on the MNI_by_4 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	56.60	65.74	65.22	66.27	67.29	69.77	66.78	64.69	66.27	76.09	71.66	66.78	72.58	72.58	72.12	70.72	70.72	67.29
Run 2	61.37	64.69	65.22	57.84	65.74	66.78	67.80	73.48	71.19	71.66	73.48	71.66	70.25	66.78	68.30	60.22	67.80	69.28
Run 3	52.06	60.22	65.74	66.78	66.27	71.19	70.25	73.93	70.72	64.69	77.74	67.80	73.03	73.93	70.25	65.74	76.09	65.74
Run 4	59.04	64.15	60.80	67.80	60.22	57.84	69.77	72.58	68.79	70.25	76.09	68.79	74.80	68.30	71.19	69.77	69.77	70.72
Run 5	58.44	66.78	64.15	61.37	67.80	63.60	71.66	68.79	64.15	78.95	76.92	66.27	71.19	70.25	71.66	69.28	70.25	67.80
Run 6	47.86	59.04	63.06	59.04	70.25	60.22	74.37	70.72	65.74	77.34	68.30	71.66	74.37	70.72	61.94	67.80	71.19	66.78
Run 7	47.86	64.15	63.06	57.22	69.77	71.19	64.15	66.78	77.34	69.28	76.09	70.72	70.72	65.22	70.72	70.25	69.28	66.78
Run 8	47.13	60.22	60.22	64.69	66.78	64.15	71.19	71.19	73.48	67.29	72.58	80.50	67.29	64.15	69.77	66.27	63.06	67.80
Run 9	62.50	59.04	63.60	66.27	61.37	66.78	70.72	65.74	72.12	75.66	77.74	70.25	65.74	65.22	70.25	74.37	64.69	64.15
Run 10	50.70	60.80	61.37	66.27	67.80	70.25	70.72	73.48	73.93	76.51	76.09	68.79	67.80	64.69	72.58	65.22	61.94	65.74
Run 11	66.27	64.69	66.27	73.93	71.66	75.66	71.19	65.74	69.28	74.37	72.58	73.93	74.80	71.19	73.48	73.93	67.80	69.77
Run 12	54.05	60.22	63.60	61.94	72.58	63.60	71.19	74.80	72.58	77.34	71.66	67.80	72.12	68.79	68.30	64.15	66.78	66.78
Run 13	50.70	61.37	63.60	68.30	69.77	69.77	64.15	70.25	73.03	72.58	66.78	68.30	67.80	74.80	65.22	63.60	69.77	63.60
Run 14	53.40	63.06	59.63	62.50	64.69	65.74	68.79	71.66	69.28	72.58	75.66	70.72	70.72	70.25	67.29	60.80	68.30	63.60
Run 15	54.70	62.50	59.04	65.22	69.77	72.12	65.74	63.60	75.24	71.19	72.58	72.12	69.77	69.77	60.22	56.60	65.22	65.22
Run 16	52.73	66.78	61.37	59.04	64.69	71.66	72.58	65.74	64.15	65.74	70.25	71.66	77.34	73.03	69.77	68.30	71.19	69.77
Run 17	55.98	60.22	60.22	64.69	67.29	69.77	67.80	66.78	69.77	72.58	76.51	76.51	67.80	67.80	63.60	61.37	68.79	59.04
Run 18	52.73	64.69	60.80	64.69	66.78	68.30	69.77	66.78	71.19	72.12	70.72	76.09	72.58	68.79	70.72	69.28	68.79	74.80
Run 19	57.22	64.69	67.80	75.66	63.60	71.66	73.03	66.27	68.30	68.30	69.77	71.19	66.78	69.28	63.06	68.79	72.12	72.12
Run 20	52.06	66.27	62.50	66.78	69.28	64.69	68.79	75.66	73.03	72.58	76.92	71.66	76.51	71.66	72.58	71.19	73.93	64.69
Run 21	54.70	59.04	51.38	67.29	56.60	66.78	72.58	71.19	75.24	65.74	74.37	69.28	73.48	67.80	63.60	71.66	66.27	63.60
Run 22	54.05	72.12	61.37	70.72	66.27	63.60	75.66	63.60	71.66	71.19	76.51	77.74	79.73	69.28	67.29	64.69	63.60	66.27
Run 23	54.70	64.15	65.74	78.15	68.79	69.77	69.77	72.12	75.24	68.30	73.48	76.09	67.29	72.12	69.77	69.28	71.19	61.94
Run 24	52.73	57.22	59.63	61.94	69.28	70.25	71.19	68.30	74.37	75.24	65.74	68.79	68.79	73.93	75.66	59.63	63.60	66.27
Run 25	57.84	59.04	59.63	66.78	66.78	65.74	73.48	75.24	66.78	77.74	70.72	70.72	72.12	76.92	67.80	67.80	69.77	57.22
Run 26	61.94	61.94	67.29	64.15	68.30	71.66	71.66	67.29	71.19	69.77	69.77	70.25	60.22	69.77	64.15	71.19	64.69	65.22
Run 27	60.22	60.22	61.37	66.27	67.80	71.19	62.50	71.19	73.03	73.03	80.12	73.03	74.80	71.66	69.77	71.66	64.69	67.29
Run 28	48.58	62.50	64.15	62.50	68.30	63.60	70.72	73.93	69.77	73.48	69.28	61.37	71.66	76.09	69.28	76.09	66.78	66.78
Run 29	55.98	57.22	64.15	65.74	73.48	68.79	71.19	75.66	75.66	73.93	65.74	73.03	71.66	60.22	77.34	73.48	65.22	66.27
Run 30	55.98	61.37	64.15	68.30	68.30	74.80	75.66	71.19	69.28	76.09	74.80	67.80	73.48	65.74	73.93	68.79	60.22	69.77
Mean	55.00	62.47	62.54	65.60	67.24	68.03	70.16	69.95	71.06	72.39	73.02	71.04	71.24	69.69	69.05	67.73	67.78	66.40
Std dev	0.0456	0.0325	0.0315	0.0472	0.0346	0.0407	0.0315	0.0370	0.0339	0.0376	0.0372	0.0378	0.0388	0.0371	0.0402	0.0467	0.0359	0.0348

