Talking Books:

The Development of an Interactive, Educational, Digital Application

Emile Drescher

This exegesis is submitted to the Auckland University of Technology in partial fulfilment of the requirements for the degree of Master of Creative Technologies

July 2015

Abstract

This project explores the development of a digital application that pairs speech recognition technology with children's storytelling on tablet devices. The system engages children with active reading, positively reinforcing the reader through story progression, animations and sound effects when the narrative is correctly read out loud.

The research is located through a literature and contextual review, which establishes the significance and potential benefits of this system. Further, the practical development is coordinated through collaboration with co-creators, following a user-centred design methodology to create iterations of prototypes. These prototypes were developed, tested and refined through user feedback.

User feedback from the testing sessions revealed encouraging and positive feedback from educators and children users, but highlighted the crucial need for a robust speech recognition system that performs well with the intended child audience.

The outcomes of this research project include the creation of a digital storybook system that has potential for supporting early literacy development, commercial development opportunities, and a platform that facilitates co-creators to develop stories for the next generation of early readers. Documentation and insights into the development process, gained through this research may be of value to other application developers.

Table of contents

List of Figures

Attestation of authorship

I hereby declare that the submission is my own work and that to the best of my knowledge and belief, it contains no material previously published or written by any other person nor material which to a substantial extent has been accepted for the award of any other masters, degree or diploma of a university or other institution of higher learning, except where due acknowledgement is made in referencing.

Signed

Acknowledgements

Firstly I would like to acknowledge the support throughout this research project from my primary supervisor, Dr Frances Joseph. Furthermore, I am grateful to my secondary advisor Laurent Antonczak for his insightful perspectives.
Additionally my friend and research partner Sam Joe who endured my company over the duration of this project, and Lauren and Tom for their lovely support and encouragement.

This Thesis received ethics approval on the 17th of February 2015 filed under the application number 15/09.

1.0 Introduction

This research investigates, identifies and implements speech recognition, multimedia and storytelling for an interactive application on a tablet device. This software is aimed at early readers aged 6 - 8, and is designed to simultaneously engage and encourage users in their reading and comprehension of text. The software advances the stories narrative, synchronised with the users reading speed. The speech recognition recognises when the user has read the words correctly and updates the readers narrative progress in real-time. When the user finishes reading a sentence, they are rewarded with story progression in the form of animations and sound effects replacing the traditional page turning mechanism.

The investigation has taken a human centred design approach, informed through the development of prototypes, user testing and feedback. The findings from the research process have been incorporated into iterations of software prototypes. These iterations have been continually critiqued and tested to determine the most effective solutions.

While I have instigated and guided the project I have also worked in collaboration with other researchers, content creators and software developers. The outcomes of the research and collaboration include this written thesis and a proof of concept prototype application, which explores this new precedent for interactive storytelling.

This practice led research project addresses the following questions:

What are the uses and potential benefits of this system?
How does the user group respond to this system?
How does this system perform as an interface for experiencing interactive media?
What is the best way of designing this system for the intended user group?

This exegesis is organised in six parts. In Section 1 the research project is introduced and the rationale presented. In Section 2 the research is situated through a literature and contextual review, analysing surrounding areas including: Children's reading in relation to this project, Speech recognition technology iPad's and education, and case studies of relevant media.

In Section 3 the research methodology is discussed, explaining the process and overall plan behind the research project. Section 4 considers the project development through chronologically organised iterations of prototypes. The significance of these prototypes and the practical development related to them are discussed. In Section 5, the user testing and results are discussed, and Section 6 concludes the research, and considers future developments.

1.1 Rationale

Kate Wilson (2013) argues the need for beneficial content for younger audiences as children are increasingly spending more time with touch-screen devices. This research project explores a technological system that combines speech recognition with children's media on a tablet device. The proposed system does not exist in the public market for tablet software, and sets to create a new genre of digital storybooks for children.

Significant technical advances in the fields of speech recognition, and personal smart devices have enabled the system's viability. Adams (Adams, 2011) recognises the potential that exists for speech recognition being applied to education, noting that it could provide "corrective and enriched learning opportunities that are beyond the human capacities of the conventional classroom". This research expands on Adam's arguments by developing a practical application of this technology, with a focus on supporting early literacy development.

Through iterative cycles of research, design, practical development, user testing, and evaluation, the project looks to add to the body of knowledge surrounding digital application development, speech recognition, education and entertainment.

2.0 Literature and Contextual Review

2.1 Children's reading

The storybook content within the *Talking Books* application targets children as the main user group, though it could have important parallel functions in the field of adult literacy. The application aims to support early literacy development. Within this section I discuss current theories and ideas about children's reading and media in relation to the *Talking Books* project.

2.1.1 Reading, Storytelling and Narrative

Learning to read is recognised as an important precursor to language and literacy development. It is considered one of the most important attainments of the first years of formal education and is essential to success in schooling and virtually any future vocation. The American Academy of Pediatrics recommends reading regularly with young children, beginning from infancy, as it "stimulates optimal patterns of brain development, strengthens parent-child relationships at a critical time in child development, which, in turn, builds language, literacy, and social-emotional skills that last a lifetime" (Council on Early Childhood, 2014).

"For beginner readers, sounding out or decoding even the simplest words can be an arduous task. Yet, the very process of decoding a word leaves an interconnected trace in memory of its spelling and pronunciation. As these traces strengthen and overlap, they gradually enable the reader to recognize whole words and spelling patterns and to map them instantly to pronunciations" (Adams, 2011).

Talking Books engages children with interactive storybooks through the act of reading aloud. Children's storybooks are a familiar and accessible form of content that provide the context for children to develop language and literacy competencies. They have the potential of exposing children to rich language, meaningful plots, compelling characters and engaging illustrations.

Children's stories are usually comprised of short sentences, contain basic words and are read in a quiet setting. They are intentionally created this way so children can easily comprehend the story's narrative. Chaitin (2003) defines a story or narrative as a combination either of real or

imagined events that connect in such a way to provide a chain of events that are recounted to others. Storytelling succeeds in transporting children into worlds depicted by words and images, immersing and engaging them with content.

2.1.2 Children's interests and aesthetic

As children grow up their interests change. A storybook designed for a child aged 4 is unlikely to be suitable for a 6 year old, as cognitive development in children changes rapidly at this early stage in life. Storybook content for children is categorised by age, followed by the story's genre. Age takes precedent as it determines the approximate literacy level and cognitive capacity of the reader. Second to age, the child's gender significantly influences their reading interest[i]. Norvell & George (1958) states that as the child matures, the rate of change in his/her interests diminish and the child matures, the broader and deeper his/her range of interests.