Table B-65: Detailed results for the 30 runs of the cross-validation experiments on the sign language dataset using only the sound data for training on the TAL_by_8 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	49.30	50.70	41.76	40.95	47.13	55.98	50.00	54.70	55.34	55.98	60.22	55.98	57.22	52.73	52.73	37.61	44.89	37.61
Run 2	43.34	43.34	47.13	35.87	47.86	48.58	56.60	55.34	60.22	57.22	49.30	60.22	54.70	52.73	54.70	44.89	46.39	40.13
Run 3	42.55	50.00	54.70	47.13	48.58	55.34	51.38	54.70	56.60	59.63	65.74	50.00	49.30	55.98	49.30	44.12	45.64	44.12
Run 4	34.09	46.39	47.13	50.70	49.30	47.86	51.38	55.34	54.05	61.37	57.22	52.73	54.05	47.86	57.84	52.06	48.58	50.00
Run 5	49.30	45.64	36.75	46.39	52.06	50.70	54.70	52.06	63.06	60.22	57.22	57.22	49.30	52.06	54.70	44.89	50.00	40.13
Run 6	49.30	51.38	44.12	54.05	47.13	52.06	56.60	50.00	52.06	59.63	61.37	54.70	51.38	52.06	52.06	59.04	58.44	49.30
Run 7	50.70	49.30	47.13	46.39	41.76	47.13	54.70	52.06	52.06	54.70	57.84	56.60	61.94	50.00	44.89	50.00	54.05	45.64
Run 8	49.30	52.06	44.89	43.34	53.40	49.30	58.44	54.70	65.22	57.84	56.60	52.06	54.05	56.60	47.86	41.76	45.64	47.86
Run 9	42.55	51.38	53.40	37.61	54.05	43.34	51.38	54.05	54.70	63.06	64.69	50.70	47.86	50.00	42.55	44.89	40.95	41.76
Run 10	40.13	46.39	47.86	42.55	50.70	49.30	49.30	52.73	61.94	54.70	55.34	53.40	54.05	54.05	51.38	40.13	48.58	47.86
Run 11	53.40	44.12	55.98	40.95	48.58	51.38	49.30	52.73	58.44	63.06	62.50	61.37	52.06	50.00	55.34	55.34	43.34	40.13
Run 12	44.12	40.13	47.13	47.86	56.60	54.05	55.98	54.70	49.30	58.44	58.44	56.60	56.60	57.22	47.13	47.13	47.86	37.61
Run 13	53.40	43.34	48.58	35.87	52.06	53.40	50.00	52.06	48.58	53.40	60.22	59.04	55.98	52.73	58.44	52.06	40.13	42.55
Run 14	47.86	48.58	41.76	40.13	55.34	53.40	46.39	52.06	65.22	54.70	53.40	53.40	55.98	53.40	54.05	45.64	44.89	43.34
Run 15	47.13	52.73	45.64	52.06	51.38	50.00	58.44	57.84	53.40	53.40	63.60	62.50	48.58	57.22	52.73	54.70	47.13	45.64
Run 16	43.34	50.00	39.30	42.55	47.86	52.73	53.40	52.73	59.04	52.73	61.37	50.00	58.44	59.04	53.40	50.00	37.61	46.39
Run 17	47.86	46.39	51.38	38.46	40.95	48.58	50.70	48.58	54.05	68.30	59.04	64.69	55.34	55.34	47.86	43.34	50.00	47.13
Run 18	48.58	47.13	46.39	44.12	33.18	47.13	53.40	54.05	54.05	57.84	60.80	57.22	57.22	46.39	48.58	47.13	46.39	52.73
Run 19	40.95	48.58	46.39	40.13	50.70	50.70	57.84	52.73	54.70	56.60	62.50	61.94	58.44	47.86	49.30	52.73	53.40	45.64
Run 20	39.30	44.12	46.39	46.39	52.73	44.89	52.73	54.70	53.40	61.94	55.98	52.06	48.58	52.06	37.61	49.30	38.46	49.30
Run 21	41.76	42.55	44.12	43.34	50.00	55.34	50.70	55.34	58.44	57.84	61.94	52.73	60.22	51.38	43.34	51.38	47.13	44.12
Run 22	41.76	44.12	45.64	41.76	47.86	53.40	40.95	54.70	50.00	58.44	60.22	50.70	52.73	54.05	54.05	55.34	51.38	40.95
Run 23	44.89	52.06	50.00	42.55	53.40	53.40	52.73	49.30	51.38	56.60	59.63	49.30	55.34	54.70	39.30	45.64	44.12	44.89
Run 24	32.26	44.12	51.38	43.34	47.86	52.73	54.05	52.73	57.22	52.73	65.22	58.44	52.73	55.98	52.06	47.86	44.12	37.61
Run 25	52.06	48.58	44.89	45.64	47.86	49.30	49.30	56.60	55.34	60.80	65.74	58.44	54.05	51.38	55.34	59.63	45.64	46.39
Run 26	49.30	49.30	44.12	37.61	50.00	46.39	46.39	55.98	56.60	63.60	65.74	54.05	53.40	42.55	54.70	44.89	54.70	34.99
Run 27	48.58	47.86	52.73	47.86	36.75	61.94	54.05	52.06	59.63	58.44	60.80	52.73	55.34	50.70	46.39	50.00	40.95	47.86
Run 28	52.06	42.55	40.13	47.13	50.70	41.76	56.60	51.38	53.40	61.37	63.06	54.70	53.40	52.73	49.30	54.70	53.40	47.13
Run 29	52.73	54.05	49.30	42.55	47.86	50.70	54.70	56.60	55.98	49.30	60.80	57.22	55.34	49.30	52.73	43.34	44.89	52.73
Run 30	46.39	46.39	41.76	35.87	52.06	47.86	53.40	55.98	62.50	54.05	59.04	54.70	49.30	52.06	44.12	51.38	54.70	47.13
Mean	45.94	47.44	46.60	43.37	48.86	50.62	52.52	53.62	56.20	57.93	60.19	55.52	54.10	52.34	50.13	48.70	47.11	44.62
Std dev	0.0529	0.0350	0.0446	0.0463	0.0501	0.0405	0.0385	0.0217	0.0438	0.0398	0.0376	0.0399	0.0347	0.0347	0.0515	0.0533	0.0502	0.0445

Table B-66: Detailed results for the 30 runs of the cross-validation experiments on the sign language dataset using only the video data for training on the TAL_by_8 template in %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	72.12	67.29	78.15	74.80	73.48	78.95	78.95	80.50	77.74	82.37	73.03	76.51	76.09	78.55	64.15	68.79	72.12	71.66
Run 2	74.37	73.93	78.55	76.92	76.09	80.12	82.73	80.88	79.34	78.95	81.63	81.26	80.88	77.34	75.24	65.74	72.12	69.28
Run 3	71.19	70.25	78.55	78.15	75.66	75.66	78.55	80.12	77.34	81.63	76.09	75.24	74.37	75.24	77.74	74.80	74.37	70.25
Run 4	73.03	77.34	80.88	74.80	79.34	78.15	80.12	79.73	79.73	80.12	80.50	75.24	78.95	78.15	73.93	76.51	70.72	70.72
Run 5	72.58	71.66	81.63	69.28	77.74	78.15	78.15	76.92	74.37	81.63	83.45	81.26	78.15	75.24	70.72	71.19	73.48	70.25
Run 6	70.72	70.72	74.37	76.09	78.55	82.73	78.15	79.34	75.66	82.37	79.73	76.51	79.34	75.66	72.58	71.66	72.12	72.58
Run 7	73.93	69.77	79.34	77.34	73.93	82.37	82.37	77.74	81.26	82.00	78.15	78.95	78.15	79.73	75.66	71.19	71.19	71.19
Run 8	73.48	73.03	79.34	78.15	77.34	71.66	80.50	77.34	82.37	78.95	76.09	79.34	76.92	80.12	77.74	75.24	73.93	66.27
Run 9	71.19	74.80	80.12	73.03	76.92	77.34	77.74	80.88	78.55	80.50	74.37	78.55	75.24	76.09	76.09	70.72	73.48	66.78
Run 10	72.58	73.93	75.24	74.37	72.12	75.24	75.24	82.37	80.88	77.34	75.24	74.80	76.09	80.12	73.93	75.66	74.80	72.58
Run 11	66.27	68.79	80.12	71.19	76.09	77.74	80.88	75.24	76.09	79.73	76.09	80.50	77.34	70.25	73.03	69.77	67.29	72.58
Run 12	68.30	68.79	73.03	81.63	74.37	73.03	80.12	79.34	78.95	82.00	77.34	77.34	80.50	77.34	73.48	69.77	67.80	73.93
Run 13	71.66	69.28	80.50	77.34	82.73	77.74	81.63	77.34	79.73	78.55	74.80	80.50	72.58	73.93	77.34	74.37	72.58	73.48
Run 14	70.72	69.77	81.26	73.48	73.93	80.50	79.34	78.95	74.80	80.12	82.73	78.55	74.80	77.34	74.37	73.93	70.25	70.72
Run 15	74.37	69.77	79.34	78.15	79.34	70.25	80.12	78.95	79.73	80.12	75.66	79.34	76.92	76.51	70.72	70.25	70.72	74.37
Run 16	70.25	68.79	80.12	77.34	76.09	78.95	82.37	76.92	74.80	81.63	74.37	80.50	82.00	73.93	74.37	73.93	66.78	70.25
Run 17	69.28	67.80	77.34	75.66	77.34	77.34	78.15	79.34	76.92	75.66	78.55	81.26	72.12	78.95	73.03	69.77	71.19	72.12
Run 18	73.93	72.58	76.09	76.51	79.73	77.74	78.55	81.26	74.37	75.24	79.34	74.37	76.51	76.51	71.66	70.25	72.58	69.28
Run 19	70.72	68.79	76.09	75.66	75.66	77.34	77.34	82.00	76.92	80.88	77.74	83.45	79.34	79.73	73.03	71.66	74.80	72.58
Run 20	71.66	73.03	80.12	78.55	79.34	82.73	78.95	78.55	79.34	78.95	76.09	79.73	75.24	73.48	74.80	70.72	72.12	74.37
Run 21	73.48	70.25	78.95	74.80	73.93	78.15	76.51	80.50	75.66	78.15	76.51	79.34	74.80	77.74	76.51	73.48	74.80	73.03
Run 22	71.66	75.66	75.66	75.66	80.88	80.88	78.15	80.50	76.51	75.24	77.34	78.55	78.55	74.80	73.03	78.55	68.79	65.22
Run 23	72.12	67.80	76.09	75.24	73.03	76.51	76.92	83.09	70.25	72.58	78.95	77.34	78.55	78.15	68.79	71.66	71.19	67.29
Run 24	73.48	73.93	79.34	77.34	80.88	73.93	82.37	77.34	78.15	78.95	78.15	76.51	73.03	79.34	76.51	74.80	67.80	66.27
Run 25	72.58	67.29	78.15	72.58	81.63	82.37	79.34	80.50	78.55	83.45	77.34	78.95	80.50	78.15	76.51	75.66	68.79	73.48
Run 26	72.12	70.72	75.24	79.34	78.55	76.09	73.03	81.63	73.48	80.50	76.51	77.74	78.55	75.24	69.77	71.19	69.77	71.19
Run 27	71.66	68.30	80.88	76.09	74.80	80.50	83.45	78.95	77.74	83.09	76.92	78.95	78.55	75.66	66.27	70.72	70.25	74.80
Run 28	73.48	71.19	78.15	78.95	71.66	80.50	76.92	80.88	77.34	77.34	81.26	80.50	76.92	73.03	73.93	74.37	70.25	69.77
Run 29	74.80	70.25	80.50	76.09	72.58	76.09	78.55	80.12	75.24	75.24	77.34	75.66	81.63	76.51	75.66	70.25	68.79	75.66
Run 30	73.48	73.93	79.73	76.09	80.12	75.24	78.55	76.51	79.73	81.26	76.51	78.15	74.37	74.37	77.74	71.19	70.25	64.15
Mean	72.04	70.98	78.43	76.02	76.80	77.80	79.12	79.46	77.38	79.48	77.59	78.50	77.23	76.57	73.61	72.26	71.17	70.87
Std dev	0.0186	0.0262	0.0222	0.0247	0.0299	0.0310	0.0229	0.0187	0.0258	0.0265	0.0246	0.0218	0.0260	0.0234	0.0324	0.0269	0.0228	0.0295

TABLE B-67: DETAILED RESULTS FOR THE 30 RUNS OF THE CROSS-VALIDATION EXPERIMENTS ON THE SIGN LANGUAGE DATASET USING THE COMBINED SOUND AND VIDEO DATA FOR TRAINING ON THE TAL_BY_8 TEMPLATE IN %.

Threshold	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Run 1	49.30	48.58	47.86	53.40	63.06	61.94	72.58	62.50	68.30	68.79	79.73	69.28	75.66	77.34	73.03	76.92	76.92	75.66
Run 2	52.06	46.39	51.38	39.30	58.44	71.66	62.50	70.72	67.80	71.66	77.74	67.29	73.48	72.12	71.66	69.77	70.72	72.12
Run 3	51.38	46.39	54.70	42.55	62.50	61.94	74.80	69.28	75.24	67.80	76.92	73.93	72.12	77.74	74.80	72.12	75.66	76.51
Run 4	44.12	44.89	48.58	55.34	64.15	66.27	66.78	70.72	70.72	67.80	78.15	72.12	79.34	73.03	78.15	63.06	70.72	73.03
Run 5	49.30	50.00	40.13	52.73	65.22	66.27	67.29	70.72	66.78	72.58	74.37	76.92	80.50	78.55	78.55	73.93	78.15	72.12
Run 6	46.39	50.00	53.40	54.70	63.60	67.80	68.30	76.09	74.37	70.25	73.03	76.51	76.51	79.73	76.09	81.63	76.09	71.19
Run 7	51.38	49.30	54.70	52.06	61.37	61.94	63.06	73.93	67.29	73.03	75.24	72.12	73.48	70.25	78.15	77.74	74.80	76.09
Run 8	54.05	41.76	50.70	47.13	57.22	62.50	69.28	68.79	66.27	73.48	73.03	72.12	76.92	70.25	74.37	75.66	74.37	73.03
Run 9	52.73	49.30	54.70	43.34	58.44	71.19	65.74	72.58	72.58	76.51	72.58	73.03	76.09	79.34	80.12	74.80	63.06	65.74
Run 10	57.22	55.98	44.12	47.13	60.22	61.37	65.74	64.69	66.78	73.48	73.93	73.93	79.73	74.80	75.66	78.55	76.92	73.03
Run 11	54.05	44.12	48.58	50.70	57.22	57.22	70.25	70.72	66.27	73.93	80.88	76.51	77.74	79.34	75.66	75.24	75.24	78.15
Run 12	51.38	53.40	51.38	41.76	63.60	69.28	69.77	64.69	71.66	68.30	70.72	76.51	78.15	75.24	78.95	75.66	71.19	68.79
Run 13	48.58	51.38	47.13	47.86	66.27	62.50	69.28	66.27	67.80	71.66	78.95	71.66	75.66	72.12	77.34	77.74	74.80	81.26
Run 14	47.86	57.84	39.30	52.06	60.22	61.94	65.22	71.66	69.77	71.66	77.74	75.66	76.09	77.74	70.72	77.74	71.66	68.30
Run 15	50.70	52.06	49.30	54.05	55.98	72.12	61.94	76.92	73.48	72.12	78.15	73.48	76.09	78.15	73.03	73.93	71.66	67.80
Run 16	49.30	52.73	44.89	54.70	64.15	70.72	76.09	72.12	73.93	70.72	76.92	76.92	76.51	76.09	81.26	69.77	70.72	71.66
Run 17	47.13	50.70	43.34	43.34	59.63	66.27	65.22	68.30	69.28	66.78	77.34	78.55	73.03	77.34	79.34	74.37	75.66	70.72
Run 18	48.58	44.12	54.05	49.30	62.50	66.78	69.77	71.66	65.74	76.92	73.48	79.34	72.58	79.73	82.00	71.19	76.09	71.19
Run 19	52.06	49.30	47.86	49.30	60.22	64.15	67.80	66.27	67.29	68.79	76.09	78.15	70.25	78.55	81.26	74.80	78.15	74.37
Run 20	46.39	39.30	44.12	50.00	63.60	62.50	70.72	65.74	67.29	69.77	80.50	74.37	79.34	77.74	82.37	75.66	72.12	72.12
Run 21	54.05	47.13	48.58	45.64	62.50	64.15	70.72	67.29	69.28	76.51	76.51	74.37	80.88	74.80	76.09	72.58	71.66	66.27
Run 22	59.04	49.30	47.13	47.13	64.69	65.22	62.50	65.74	76.09	69.28	77.74	76.09	72.58	75.66	71.66	77.74	69.28	66.78
Run 23	42.55	52.06	50.70	48.58	62.50	67.80	72.12	69.28	64.15	66.78	71.19	78.55	79.34	80.12	72.12	68.30	76.51	75.66
Run 24	55.34	39.30	48.58	48.58	66.27	69.77	76.51	66.78	65.22	68.79	71.66	74.80	75.24	69.28	73.48	69.28	73.93	65.22
Run 25	43.34	41.76	54.70	46.39	67.80	69.28	71.19	71.19	67.29	73.03	73.93	78.95	72.12	78.55	78.55	73.93	73.03	70.25
Run 26	47.86	48.58	55.34	45.64	65.74	66.78	68.79	67.80	72.12	73.03	70.72	68.79	76.09	78.55	75.66	70.72	75.66	76.92
Run 27	53.40	51.38	51.38	47.13	64.15	64.69	73.48	64.15	71.66	68.79	69.28	72.12	73.03	74.37	71.66	76.51	73.93	73.03
Run 28	46.39	52.73	50.70	47.86	63.06	69.77	72.12	73.48	69.28	71.66	77.74	72.58	71.19	76.92	82.37	74.80	71.19	73.03
Run 29	47.86	46.39	47.13	50.70	55.34	64.15	65.22	73.48	73.48	68.79	72.58	74.80	81.26	77.34	73.48	74.37	73.93	73.48
Run 30	49.30	50.00	50.00	48.58	59.04	66.78	71.66	74.80	69.77	67.80	80.88	75.66	76.09	77.74	71.19	72.12	71.66	76.09
Mean	50.10	48.54	49.15	48.57	61.96	65.82	68.88	69.61	69.57	71.02	75.59	74.50	75.90	76.28	76.29	74.02	73.52	72.32
Std dev	0.0386	0.0442	0.0418	0.0401	0.0317	0.0361	0.0390	0.0366	0.0317	0.0281	0.0323	0.0299	0.0299	0.0300	0.0360	0.0365	0.0309	0.0379