Providing suitable reading material that matches children's interests is critical as it provides motivation to read. When a child is reading material of interest, engagement and comprehension increases, which improves achievement and reading levels (Guthrie & Wigfiels 2000). Research has shown that children who read about their interests are more likely to become life-long readers (Shafer, 2003). Children's interests are further influenced by their immediate social worlds including their parents, siblings and peers (Smith, 2004) as well as surrounding cultural images including popular culture and media. Bloomer (1960) identifies that the illustration style, use of colour, text formatting and front cover design affect book choice. Recognised authors, evidence of awards and recommendations by friends and family also influence a child's book decision process.

Reader's interest influences whether a child invests themselves in the application or not. Regardless of the applications potential benefits, a reader will not engage with the application if the content does not interest them. Therefore providing content that aligns to the target users' interests is an important consideration.

2.1.3 Contemporary children's storytelling

New forms of media are enabling innovative ways for children to engage with storytelling. With the advent of screens and personal tablet devices, the capabilities of storytelling have changed drastically.

Traditionally storytelling existed through oral dialogue passed from generation to generation. This served as a means of communication, a form of entertainment, and a way to share

common values and cultural history. Storytelling was also seen as a useful educative tool to teach children morals. It is an effective method of communicating a message, as the moral of the story is easily understood and remembered. (Colwell, 1980) The introduction of the printing press in the 15th century enabled oral stories to be transcribed and reproduced, advancing the education and growth of the reading public. Categories and genres of storytelling developed, promoting reading for pleasure and entertainment, as well as education.

The *Golden Age* for children's literature is considered to be in the mid-19th century, when there was a sudden surge in books created for children.[ii] Children's television emerged in the 1950's as a versatile medium for storytelling due to its audio-visual capability. Television became a highly popular form for children to experience media for decades. Today children experience storytelling through a variety of platforms. Devices such as smartphones and tablets that offer engaging interactive capabilities are becoming ubiquitous and are changing the way children experience storytelling (Common Sense Media & Rideout, 2011).

Analysing the current trends in children's media usage, and exploring the capabilities that these media platforms offer has been an important preliminary step in the development of this project. This project focuses on tablet technology, specifically the iPad (see section 2.3 for more information). Hall (2012, unpaged) speculates that a new generation may become more used to reading from an iPad or Kindle screen than from a traditional book. This reiterates the magnitude of the digital shift and its influence on children's storytelling.

2.2.0 Speech recognition technology

The core functionality within the *Talking Books* application uses speech recognition. In this section I will discuss the current state of this technology, and its role within the research project.

2.2.1 Introduction to speech recognition

Speech recognition is the translation of spoken word into text through computers. Communication through speech is the most immediate way for human beings to communicate with one another and is faster and more efficient than any alternative form of communication. Since initial research into a speech analysis and synthesis systems by Homer Dudley, speech recognition technology has been steadily developing.

The accuracy of speech recognition has been improving, corresponding with the exponential growth in computing power (Schaller, 1997) that has enabled speech recognition researchers to tackle harder problems, leading to a widespread adoption of this technology. The Chief Technology Officer of leading speech recognition software company *Nuance Communications,* Vlad Sejnoha states:

"I think we have passed a magical threshold of usability that means that you can pick up a device today and speak to it and expect to be understood" (Schneiderman, 2014).

Adams (2011) describes the wide range of applications of speech recognition technology today, however she argues that there is a salient exception where speech recognition is needed most; education. Adams suggests that speech recognition technology has great potential to be integrated into the classroom to provide ample reading and learning support for every student.

2.2.2 Understanding children as speakers.

Human speech is enabled through anatomical functions of the human body. The language we use to communicate is also formed and influenced by social and environmental factors. Human's speech is unique to the speaker as the factors that influence and construct it are all variable. Gupta & Wadhwa (2014) explain that vocalisation varies widely in terms of accent, pronunciation, articulation, roughness, nasality, pitch, volume and speed. These variability's in

human speech make speech analysis a highly complex problem. Despite these limitations, speech recognition can be a useful tool for a variety of applications, as long as designers and users fully understand the boundaries and weaknesses of such systems (Gybels, 2010).

Automatic speech recognition (ASR) technology has achieved highly accurate results for adult speech, but the field of ASR for children has lagged behind. This is due to the variability in the acoustic and linguistic characteristics of younger users. Conventional speech recognition systems are modelled on adult data and fail to perform satisfactorily with children's speech input. Children's voices are higher pitched, their elocution is idiosyncratic and their syntax is immature (Adams, 2011). The vocal tract length of the speaker affects the sound of the vowel production, and children's vocal tracts are still developing and are *much* shorter than adults. However there are mathematical techniques to normalise the effects of vocal tract length and adapt the adult acoustic model towards children speakers (Potamianos & Narayanan, 2003). Researchers from Nuance Communications developed a child specific recognition system and reported a word error rate (WER) improvement of 27.2% compared to the adult targeted model when testing with children (Gray, Willett, Pinto, Maergner, Bodenstab, 2014).

2.2.3 Choosing an appropriate speech recognition engine

The *Talking Books* application is being developed for Apple's iPad device, so a speech recognition engine that was compatible with Apple's mobile operating system was sought. Unfortunately, Apple's native speech recognition engine *Siri* is not available to developers, so an alternative third-party speech recognition engine was required. The device that the application is run on has its own technical restrictions, specifically a limited amount of memory and processing power. A speech recognition engine that was relatively small and efficient to account for these limitations was needed. Additionally, access to the speech recognition engine's phonetic dictionaries, acoustic models and language models to optimise the speech recognition to work with the intended user group was required. [iii]

The two main third party speech recognition possibilities considered were *Nuances* 'Dragon Naturally Speaking' (Nuance, 2015), and *Politepix's* 'OpenEars' (PolitePix, 2015). Nuance offers a wide variety of speech recognition resources including automated telephone directory systems, speech recognition productivity software, and a toolkit for developers to integrate

speech recognition into custom software. Nuance's speech recognition software is comparatively more expensive and marketed towards large commercial businesses. The main drawback of this engine was the lack of customisation. Nuance does not allow developers access to modifying the speech recognition dictionaries or language models (Rubin & Kurniawan, 2013).

Politepix's OpenEars platform is the open-sourced alternative. OpenEars offers a free offline speech recognition library with additional purchasable plugins that improve accuracy and recognition speed. OpenEars is based on Carnegie-Mellon's SPHINX speech recognition engine and has a strong developer community behind it. As the engine is open-sourced, the developer has access to modify the source code, including the language model, acoustic model and phonetic dictionary. Additionally the engine allows for offline speech recognition where no connection to the Internet is required.