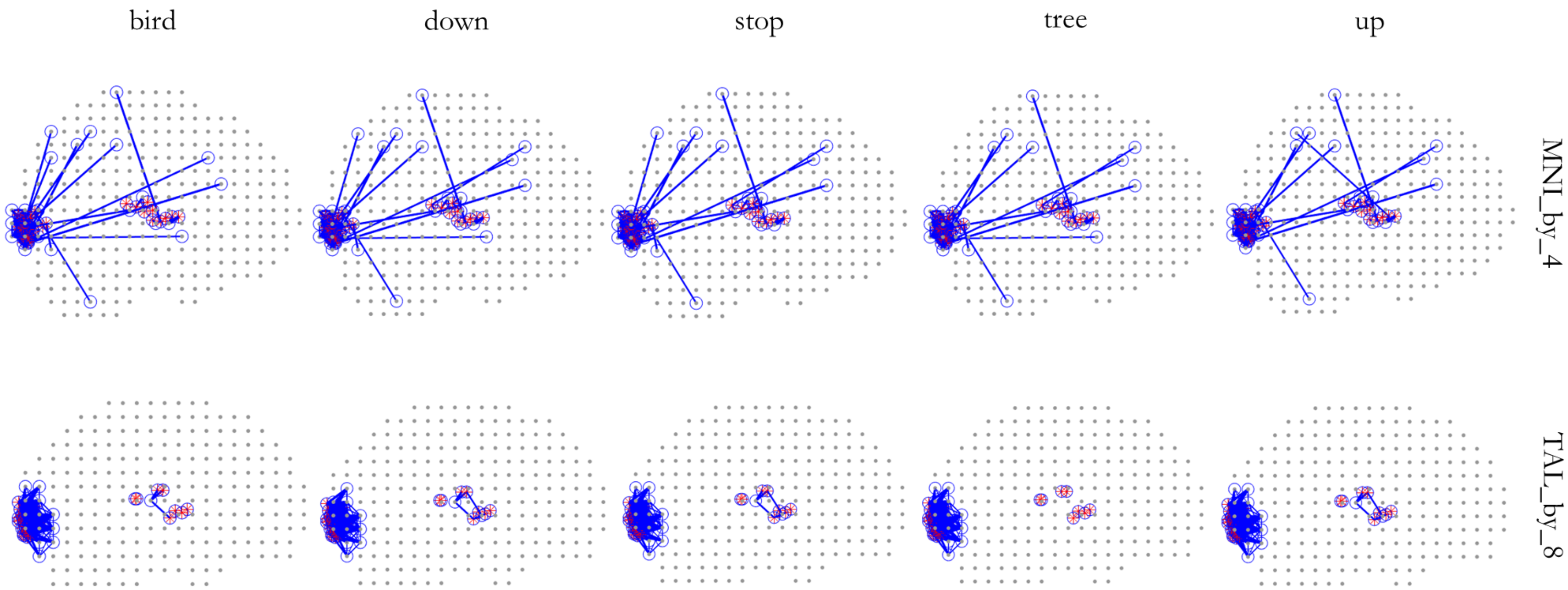


FIGURE B-1: VISUALISATION OF THE ACTIVATED NEURONS AND CONNECTIONS FOR THE FIVE CLASSES OF THE SIGN LANGUAGE DATASET IN THE TWO BEST MODELS TRAINED ON THE COMBINED DATA. THE VIEWING ANGLE IS FROM THE RIGHT SIDE OF THE BRAIN.

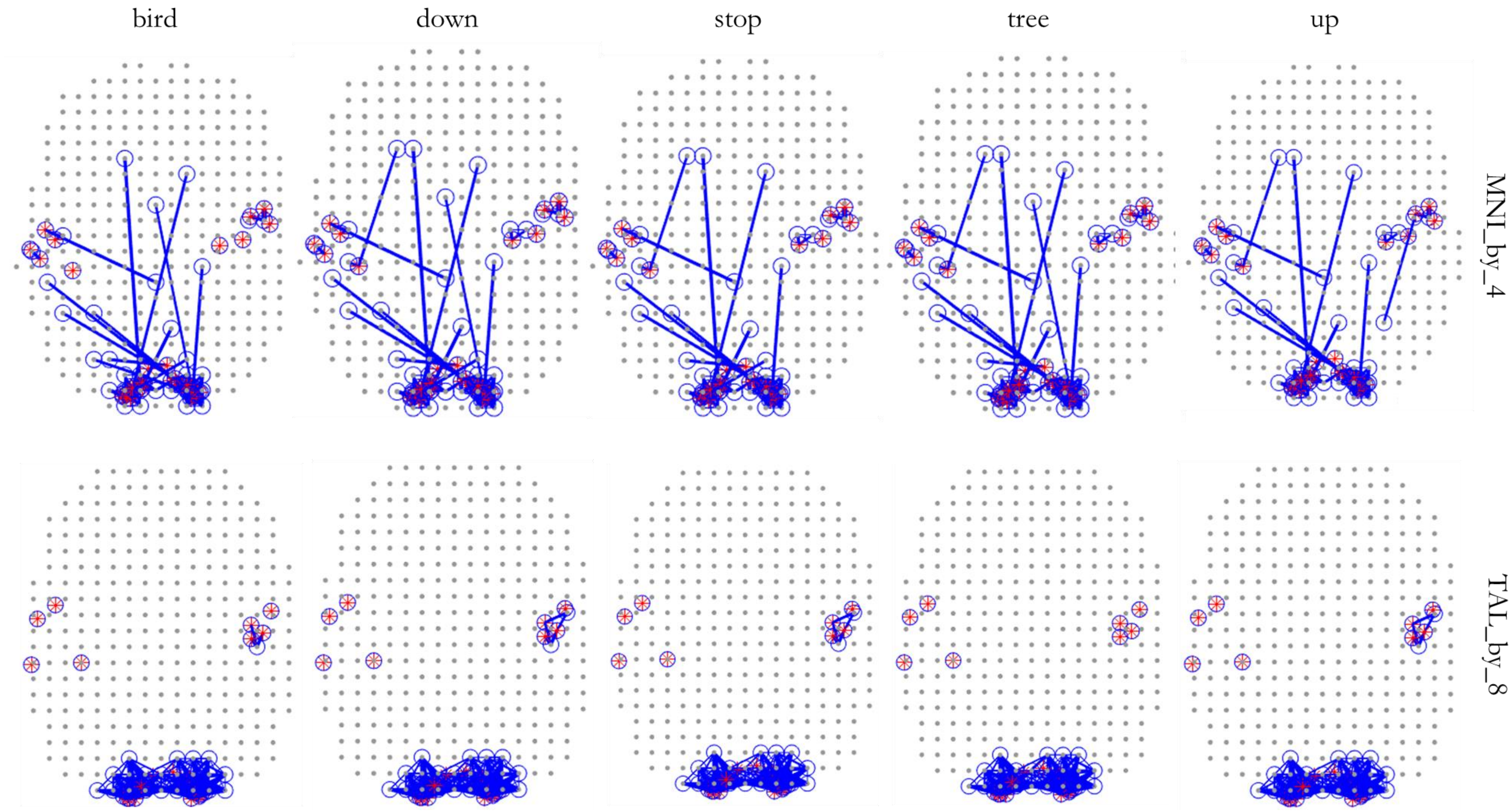


FIGURE B-2: VISUALISATION OF THE ACTIVATED NEURONS AND CONNECTIONS FOR THE FIVE CLASSES OF THE SIGN LANGUAGE DATASET IN THE TWO BEST MODELS TRAINED ON THE COMBINED DATA. THE VIEWING ANGLE IS FROM THE TOP OF THE BRAIN.