The readily accessible, well-documented and open sourced OpenEars was a good solution in the early development stages of this project. Through early prototyping and testing with this engine a correlation between the speech recognition requirements of the project with the capabilities of the OpenEars platform was realised. The majority of current speech recognition operates by sending speech input to servers, where the speech is processed and results sent back to the device. The OpenEars platform provides offline speech recognition capability so there is no dependency on a fast Internet connection. This reduces the processing time but limits the processing power, as the external server's computing power is comparatively stronger then a mobile device. However the mobile device can compensate this by using smaller vocabularies to get accurate recognition. OpenEars' recommended maximum vocabulary size is 500 words, which is suitable for a children's storybook (PolitePix, 2015).

2.2.4 Speech recognition and Talking Books.

Adams (2011) discusses the potential for speech recognition technology to be used in an educational context to support children's literacy development. The *Talking Books* project expands on Adams' argument with a practical application of this technology. Speech recognition accuracy improves with a smaller vocabulary size to compare against. In the case of a

children’s storybook, the size of the vocabulary is determined by the narratives vocabulary size, which is relatively small compared to other applications of speech recognition.

Children’s stories have a predefined narrative, so the speech recognition can anticipate what words the reader will speak. Unlike other examples of speech recognition where the users input is unpredictable, a predefined narrative simplifies the speech recognition task and improves accuracy. These circumstances support the rationale for a system that pairs speech recognition technology with children's storybooks and reading out loud. In the *Talking Books* system, many of the technical limitations that speech recognition faces are mitigated.

2.3 The iPad and education

The *Talking Books* application is located and built for the iPad device, encouraging children to engage in active reading. In this section the iPad device, and its relationship to education are discussed.

2.3.1 Overview of the iPad

The introduction of the iPad in 2010 marked a significant step in the development of consumer tablet technology. Although tablets existed prior to the release of the iPad, it was the iPad that launched tablets into the mass consumer market (Zamani, Giaglis, & Pouloudi, 2013). The iPad offered a large multi-touch screen, a simple design with no cables or peripheral attachments, connectivity to the Internet, and high mobile capability. Additionally it has many inbuilt sensors including a camera, microphone, and gyroscope, and had access to a universe of applications (apps) on the App Store.

The iPad's hardware defines its technical capabilities and limitations. Functionality is defined by the software that it runs in the form of apps. Apps are self-contained programs that fulfil a particular purpose. For example, a user may use a map app for getting directions, or a weather app for finding information on the day's forecast. The App Store is the supporting software infrastructure that provides a marketplace for users to acquire apps for their devices. By June 2014, 75 billion apps had been downloaded from Apple's App Store (Statista, 2014a)

A significant factor behind the success of the iPad is the simplicity of the device. Cumbersome instruction manuals are omitted from the iPad, instead intuitive and natural operation takes their place. The software design has incorporated reliable affordances into the iPad and has broadened the diversity of the users to include unlikely demographics, bridging the gap between young and old. Children are naturally attracted to the large touchscreen and generally find the interaction with the iPad natural, readily adopting the devices and grasping the interaction techniques quickly, becoming enthusiastic and competent users (Lynch and Redpath, 2012).

2.3.2 iPads and education

Throughout history, educational milestones have had a close relationship with technological developments with educational aspirations and opportunities driven by advances in technology (Adams, 2011). Webster's Dictionary defines technology as the "practical application of knowledge in a particular area" (Merriam-Webster, n.d). This definition encompasses the creation of cave paintings to communicate stories in early human civilisations, and the invention of paper and the printing press. In the 15th century the invention of the printing press enabled widespread circulation of information and ideas, and enabled the development of a literate society. Televisions, slide projectors, overhead projectors, and radios are further examples of technology that have been used as educational tools. We now find ourselves in an age where information is shared through sophisticated communication infrastructures, notably through the Internet.

Technology has become intertwined in our lives and integral to how we function in modern society. Children born in the 21st century are considered digitally native, having no conception of a world without technology (Gelman, 2014, p. 3). However this notion is globally inconsistent and generally first and second world centric, but is developing rapidly in all cultures with access to this technology. Educational institutes have been integrating technology into the learning environment to provide the best learning experiences for children. "The iPads strong penetration into the classroom is due as much to its appeal as to the oft-claimed potential of technology for education: that it motivates students to learn"(Karsenti & Fievez, 2013, p. 2).

However, as with any form of new technology, the iPad is receiving a tentative integration into the classroom with many arguments presented from different pedagogical perspectives.

The education category on the App Store is the third most popular by share of active applications (Statista, 2014b). These statistics reflects the significant relationship that the iPad has with education. In 2014 the App Store introduced the 'Kids' category, which organises apps that target children into three age categories. This move validated their younger user group, and established a secure environment for parents to acquire suitable apps for their children.

2.3.3 The iPad and children's storybooks

Children's storytelling has found its place on the iPad with a large variety of storybooks available on the App Store. Transposing children's stories from books onto the iPad has enabled complex, interactive storytelling that was previously impossible. In this section I will discuss some capabilities that the iPad brings to children's storytelling.

Storytellers are able to incorporate the iPad's interactive capabilities into stories, providing engaging experiences. Static images are replaced by multimedia with animations and sound. The narrative is no longer restricted to linearity and can instead be interactive and directed by the user. Complicated narrative structures are possible, enabling a personalised storytelling experience. The iPad's sensors can be creatively incorporated into a story to provide interaction between the user and the story. If done well, these features have the potential to enhance the story and the reader's comprehension. However, implemented without careful consideration, these interactions can detract from the story and distract the user.

The storybook *Dandelion (Protein, 2014)* incorporates the iPad's sensors to create an interactive challenge. The reader faces the task of blowing a dandelion on the screen, with an arrow that guides the reader where to blow. When the child blows, the dandelion on screen responds to the velocity of the breath. If the reader blows hard enough, the dandelion seeds disperse into the air.
The application uses the microphone as a sensor, detecting the microphones decibel levels which triggers the animation.

Figure 1: The dandelion interactive blowing sequence

Gamification is a concept which has been applied to an educational software context to increase incentive and motivation for children to engage with a task. Through abstracting

theoretical elements taken from games and reapplying them into other forms of media, the user can become more invested into the software. “Children and young people are often highly motivated by computer games and simulations” (Higgens, et al, 2012, p. 9). This concept has been implemented into many educational settings with varied results.

UK-based publishing company Nosy Crow create well-crafted storybook apps that successfully incorporate these interactive features, and gamification elements into a storey’s narrative. In the next section their contemporary version of the popular storybook *‘Jack and the Beanstalk’* is discussed (see section 2.4).

2.3.4 Criticism of the iPad and education

There has been an ambivalent reception to the educational use of the iPad by younger audiences. While some enthusiastically embrace new media (e.g. Galloway, 2009), others argue that they have no place in early learning (House, 2012). Children’s immersion into digital media occurs at a critical time in their lives when they are developing literacy skills - speaking, listening, reading and writing. This demands extra consideration as the repercussions of careless integration can have detrimental effects on a child’s development. As the iPad is a relatively new device, no research exists on the long-term effects that these devices may have on child users.

Children’s exposure to screen time has rapidly increased with the ubiquity of digital technology in our society. “Young people now spend more time with digital media than they do in school - it is the leading activity for children and teenagers other than sleeping” (Strasburger, Hogan, Mulligan, Ameenuddin, Christakis, Cross & Swanson, 2013, p. 598). Guidelines released in 2013 by the American Pediatrics Association recommends parents limit screen exposure for children to under 1-2 hours, and discourage screen exposure for children under the age of 2 years old. Excess screen time can have adverse effects on children such as developing attention problems, school difficulties, sleep and eating disorders, and obesity (American Academy of Pediatrics, n.d, para. 3). With the popularity of the iPad with young children, and special categories curated specially for younger audiences on the App Store, the iPad seems to be exacerbating the growing problem of children and screen-time.

When analysing the educational significance of the iPad, the critique is mainly focused on hardware aspects. However the software that is run on the device is a major contributing factor that affects the users experience and value that the device offers. However, it is difficult to evaluate the educational potential and the importance of software in childrens' lives (Kucirkova, Messer, Sheehy & Flewitt, 2014). Additionally there is no objective prescription of a good education for a child, as many philosophical standpoints differ on this subject. Some educators argue that the problem is not with this new technology, but instead how we integrate it, and use it with our children. "There is no doubt that technology engages and motivates young people. However this benefit is only an advantage for learning if the activity is effectively aligned with what is to be learned" (Higgins et al, 2012, p. 15).

Critical discussion about the iPad in the field of education is important because designers of children's educational software need to be aware of contextual criticism about areas that can be improved through the development of the software. Despite these debates and criticism of iPad's in educational contexts, this technology is in high demand and is increasingly used by young children. Kate Wilson (2013), the founder of *Nosy Crow* states that recent studies show children are increasingly spending more time with touch screen devices, which suggests a need for beneficial content to be developed for these devices.

2.4 Reviews of relevant media

In this section four texts are critically analysed in relation to the design process and creative development of the *Talking Books* project. The first text is a physical storybook written by author and illustrator Maira Kalman called *Max Makes a Million* (Kalman, 1990). Next I a storybook app version of the classic fairy tale *Jack and the Beanstalk* developed by Nosy Crow (Nosy Crow, 2014) will be examined. In the last section two texts, a research paper entitled *Reader-Animated Stories* by Chris Johnson (Johnson, 2012) and a children's reading app called *The Boy and the Bears Read Aloud* (Richards, 2012) are discussed.

2.4.1 Max Makes a Million

Physical storybooks have a significant place in children's literature and have established a strong foundation for children's storybook culture (Maltepe, 2014). They precede their digital counterparts and remain a popular form of children's literature today. Storybooks are creative works crafted by the author/s using the tools of language and image to communicate a story from the author's perspective. The resulting storybook can evoke a wide range of reactions and feelings depending on the reader.

If a child can relate to a story's narrative, story comprehension and engagement in the reader will be increased (Kauffman, 2005). *Max Makes a Million* is a storybook that I read in my childhood and a storybook that I feel a personal connection to. The language and illustration style feel closely aligned to my ingrained aesthetic interests. Kalman's eccentric characters depict the bohemian feel of New York City in the early 1990's. I share the same cultural context with this book's setting and the date of the books publishing is close to the year I was born in New York. My parents would have read this book to me as a child, influencing my cultural standing and familiarising myself with Kalman's visual aesthetic.

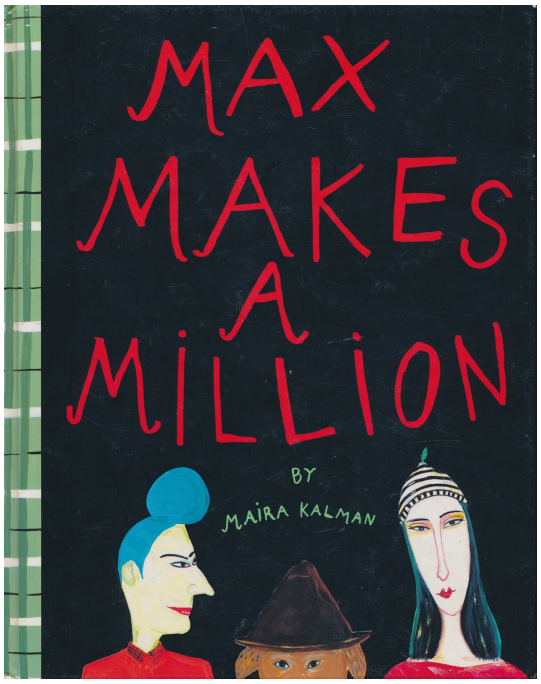

Figure 2: Max Makes a Million front cover.

Max Makes a Million is a children's storybook full of colourful paintings with an accompanying narrative that I found to be engaging and inspiring. It has the qualities of a classic book where the reader is able to revisit the story, over and over again. "Although there is much to glean from an unhurried single reading, this fanciful creation yields its greatest treasures through repeated visits." Publishers Weekly, *Copyright 1990 Reed Business Information, Inc.*

Kalman illustrates a story about a bohemian dog called Max who is broke and lives in New York City with dreams of making it 'big' as a poet in Paris. We hear about Max's eccentric life with his friends and family. The book has underlying messages on the importance of dreaming, having strong ambitions and a wild imagination. The fictional protagonist Max and his crazy life is interesting subject matter that playfully draws readers in alluring their imagination.

"I want to say, before anything, that dreams are very important" (Kalman, 1990).

Kalman includes proverbs, practical knowledge and interesting ideas into this book for children to take away with them. Storybooks that impart some form of knowledge, or communicate a beneficial message through the narrative are recommended as they potentially increase the educational value that a book provides.

“Morris and Ida don’t have any children.

But they have me, Max.

And they start to cry every time I bring up the subject of moving to Paris.

They will have to face the facts.

There is an old Chinese proverb that says parents must give their children two things, roots and wings. I have the roots. Now I want the wings.”

Kalman tantalises the reader with poetic language that is loaded with sensuous connotations ‘I closed my eyes while the candy lay on my tongue’ - (Kalman, 1990).

The text layout of the narrative is unorthodox in its formatting. It playfully jumps around the page, following the curves of the paintings, and squeezes into shapes forming part of the illustrations. In an early page in the book, Kalman uses a typographical formatting technique known as *concrete poetry* where the text discussing Paris is shaped into the Eiffel Tower. Through altering the text layout Kalman creates a dynamic reading experience, and this may act as a catalyst for conversation related to the book, such as: “Do you recognise that shape?” during dialogic reading with a child and adult.

Figure 3: A book spread from Max Makes a Million

The text formatting establishes a close relationship between the narrative and the illustrations. Kalman also plays with individual words, creating emphasis for the reader by stretching out words and increasing their weight (see Figure 4). To the child reader, Kalman's playful creative license may inspire and encourage creativity and experimentation in contrast to the standard conventions that they are taught.