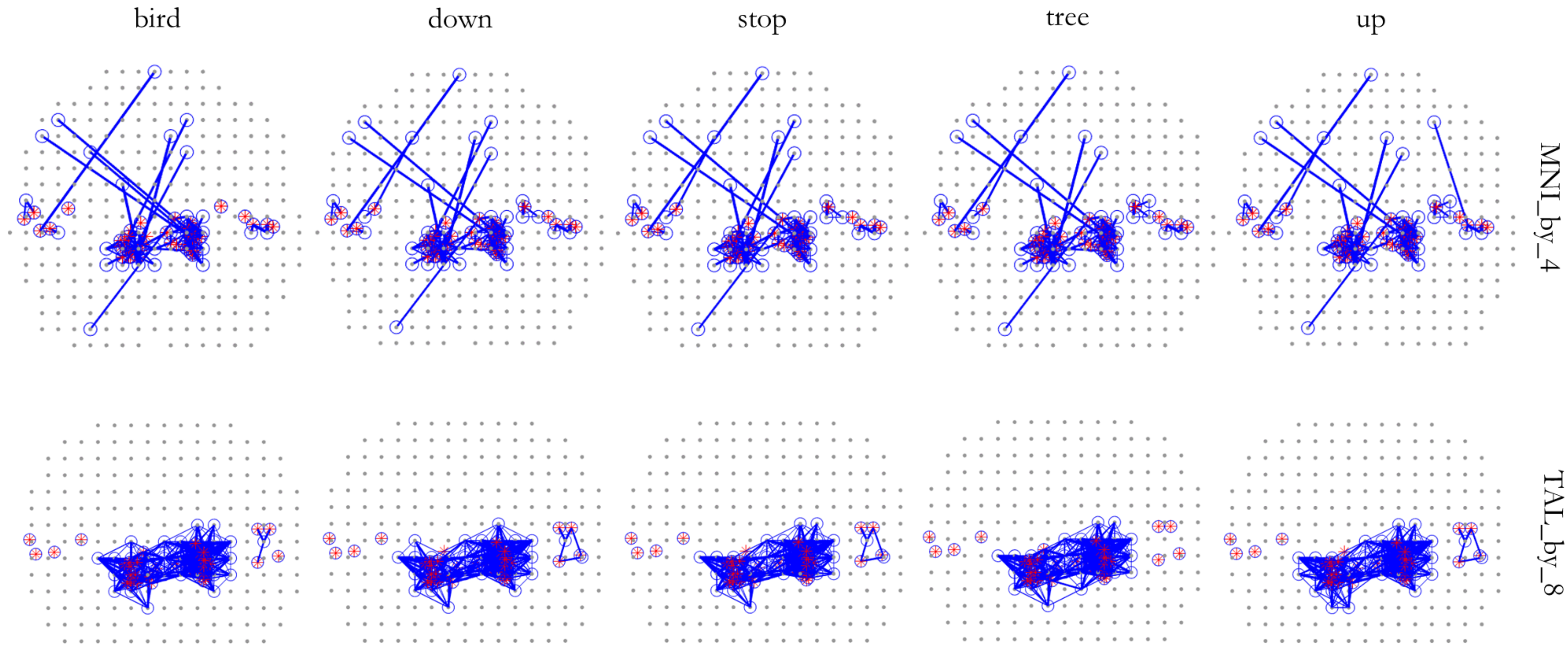


FIGURE B-3: VISUALISATION OF THE ACTIVATED NEURONS AND CONNECTIONS FOR THE FIVE CLASSES OF THE SIGN LANGUAGE DATASET IN THE TWO BEST MODELS TRAINED ON THE COMBINED DATA. THE VIEWING ANGLE IS FROM THE BACK OF THE BRAIN.

APPENDIX C   COPYRIGHT LETTERS FOR FIGURES

Figure 3-1 (page 55)

Copyright Clearance Center RightsLink®

Home | Help | Email Support | Anne Wendt

Chapter: Chapter 36 Auditory System
Book: The Human Nervous System
Author: Katrin Amunts,Patricia Morosan,Heidegard Hilbig,Karl Zilles
Publisher: Elsevier
Date: 2012

Order Completed

Thank you for your order.

This Agreement between Ms. Anne Wendt ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4776940234834
License date	Feb 27, 2020

Printable Details

Licensed Content

Licensed Content Publisher	Elsevier
Licensed Content Publication	Elsevier Books
Licensed Content Title	The Human Nervous System
Licensed Content Author	Katrin Amunts,Patricia Morosan,Heidegard Hilbig,Karl Zilles
Licensed Content Date	Jan 1, 2012
Licensed Content Pages	31
Journal Type	S&T

Order Details

Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
Format	print
Are you the author of this Elsevier chapter?	No
Will you be translating?	No

About Your Work

Title	Brain-inspired audio-visual information processing using spiking neural networks
Institution name	Auckland University of Technology
Expected presentation date	Nov 2020

Additional Data

Portions	Figure 36.1

Requestor Location

Requestor Location	Ms. Anne Wendt KEDRI - AUT Tower 2 Wakefield Street Auckland, 1010 New Zealand Attn: Ms. Anne Wendt

Tax Details

Publisher Tax ID	GB 494 6272 12

Figure 3-2 (page 56)

Copyright Clearance Center RightsLink®

Home | Help | Email Support | Anne Wendt

Chapter: 8 Head and Neck
Book: Gray's Anatomy for Students
Author: Richard L. Drake,A. Wayne Vogl,Adam W.M. Mitchell
Publisher: Elsevier
Date: 2020

Order Completed

Thank you for your order.

This Agreement between Ms. Anne Wendt ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4784600464031
License date	Mar 09, 2020

Printable Details

Licensed Content

Licensed Content Publisher	Elsevier
Licensed Content Publication	Elsevier Books
Licensed Content Title	Gray's Anatomy for Students
Licensed Content Author	Richard L. Drake,A. Wayne Vogl,Adam W.M. Mitchell
Licensed Content Date	Jan 1, 2020
Licensed Content Pages	303
Journal Type	S&T

Order Details

Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
Format	both print and electronic
Are you the author of this Elsevier chapter?	No
Will you be translating?	No

About Your Work

Title	Brain-inspired audio-visual information processing using spiking neural networks
Institution name	Auckland University of Technology
Expected presentation date	Nov 2020

Additional Data

Portions	Figure 8.130

Requestor Location

Requestor Location	Ms. Anne Wendt KEDRI - AUT Tower 2 Wakefield Street Auckland, 1010 New Zealand Attn: Ms. Anne Wendt

Tax Details

Publisher Tax ID	GB 494 6272 12

Figure 3-5 (page 60) and Figure 3-11 (page 71)

[Posit Science] Re: Dear Team at brainHQ, I am a doctoral student in computer science at Auckland University of Technology in New Zealand. I am currently writing my...

Seth Oakley <support@positscience.com>
Fri 25/05/2018 02:16
To: Anne Wendt <anne.wendt@aut.ac.nz>

Please do not write below this line

Dear Anne Wendt,

We believe we have successfully addressed your recent request for support from Posit Science (#144665). The complete ticket history with all comments can be reviewed below. If we have not successfully addressed your help submission, our apologies. To re-open your ticket please reply to this message.

Seth Oakley (Posit Science)
May 24, 7:16 AM PDT

Hello Anne Wendt,

Thank you for contacting us! You have our permission to use these images in the manner that you described. Please let us know if you have any other questions about this.

Regards,
Seth Oakley
Posit Science
San Francisco, CA

Anne Wendt
May 23, 4:55 PM PDT

Dear Team at brainHQ,

I am a doctoral student in computer science at Auckland University of Technology in New Zealand. I am currently writing my thesis on the topic of Brain-Inspired Audio-Visual Information Processing using Spiking Neural Networks, which will include a chapter describing the biological processes of hearing and vision. I was wondering if it would be possible for me to use two of your images in my thesis, in particular of the auditory (https://www.brainhq.com/sites/default/files/auditory-pathway.jpg) and visual pathways (https://www.brainhq.com/sites/default/files/visual-pathway%281%29.jpg) in the brain. The thesis itself will eventually be made available to two or three examiners, and then through our university's library website. There will be no other publications using your images. I would, of course, attribute the source of the images and link to your website in the figure captions. Please let me know if this would be okay for you.

With kind regards,

Anne Wendt

This email is a service from Posit Science. Delivered by Posit Science Customer Support.

Figure 3-6 (page 61)

Copyright Clearance Center RightsLink® · Help Email Support Anne Wendt

Mechanisms of Sound Localization in Mammals

Author: Benedikt Grothe, Michael Pecka, David McAlpine
Publication: Physiological Reviews
Publisher: The American Physiological Society
Date: Jul 1, 2010

Order Completed

Thank you for your order.

This Agreement between Anne Wendt ("You") and The American Physiological Society ("The American Physiological Society") consists of your order details and the terms and conditions provided by The American Physiological Society and Copyright Clearance Center.

License number	Reference confirmation email for license number
License date	Mar, 11 2020

Licensed Content

Licensed Content Publisher	The American Physiological Society
Licensed Content Publication	Physiological Reviews
Licensed Content Title	Mechanisms of Sound Localization in Mammals
Licensed Content Author	Benedikt Grothe, Michael Pecka, David McAlpine
Licensed Content Date	Jul 1, 2010
Licensed Content Volume	90
Licensed Content Issue	3

Order Details

Type of Use	Thesis/Dissertation
Requestor type	non-profit academic/educational
Readers being charged a fee for this work	No
Format	print and electronic
Portion	figures/tables/images
Number of figures/tables/images	2
Will you be translating?	no
World Rights	no

About Your Work

Title	Brain-inspired audio-visual information processing using spiking neural networks
Institution name	Auckland University of Technology
Expected presentation date	Nov 2020

Additional Data

Portions	Figures 2B and 2C
Attachment	Adapted_Figure.png

Requestor Location

Requestor Location	Ms. Anne Wendt KEDRI - AUT Tower 2 Wakefield Street Auckland, 1010 New Zealand Attn: Ms. Anne Wendt

Tax Details

Figure 3-7 (page 63)

Copyright Clearance Center Marketplace™

Order Number: 1025111
Order Date: 27 Mar 2020

Payment Information

Anne Wendt
anne.wendt@aut.ac.nz
Payment method: Invoice

Billing Address:
Ms. Anne Wendt
KEDRI - AUT Tower
2 Wakefield Street
Auckland, 1010
New Zealand

+64 223919063
anne.wendt@aut.ac.nz

Customer Location:
Ms. Anne Wendt
KEDRI - AUT Tower
2 Wakefield Street
Auckland, 1010
New Zealand

Order Details

1. Principles of neural science

Billing Status: Open

Order license ID	1025111-1
Order detail status	Completed
ISBN-13	9780071390118
Type of use	Republish in a thesis/dissertation
Publisher	MCGRAW-HILL
Portion	Chart/graph/table/figure

0.00 USD

LICENSED CONTENT

Publication Title	Principles of neural science	**Language**	English
Author/Editor	KANDEL, ERIC R.	**Rightsholder**	McGraw Hill LLC
Date	01/01/2013	**Publication Type**	Book

REQUEST DETAILS

Portion Type	Chart/graph/table/figure	**Distribution**	Worldwide
Number of charts / graphs / tables / figures requested	1	**Translation**	Original language of publication
Format (select all that apply)	Print,Electronic	**Copies for the disabled?**	No
Who will republish the content?	Academic institution	**Minor editing privileges?**	No
Duration of Use	Life of current edition	**Incidental promotional use?**	No
Lifetime Unit Quantity	Up to 499	**Currency**	USD
Rights Requested	Main product		

NEW WORK DETAILS

Title	Brain-inspired audio-visual information processing using spiking neural networks	**Institution name**	Auckland University of Technology
		Expected presentation date	2020-11-01
Instructor name	Anne Wendt		

ADDITIONAL DETAILS

The requesting person / organization to appear on the license	Anne Wendt

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Figure 31-13	**Title of the article/chapter the portion is from**	The auditory central nervous system
Editor of portion(s)	Kandel, Eric R	**Author of portion(s)**	Oertel, Donata & Doupe, Allison J
Volume of serial or monograph	N/A	**Publication date of portion**	2013-01-01
Page or page range of portion	704		

McGraw Hill LLC Terms and Conditions

It is the responsibility of the user to verify that the text, figures, images and other material that they wish to use are owned by McGraw Hill LLC. McGraw Hill LLC can only grant permission for the material that they own. Please check the caption of any figures, images, charts, etc., as well as any credits pages at the front or back of the book. McGraw Hill LLC may restrict the portion of the title that may be used.

Total Items: 1

Subtotal: 0.00 USD

Order Total: 0.00 USD

Figure 3-8 (page 66)

Copyright Clearance Center RightsLink®

Home | Help | Email Support | Anne Wendt

Chapter: 8 Head and Neck
Book: Gray's Anatomy for Students
Author: Richard L. Drake,A. Wayne Vogl,Adam W.M. Mitchell
Publisher: Elsevier
Date: 2020

Order Completed

Thank you for your order.

This Agreement between Ms. Anne Wendt ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number 4786150370603 — Printable Details

License date Mar 11, 2020

Licensed Content

Licensed Content Publisher	Elsevier
Licensed Content Publication	Elsevier Books
Licensed Content Title	Gray's Anatomy for Students
Licensed Content Author	Richard L. Drake,A. Wayne Vogl,Adam W.M. Mitchell
Licensed Content Date	Jan 1, 2020
Licensed Content Pages	303
Journal Type	S&T

Order Details

Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
Format	both print and electronic
Are you the author of this Elsevier chapter?	No
Will you be translating?	No

About Your Work

Title	Brain-inspired audio-visual information processing using spiking neural networks
Institution name	Auckland University of Technology
Expected presentation date	Nov 2020

Additional Data

Portions	Figure 8.108

Requestor Location

Requestor Location	Ms. Anne Wendt KEDRI - AUT Tower 2 Wakefield Street Auckland, 1010 New Zealand Attn: Ms. Anne Wendt

Tax Details

Publisher Tax ID	GB 494 6272 12

Figure 3-9 (page 67)

Copyright Clearance Center RightsLink®

Home | Help | Email Support | Anne Wendt

Diabetes and retinal vascular disorders: role of the renin–angiotensin system

CAMBRIDGE JOURNALS

Author: Jennifer L. Wilkinson-Berka
Publication: Expert Reviews in Molecular Medicine
Publisher: Cambridge University Press
Date: Sep 23, 2004

Order Completed

Thank you for your order.