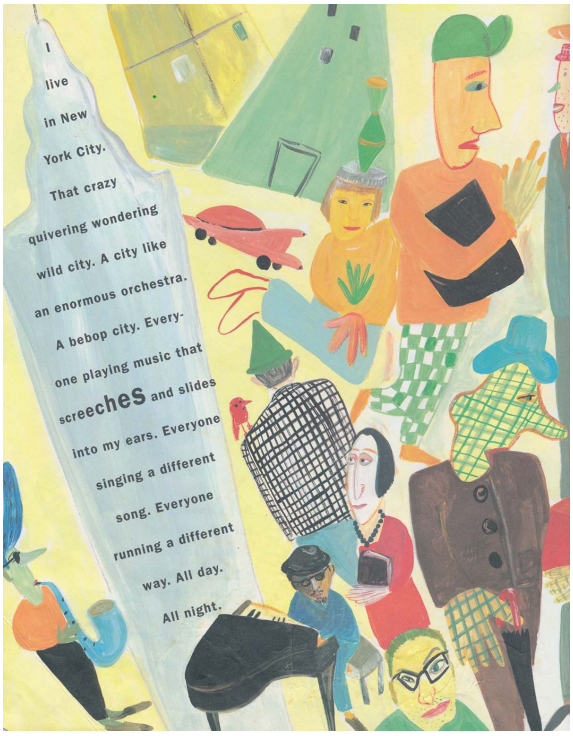

Figure 4: Max Makes a Million page detail.

Kalman's painting style is crudely elegant. The artist's brush-strokes are noticeable giving the illustrations a tactile feel, with the underlying pencil marks still visible in some pages. To an adult, the simple painterly style is charming, and gives the book a sense of approachability. Kalman's visual imagery is not technically correct - the perspective and anatomical proportions are abstract, but this does not detract from their ability to communicate the story. To a child these paintings may feel familiar and attainable. Kalman's illustrations are a strong example of how technical expertise is not required to create strong images to support a narrative. The handmade style of the illustrations leaves space for the child's imagination and room to engage with the content.

Figure 5: Max's best friend, Bruno.

Max Makes a Million is a strong example of a physical storybook that contributes an inspiring and whimsical story to children's literature. Kalman's poetic narrative and painterly illustration style create a story that resonates and has influenced me as a content creator. Through analysing Kalman's visual and narrative conventions in *Max Makes a Million* I have positioned myself as a content creator and further contextualised the development of the *Talking Books* stories.

2.4.2 Jack and the Beanstalk

Nosy Crow's version of the classic fairy tale *Jack and the Beanstalk* is an interactive storybook app for the iPad. This app sits in the same category as the *Talking Books* app and targets the same early reader age group. It is a good example of how the iPad's technical capabilities can be integrated into the narrative to create a highly engaging reading experience (Common Sense Media, 2014). The story offers two modes: 'Read and Play' and 'Read by Myself'. For this review I will be discussing the app in the 'Read and Play' mode as it exhibits the applications full functionality.

The visual animations are colourful and stylised, the characters have a cartoon feel, with long noodle-like arms and large heads and eyes. The visuals are composed of three 2D planes (a foreground, middle and background plane) that are situated behind one another. The middle plane is where the main content is displayed and where the stories interaction occurs. The background and foreground planes provide context to the scenes and create a sense of depth. These planes shift around controlled by the physical orientation of the device, creating an interactive perspective effect to the story.

However, it is not the style of illustration but the integration of interactive functionality that merits this *Nosy Crow* work's inclusion as a case study. It provides an exemplar for interactive storytelling on the iPad device.

Nosy Crow's *Jack and the Beanstalk* uses procedural animation that enables dynamic content interaction. Procedural animation is a type of computer animation where animations are generated in real time and allow for flexible user generated animation as opposed to pre-defined rendered animation. Within this app the reader helps Jack through various challenges to progress through the story, such as in Figure 6 below. The reader is required to drag 10 magic beans out of a suitcase from the strange bean seller into Jack's backpack. When the reader finishes moving each bean the narrator counts with the bean collection. One.. Two.. Three.. etc.

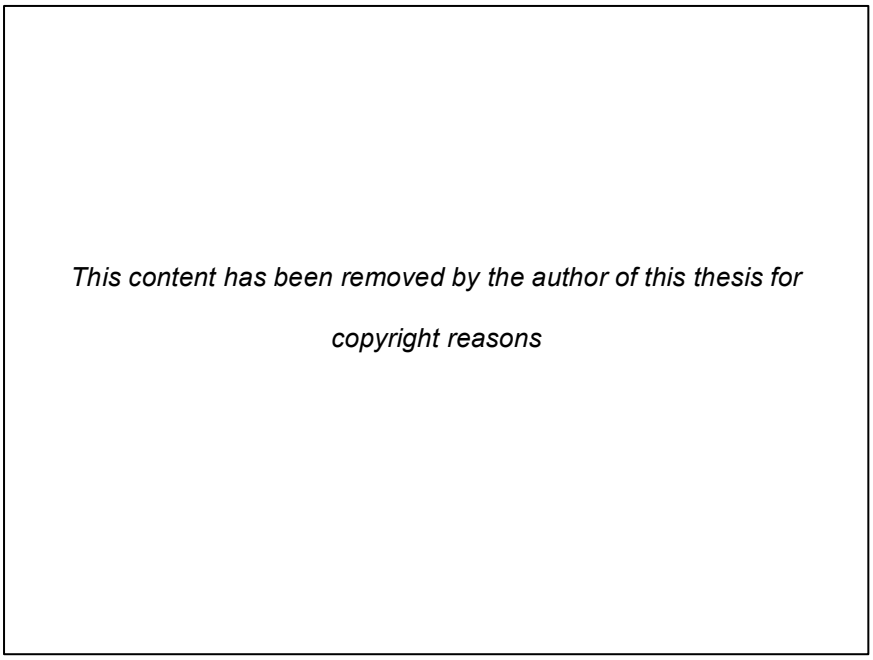

Figure 6: Jack meets a strange-looking man

Due to the stories interactivity and gamefull challenges, the distinction between the user being a reader or a player is hard to determine. The integration of interactive elements into the narrative is done with careful consideration, and reinforces the stories narrative rather than distracting the reader from the story.

The story implements multiple narrative paths through a game like mechanism. The user progresses through the narrative by collecting coloured keys - the order that these keys appear is randomised, and alter the order of the challenges. Multiple narrative paths encourage reruns of the story and additional gameplay to discover alternative endings to the story.

One of the challenges the reader encounters is a puzzle consisting of a shattered mirror that needs to be reassembled back into its original shape. This puzzle uses the front facing camera of the iPad to transposes the camera feed onto the fragments of glass, for the user to reassemble using the touchscreen.

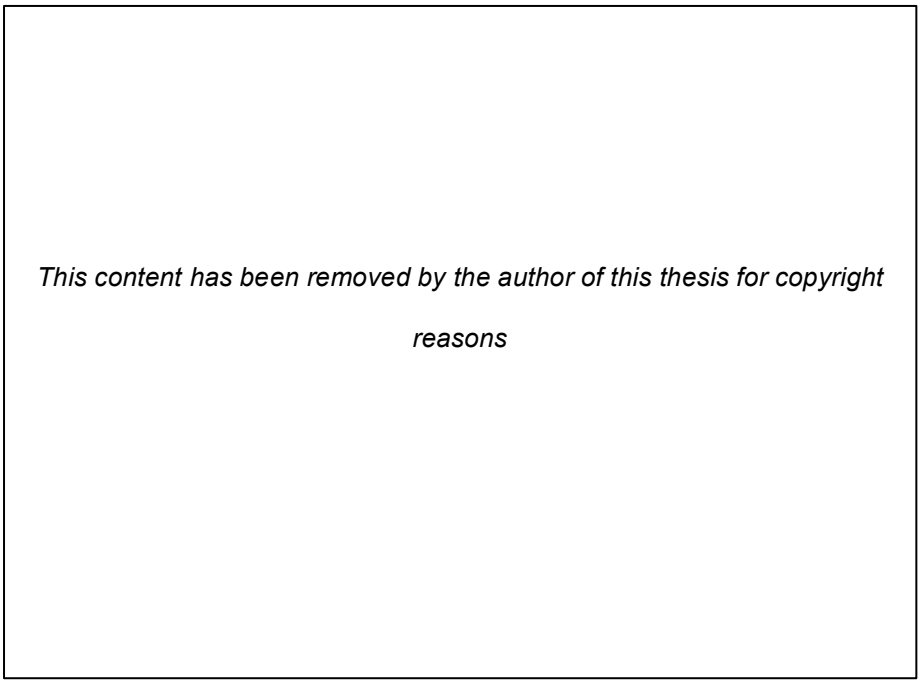

Figure 7: The broken mirror puzzle

Another challenge uses the inbuilt gyroscope (a sensor that detects the orientation of the device) when directing a bucket to the bottom of a well, avoiding obstacles to collect a golden egg. This action is controlled by physically tilting the iPad device.

The user experience and interaction design of *Jack and the Beanstalk* has been carefully considered. Children are able to progress through the story with inbuilt prompts that support their progress through the story. In the title sequence the reader is introduced to a blue circle that acts as a guide by prompting the reader with touch points. When a character speaks they create a speech bubble displaying the sentence and highlighting the words as they are spoken. The characters also give the reader helpful tips through informal dialogue which appears in speech bubbles. Overall, the user-interaction is very polished and demonstrates how procedural animation can provide highly dynamic interaction through user touch. (Common Sense Media, 2014, unpaged) reviews this application stating: “Really impressive use of the device's tools -- multi-touch, forward-facing camera, graphics - makes the design first rate. Kids will have a blast working their way through the adventure and seeing how their actions impact the story”.
The animation within the *Talking Books* story does not use procedural animation and as a result does not offer interaction capabilities as dynamic and engaging as the *Jack and the Beanstalk* app. However through analysing *Nosy Crow’s* application, the technical horizon has been

expanded and future stories will incorporate similar interactive capabilities, using *Jack and the Beanstalk's* interactivity as a strong reference.

2.4.3 Reader-Animated Storybooks and the Boy and the Bears Read Aloud

Chris Johnson is a computer scientist from the University of Wisconsin who published a paper in 2012 entitled 'Reader-Animated Storybooks' (Johnson, 2012). This research paper describes a digital storybook system for mobile computers that shares a similar concept with *Talking Books*. Specifically they both use speech recognition to control animation playback in children's storybooks. Johnson explains that Reader-Animated Storybooks combine the passive, but appealing visual stimuli of electronic media, with the active cognitive engagement of reading. A collaborative methodology is described involving a computer science student and a design student to develop the system. Johnson suggests possible benefits of this system as the potential for improved literacy for early readers and future commercial development opportunities.

Upon discovery of Johnson's work it became apparent that a lot of the previous work gone into the earlier prototypes of *Talking Books* had overlapped with his research, however Johnson's approach to this project is more technical because he comes from a computer science background. With the *Talking Books* project I am focusing on a larger context including content creation, interactive experience and educational considerations in the design of the software.

Another text that has similar functionality to the *Talking Books* app is 'The Boy and the Bears Read Aloud' (2012). This free app uses the reader's voice to progress through the stories narrative using speech recognition. The story in '*The Boy and the Bears Read Aloud'* is about a boy who climbs up a tree house, enters four rooms and encounters a bear in each room. After a brief interaction with each bear, the boy climbs back down from the tree house and says good night, ending the story.

This app embeds touch-spots into the stories scenes, which enables the readers to interact with specific objects resulting in an animation and/or sound when touched. These touch-point interactions do not contribute to the stories narrative but provide more interaction for the reader.

The reader is able to visually recognise these touch-points because they glow with a white aura around them. The app uses a guiding arrow that bounces above the text to shows the reader which word it is waiting to hear. When the reader speaks, or the microphone hears sound, the arrow changes colour from white to red.

The animations and illustrations can be described as clean and child-friendly. However the story lacks interesting or engaging features - the narrative fails to communicate any constructive message and the interactions are not creative. However the main criticism of this app is the speech recognition performance. The functionality of both *The Boy and the Bears Read Aloud and* the *Talking Books* app relies on accurate and robust speech recognition performance. If the speech recognition is not implemented correctly it is a detrimental factor to the overall user experience. Additionally *The Boy and the Bears Read Aloud* application seemed as if it failed to undergo robust user testing as basic functionality proved to be ambiguous and the system failed to give adequate feedback during the application's use. I found myself confused during parts of the storybook.

In this application the speech recognition analyses the audio input when a break is heard in the microphone input. In the operation of the app, this translates to when the reader finishes reading a sentence. The app gives feedback on the reader's progress by an arrow that signals the current word. However, due to the way in which the speech recognition functions, the arrow jumps immediately from the first word, to the last word during use, failing to provide any real-time feedback mid-sentence.

The Boy and the Bear Read Aloud shares the fundamental functionality of *Talking Books* but fails to demonstrate the potential of this system. This app reiterates the importance of having a robust speech recognition system, as it has proved to be a critical factor to the applications functionality.

3.0 Methodology

This section introduces, and contains a discussion of the methodological approach and research design followed to address the research questions set out in section 1.0. The research design involved iterative phases of practical, technical and design development, which were evaluated through user testing and analysis, and informed by contextual and theoretical inquiry.