This Agreement between Ms. Anne Wendt ("You") and Cambridge University Press ("Cambridge University Press") consists of your license details and the terms and conditions provided by Cambridge University Press and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4786221219964
License date	Mar 11, 2020

Printable Details

Licensed Content

Licensed Content Publisher	Cambridge University Press
Licensed Content Publication	Expert Reviews in Molecular Medicine
Licensed Content Title	Diabetes and retinal vascular disorders: role of the renin–angiotensin system
Licensed Content Author	Jennifer L. Wilkinson-Berka
Licensed Content Date	Sep 23, 2004
Licensed Content Volume	6
Licensed Content Issue	15
Start page	1
End page	18

Order Details

Type of Use	Dissertation/Thesis
Requestor type	Not-for-profit
Portion	Text extract
Number of pages requested	1

About Your Work

Title	Brain-inspired audio-visual information processing using spiking neural networks
Institution name	Auckland University of Technology
Expected presentation date	Nov 2020

Additional Data

Portions	Figure 3
Territory for reuse	World

Figure 3-10 (page 68)

Taylor & Francis Group LLC - Books - License Terms and Conditions

Order Date	21-Mar-2020
Order license ID	1024076-1
ISBN-13	9780849396687
Type of Use	Republish in a thesis/dissertation
Publisher	TAYLOR & FRANCIS GROUP LLC
Portion	Chart/graph/table/figure

LICENSED CONTENT

Publication Title	Cyclic nucleotide phosphodiesterases in health and disease	**Country**	United States of America
Author/Editor	BEAVO, JOSEPH	**Rightsholder**	Taylor & Francis Group LLC - Books
Date	01/01/2007	**Publication Type**	Book
Language	English		

REQUEST DETAILS

Portion Type	Chart/graph/table/figure	**Distribution**	Worldwide
Number of charts / graphs / tables / figures requested	1	**Translation**	Original language of publication
Format (select all that apply)	Electronic	**Copies for the disabled?**	No
Who will republish the content?	Academic institution	**Minor editing privileges?**	Yes
Duration of Use	Life of current edition	**Incidental promotional use?**	No
Lifetime Unit Quantity	Up to 499	**Currency**	USD
Rights Requested	Main product		

NEW WORK DETAILS

Title	Brain-inspired audio-visual information processing using spiking neural networks	**Institution name**	Auckland University of Technology
Instructor name	Anne Wendt	**Expected presentation date**	2020-11-01

ADDITIONAL DETAILS

Order reference number	N/A	**The requesting person / organization to appear on the license**	Anne Wendt

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Figure 8.1	**Title of the article/chapter the portion is from**	Photoreceptor Phosphodiesterase (PDE6): A G-Protein-Activated PDE Regulating Visual Excitation in Rod and Cone Photoreceptor Cells
Editor of portion(s)	Beavo, Joseph		
Volume of serial or monograph	N/A		
Page or page range of portion	167	**Author of portion(s)**	Cote, Rick H.
		Issue, if republishing an article from a serial	N/A
		Publication date of portion	2007-01-01

Figure 3-12 (page 73)

Copyright Clearance Center Marketplace™

Order Number: 1024624
Order Date: 25 Mar 2020

Payment Information

Anne Wendt
anne.wendt@aut.ac.nz
Payment method: Invoice

Billing Address:
Ms. Anne Wendt
KEDRI - AUT Tower
2 Wakefield Street
Auckland, 1010
New Zealand

+64 223919063
anne.wendt@aut.ac.nz

Customer Location:
Ms. Anne Wendt
KEDRI - AUT Tower
2 Wakefield Street
Auckland, 1010
New Zealand

Order Details

1. Principles of neural science

Billing Status: Open

Order license ID	1024624-1
Order detail status	Completed
ISBN-13	9780071390118
Type of use	Republish in a thesis/dissertation
Publisher	MCGRAW-HILL
Portion	Chart/graph/table/figure

0.00 USD

LICENSED CONTENT

Publication Title	Principles of neural science	**Language**	English
Author/Editor	KANDEL, ERIC R.	**Rightsholder**	McGraw Hill LLC
Date	01/01/2013	**Publication Type**	Book

REQUEST DETAILS

Portion Type	Chart/graph/table /figure	**Distribution**	Worldwide
Number of charts / graphs / tables / figures requested	1	**Translation**	Original language of publication
Format (select all that apply)	Print,Electronic	**Copies for the disabled?**	No
Who will republish the content?	Academic institution	**Minor editing privileges?**	No
Duration of Use	Life of current edition	**Incidental promotional use?**	No
Lifetime Unit Quantity	Up to 499	**Currency**	USD
Rights Requested	Main product		

NEW WORK DETAILS

Title	Brain-inspired audio-visual information processing using spiking neural networks	**Institution name**	Auckland University of Technology
		Expected presentation date	2020-11-01
Instructor name	Anne Wendt		

ADDITIONAL DETAILS

The requesting person / organization to appear on the license	Anne Wendt

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Figure 25-7 B	**Title of the article/chapter the portion is from**	The Constructive Nature of Visual Processing
Editor of portion(s)	Kandel, Eric R.	**Author of portion(s)**	Gilbert, Charles D.
Volume of serial or monograph	N/A	**Publication date of portion**	2013-01-01
Page or page range of portion	563		

McGraw Hill LLC Terms and Conditions

It is the responsibility of the user to verify that the text, figures, images and other material that they wish to use are owned by McGraw Hill LLC. McGraw Hill LLC can only grant permission for the material that they own. Please check the caption of any figures, images, charts, etc., as well as any credits pages at the front or back of the book. McGraw Hill LLC may restrict the portion of the title that may be used.

Total Items: 1

Subtotal: 0.00 USD

Order Total: 0.00 USD

Figure 3-13 (page 78)

Copyright Clearance Center Marketplace™

Order Number: 1028049
Order Date: 13 Apr 2020

Payment Information

Anne Wendt
anne.wendt@aut.ac.nz
Payment method: Invoice

Billing Address:
Ms. Anne Wendt
KEDRI - AUT Tower
2 Wakefield Street
Auckland, 1010
New Zealand

+64 223919063
anne.wendt@aut.ac.nz

Customer Location:
Ms. Anne Wendt
KEDRI - AUT Tower
2 Wakefield Street
Auckland, 1010
New Zealand

Order Details

1. Principles of neural science

Billing Status: Open

Order license ID	1028049-1
Order detail status	Completed
ISBN-13	9780071390118
Type of use	Republish in a thesis/dissertation
Publisher	MCGRAW-HILL
Portion	Chart/graph/table/figure

0.00 USD
Republication Permission

LICENSED CONTENT

Publication Title	Principles of neural science	**Language**	English
Author/Editor	KANDEL, ERIC R.	**Rightsholder**	McGraw Hill LLC
Date	01/01/2013	**Publication Type**	Book

REQUEST DETAILS

Portion Type	Chart/graph/table /figure	**Distribution**	Worldwide
Number of charts / graphs / tables / figures requested	1	**Translation**	Original language of publication
Format (select all that apply)	Print,Electronic	**Copies for the disabled?**	No
Who will republish the content?	Academic institution	**Minor editing privileges?**	No
Duration of Use	Life of current edition	**Incidental promotional use?**	No
Lifetime Unit Quantity	Up to 499	**Currency**	USD
Rights Requested	Main product		

NEW WORK DETAILS

Title	Brain-inspired audio-visual information procesing using spiking neural networks	**Institution name**	Auckland University of Technology
		Expected presentation date	2020-11-01
Instructor name	Anne Wendt		

ADDITIONAL DETAILS

The requesting person / organization to appear on the license	Anne Wendt

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Figure 18-3	**Title of the article/chapter the portion is from**	The Organization of Cognition
Editor of portion(s)	Kandel, Eric R	**Author of portion(s)**	Olson, Carl R. & Colby, Carol L.
Volume of serial or monograph	N/A	**Publication date of portion**	2013-01-01
Page or page range of portion	397		

McGraw Hill LLC Terms and Conditions

It is the responsibility of the user to verify that the text, figures, images and other material that they wish to use are owned by McGraw Hill LLC. McGraw Hill LLC can only grant permission for the material that they own. Please check the caption of any figures, images, charts, etc., as well as any credits pages at the front or back of the book. McGraw Hill LLC may restrict the portion of the title that may be used.