The project was initiated through an investigation of the potential applications of speech recognition technology. This technology was paired with the act of reading children's storybooks, and an exploration of the potential benefits of this proposed system. The early technical development followed a *hacking* approach to create prototypes through problem solving and utilising available resources. "Hacking can be described as modifying an existing system (hardware, mechanical, or software) to improve performance or create an application that differs from the device's original purpose" (Paradiso, Heidemann, & Zimmerman, 2008, p.13-15).

The broader methodological framing of this project has been influenced by the open-source movement (Lerner, J., & Tirole, J., 2001), and utilising the Internet as a collaborative resource for developing software. Programs are freely shared through websites such as *GitHub,* which provide a platform for storing user-generated code in virtual directories known as *repositories*. S*tackOverflow* is an online forum with a global community of developers that provides a platform for questions to be asked and answers to be given. *StackOverflow* documents these user-generated questions and answers and was an invaluable resource throughout the development of *Talking Books*.

A modular approach was taken in the technical development process, where the core functionality of *Talking Books* was segregated into smaller, isolated programs. This made it easier to research and manage the initial development process. Once these smaller programs achieved their intended functionality, they were combined with other functional modules, forming early prototypes. An example of this was pairing a program that achieved basic speech recognition functionality with another program that played video files. These initial technical prototypes were important stepping-stones which refined the research focus, developed a better

understanding of the project's viability and contributed both practical and theoretical knowledge to the literature and contextual review.

During the collaboration with the Computer Science students, *Xcode's* native source control - *Git* was used to organise and coordinate the external development. Git monitors changes made to the code files, saving versions at stable stages and enabling the possibility of reverting back to these versions at later stages if bugs or problems arise. Using a source control, developers can work on different *branches* of a project separately, and when appropriate, their *branch* of the project can be *merged* into the larger program. Git source control coordinates and decentralises the technical development, enabling easy, transparent and safe collaboration between the developers.

A significant outcome of the design process was the animated stories, which involved visual, audio, narrative and interactive dimensions. How users reacted to, and interacted with these stories were key considerations, with both aesthetic and practical implications.

Paper prototyping was used during the development process as a means to gain insights into how the application will function with users, and to gain early feedback on the system. Paper prototyping allows the designers to quickly discover which parts of the interface work well, and what areas need improving. As it is all on paper it can easily be modified and mitigates the risk of finding similar problems after investing time developing a technical prototype (Snyder, 2003).

3.1 Literature and contextual review

The literature and contextual review gathered research to support, inform and situate this research project. *Talking Books* is informed by developments in current technology, existing material (including apps, children's books and visual imagery) and research about reading, education, children's media and new technologies (See Figure 8).

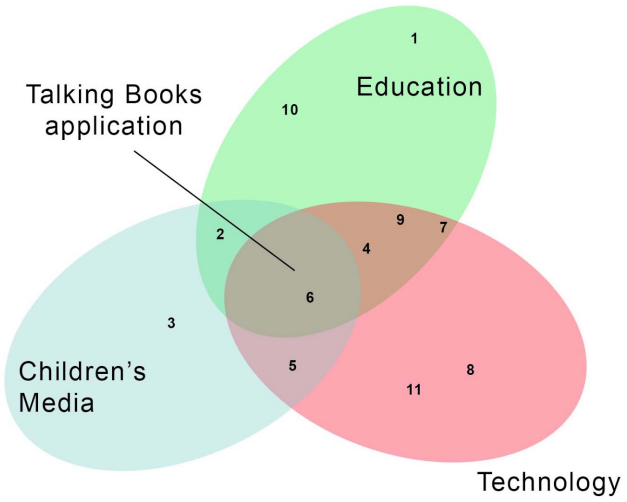

Figure 8: The contextual location of *Talking Books* and relevant texts.

Key texts considered in the review include: *Learning to Read and Write: Developmentally Appropriate Practices for Young Children (1) - which* establishes a position statement regarding early literacy development developed for the National Association for the Education of Young Children (Neuman, Copple, & Bredekamp. 2000). *The Importance of Literacy and Books in Children's Development* (2) discusses the role of children's storybooks in relation to literacy development (Stockar, 2006). *Max makes a Million (3)* by Maira Kalman is a children's storybook that I discussed in the previous section 2.4.1 in relation to effective illustration. *(Kalman, 1990). Technology for Developing Children's Language and Literacy: Bringing Speech-Recognition to the Classroom (4)* was especially pertinent in rationalising the pairing of speech recognition with early literacy development. (Adams, 2011). Common Sense Media's *Zero to Eight Children's Media Use in America 2011, (5)* is a study that reveals children's media use and trends towards digital mobile devices. (Common Sense Media & Rideout, 2011).

Jack and the Beanstalk (6) developed by *Nosy Crow* is an interactive storybook app for the iPad, I have discuss elements of this application in section 2.4.2 (Nosy crow, 2014). *Leap Frog Enterprises (7)* is an educational entertainment company that offers a wide range of technology-based learning products for children. These were considered in relation to this project.

PocketSphinx: A Free, Real-Time Continuous Speech Recognition System for Hand-Held Devices (8) is a paper introducing the open source speech recognition system used in Talking

books (Huggins-Daines, Kumar, Chan, Black, Ravishankar & Rudnicky, 2006). *iPad in Education: A case study of iPad adoption and use in a primary school* (9), provided insights into current educational developments

*(*Henderson & Yeow, 2012) *Ready to Read levelled texts (10)* introduced a colour wheel system that indicates the reading level of children's book. This system was a good reference in developing appropriate reading content for our target audience. *Child Automatic Speech Recognition for US English: Child Interaction with Living-Room-Electronic-Devices (11) is* a study into adapting adult targeted automatic speech recognition systems for children (Gray et al, 2014).

Talking Books is a technological artefact in both form and function. As technology evolves at an unprecedented rate, realising the current state of relevant technology prior to development was an important task. Specifically examined was the technology around tablets, interactive entertainment and mobile applications. This initial research stage helped to establish the project's rationale, and increased the researcher's understanding of the project's viability. Additionally the literature and contextual review explored the surrounding areas that influence and relate to the project. This information widened the researcher's holistic understanding of the project, narrowing the research focus and establishing a foundation of knowledge for the research to build on.

3.2 App Development

A user-centered design (UCD) methodology was employed throughout the development of the *Talking Books* project. This enabled the system to be designed from the perspective of how it would be understood and used by the intended end user, supporting their existing behaviour, attitudes and beliefs. UCD is an iterative process, with each iteration the design becomes more precise, gaining more detail, and the user feedback focuses on smaller and smaller problems (Diehl & Thüs, undated). Applying UCD to a systems design results in a product that offers a more efficient, satisfying and user-friendly experience for the user (Shneiderman & Ben, 2003). The iterative stages in the UCD cycle are shown in Figure 9.