Total Items: 1

Subtotal: 0.00 USD

Order Total: 0.00 USD

Figure 5-7 (page 126)

Copyright Clearance Center RightsLink®

Home | Help | Email Support | Anne Wendt

Tonotopic mapping of human auditory cortex

Author: Melissa Saenz,Dave R.M. Langers

Publication: Hearing Research

Publisher: Elsevier

Date: January 2014

Order Completed

Thank you for your order.

This Agreement between Ms. Anne Wendt ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number 4924540355414 — Printable Details

License date Oct 08, 2020

Licensed Content

Licensed Content Publisher	Elsevier
Licensed Content Publication	Hearing Research
Licensed Content Title	Tonotopic mapping of human auditory cortex
Licensed Content Author	Melissa Saenz,Dave R.M. Langers
Licensed Content Date	Jan 1, 2014
Licensed Content Volume	307
Licensed Content Issue	n/a
Licensed Content Pages	11
Journal Type	S&T

Order Details

Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No

About Your Work

Title	Brain-inspired audio-visual information processing using spiking neural networks
Institution name	Auckland University of Technology
Expected presentation date	Nov 2020

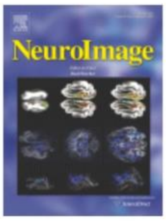
Additional Data

Portions	Figure 1

Figure 5-8 (page 127)

Copyright Clearance Center RightsLink®

Home | Help | Email Support | Anne Wendt



Neuroimaging paradigms for tonotopic mapping (II): The influence of acquisition protocol

Author:
Dave R.M. Langers,Rosa M. Sanchez-Panchuelo,Susan T. Francis,Katrin Krumbholz,Deborah A. Hall

Publication: NeuroImage

Publisher: Elsevier

Date: 15 October 2014

Order Completed

Thank you for your order.

This Agreement between Ms. Anne Wendt ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4924541020657
License date	Oct 08, 2020

Printable Details

Licensed Content

Licensed Content Publisher	Elsevier
Licensed Content Publication	NeuroImage
Licensed Content Title	Neuroimaging paradigms for tonotopic mapping (II): The influence of acquisition protocol
Licensed Content Author	Dave R.M. Langers,Rosa M. Sanchez-Panchuelo,Susan T. Francis,Katrin Krumbholz,Deborah A. Hall
Licensed Content Date	Oct 15, 2014
Licensed Content Volume	100
Licensed Content Issue	n/a
Licensed Content Pages	13
Journal Type	S&T

Order Details

Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables /illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No

About Your Work

Title	Brain-inspired audio-visual information processing using spiking neural networks
Institution name	Auckland University of Technology
Expected presentation date	Nov 2020

Additional Data

Portions	Figure 4b

Requestor Location

Requestor Location	Ms. Anne Wendt KEDRI - AUT Tower 2 Wakefield Street Auckland, 1010 New Zealand Attn: Ms. Anne Wendt

Tax Details

Publisher Tax ID	GB 494 6272 12

$ Price

Total	0.00 USD

Total: 0.00 USD

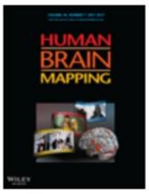
CLOSE WINDOW

ORDER MORE

Figure 5-16 (page 152)

Copyright Clearance Center RightsLink®

Home | Help | Email Support | Anne Wendt



Retinotopic mapping of the peripheral visual field to human visual cortex by functional magnetic resonance imaging

Author: Qiyong Guo, Fengzhe Jin, Zhen Zhang, et al

Publication: Human Brain Mapping

Publisher: John Wiley and Sons

Date: Mar 22, 2012

Order Completed

Thank you for your order.

This Agreement between Ms. Anne Wendt ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

Your confirmation email will contain your order number for future reference.

License Number	4927110747836	Printable Details
License date	Oct 13, 2020	

Licensed Content

Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Human Brain Mapping
Licensed Content Title	Retinotopic mapping of the peripheral visual field to human visual cortex by functional magnetic resonance imaging
Licensed Content Author	Qiyong Guo, Fengzhe Jin, Zhen Zhang, et al
Licensed Content Date	Mar 22, 2012
Licensed Content Volume	33
Licensed Content Issue	7
Licensed Content Pages	14

Order Details

Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Will you be translating?	No

About Your Work

Title	Brain-inspired audio-visual information processing using spiking neural networks
Institution name	Auckland University of Technology
Expected presentation date	Nov 2020

Additional Data

Portions	Figure 10

Figure 5-17 (page 153)

Copyright Clearance Center RightsLink®

My Orders My Library My Profile Welcome anne.wendt@aut.ac.nz Log out | !

My Orders > Orders > All Orders

The RightsLink Application will be unavailable from 10PM EST on Friday, October 16th through 12:00 AM EST Saturday, October 17th.

License Details

This Agreement between Ms. Anne Wendt ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springe Nature and Copyright Clearance Center.

Print Copy

License Number	4927080024663
License date	Oct 13, 2020
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Springer eBook
Licensed Content Title	fMRI of Human Visual Pathways
Licensed Content Author	Edgar A. DeYoe, John L. Ulmer, Wade Mueller et al
Licensed Content Date	Jan 1, 2011
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	Brain-inspired audio-visual information processing using spiking neural networks
Institution name	Auckland University of Technology
Expected presentation date	Nov 2020
Portions	Figure 26.1
Requestor Location	Ms. Anne Wendt KEDRI - AUT Tower 2 Wakefield Street Auckland, 1010 New Zealand Attn: Ms. Anne Wendt
Total	**0.00 USD**

BACK

Figure 5-20 (page 158)

From: Jonathan A Winawer
To: Anne Wendt
Subject: Re: Reuse of your figure for population receptive field sizes
Date: Saturday, 24 October 2020 9:47:57 PM

Sure - no problem.

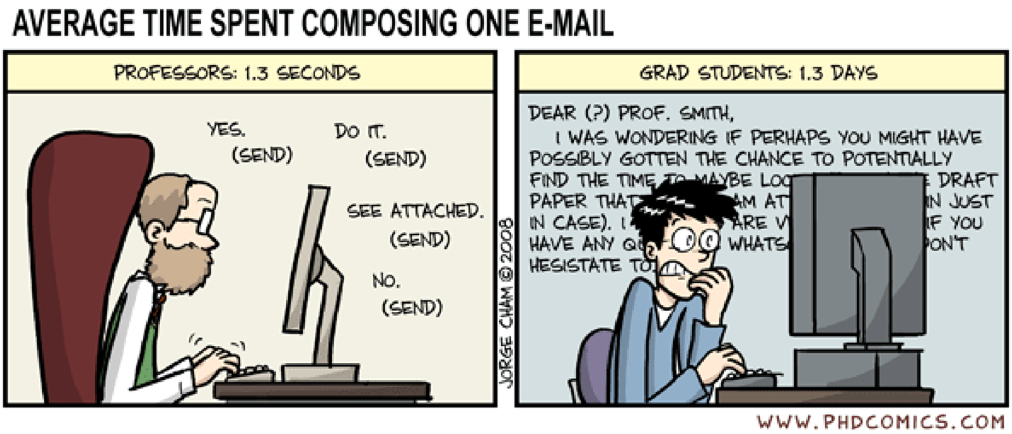
Jon

On Sat, Oct 24, 2020 at 02:30 Anne Wendt <anne.wendt@aut.ac.nz> wrote:

> Dear Professor Winawer,
>
> I am a PhD candidate at Auckland University of Technology in New Zealand and currently writing up my thesis on bio-inspired computational models for audio-visual data processing. Coming from Computer Science, I would like to include a slightly modified version (see attachment) of your figure from https://archive.nyu.edu/handle/2451/33887 in my thesis when explaining the concept of receptive fields. Would you please allow me to do this? I will, of course, include a link and citation to the original source.
>
> Kind regards
>
> Anne
>
> **Anne Wendt**
> PhD Candidate
> School of Engineering, Computer & Mathematical Sciences
> Auckland University of Technology
>
> New Zealand
>
> M +64 22 391 9063 E anne.wendt@aut.ac.nz W aut.ac.nz

--
Sent from Gmail Mobile

From "Piled Higher and Deeper" by Jorge Cham: (source; Reproduced with permission.)



The end.