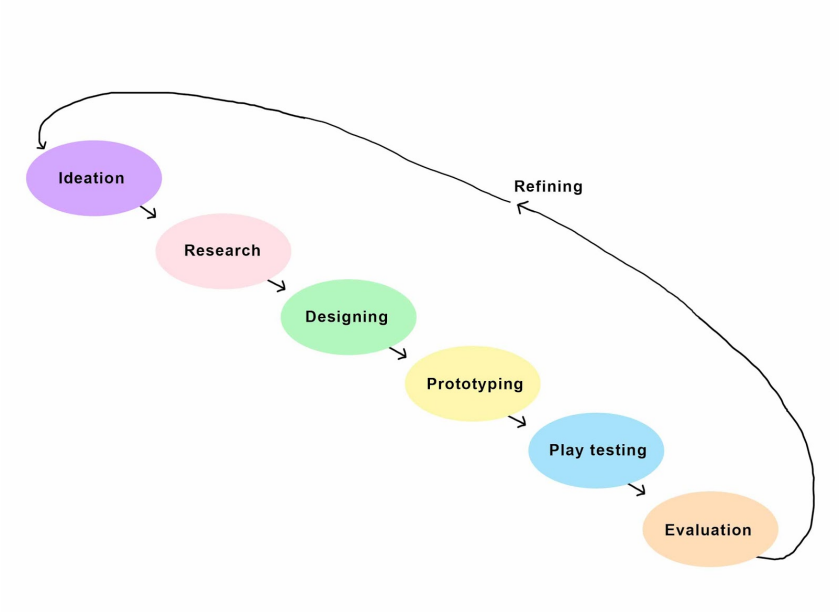

Figure 9: A typical UCD cycle

A typical cycle would include:

• **Ideation**: Theorising changes and experimenting with ideas that may improve the application and the inclusion of input from the intended users, incorporating their needs and ideas from the beginning of the development.

• **Research**: Readings on relevant topics including surrounding media. The intended users are involved in the research stage, providing insights into their specific interests, problems and capabilities to the researcher/designer.

• **Designing**: Combining the ideation and research stages to plan out elements of the design. The design is developed with constant consideration of the end user, how they will interpret, use and respond to the application.

• **Prototyping**: The tangible development of the application, guided by the design stage. Involving collaboration between technical developers and content creators to create prototypes of the application.

• **Play testing**: Testing the applications design, content, technical capabilities and function with the end users. The users and system are analysed during the play testing to gather data on how it performs, and how the users respond to the application.

• **Evaluation**: Reflecting on the play testing stage. The application is critically analysed with user feedback informing the design process, and the performance of the application is compared to the research focus.

User-centred design is a popular approach for software development, as the user's affordances[1] and interactions are taken into account (Rannikko, 2011). This is especially pertinent when designing an interactive application that requires the user's active engagement. As the *Talking Books* software targets a specific demographic - children aged 6-8 years old - their input and feedback into the design process is invaluable. Specifically within the *Talking Books* project, the user's feedback has been essential for optimising the speech recognition to work with the intended user's voices (Please see section 5 for result table)

To conduct the user testing within an academic context, applying and gaining ethics approval from the university was necessary. Preparing the ethics application form was a useful process, as I had to consider all of the possible scenarios and elements involved in the testing process. Parental/guardian permission forms were created, all of the safety aspects of the testing were considered and questions for an unstructured interview with the participants that would prompt relevant feedback from the participants were prepared.

3.3 Design Development

The design development encompasses the content and user interface development within the application. A significant portion of this stage was related to the development of the visible content. A user centred approach involved the intended user group in the development process. They provided both aesthetic and practical feedback. In the prototypes visual animations were originally developed with the purpose to test but eventually to showcase the system's functionality. These animations were created with the intended users in mind, and the animation style, technique and content were refined through iterations of prototypes. When the system was at a suitable technical level, and the designers had established an understanding of how the media playback operated in the system, the animation progressed towards communicating children's narratives.

[1] I use affordance in the context of Human Computer Interaction, describing the possibilities of the users' actions.

The narrative was developed through exploring interesting themes or visual imagery that appealed to me, as the designer. As a visual artist I have a practice of keeping visual journals with imagery and writing, collecting experiences and noting down things of interest. These visual repositories were used as references, re-examining ideas, combining elements together and developing narratives around interesting concepts. Actions can be the catalyst for a story's development, in section 4.7 I discuss my ingrained interest in the ladder construct, attracted to the repetitive climbing action which aligns to my animation practice. Visual imagery and sequences were the initial starting points for the narratives, which were developed and refined through feedback gained from the intended user group.

The animation was developed with close consideration of how it would be handled by the system. This relationship determined the synchronicity of the user's interaction with the media playback directly affecting the user experience. Different methods and techniques were employed in the animation creation including traditional techniques and digitally adapted techniques. The aesthetic direction was influenced by surrounding media, personal ideology, and practical considerations. Visual references were found and also made to support the animation development (See section 4.7, *Animation Technique*).

The user interface design was focused on clarity and simplicity, intended to facilitate an intuitive comprehension of the systems functionality and operation. As children are the intended audience, the book *Design for Kids* (Gelman, 2014) was a useful resource that provided detailed recommendations and practical examples of design elements specialised for children. Additionally user testing was important in the development of the user interface as it revealed insights into how the users interacted with the system, and directed the development of the system's interface.

3.4 Collaboration

Methodologically, collaboration has been integral throughout the development of the *Talking Books* project, including co-creation in the design process (Saunders and Stappers, 2008), which can provide a well-rounded and informed project as alternative perspectives and specialised skills are included into the development process. Collaboration enabled the project

to progress in a way that would not be feasible with a sole contributor within a similar timeframe. The identification of the contributions made by collaborators helps clarify my own contribution to the project.

A significant collaborator has been my friend and research partner Sam Joe. Sam and I worked together on the development of some early *Talking Books* prototypes in 2013 and has been close at hand to discuss the projects focus and development. We have established a working relationship where we contribute specific elements to this project and can discuss the project holistically. Sam has specialised in the educational aspects of the project, contributing research and applying theory, whereas I have specialised in the content and interaction development of the project. Together we share the same project vision and have been able to consult one another throughout the project's development.

The content development of the app has included a number of collaborators as co-creators including animators, computer programmers and sound designers. Their various contributions are discussed in section 4 – Project Discussion. Throughout the project I directed, analysed, provided feedback and managed these particular developments. A user centred design process can also be considered as collaborative in that it involves the participation of users in the ideation, testing and evaluating the artefact. Saunders and Stappers note "... we have seen that co-creation practiced at the early front end of the design development process can have an impact with positive, long-range consequences" (2008, p.9). Figure 10 illustrates the specialist areas and collaborative elements that inform the *Talking Books* project.

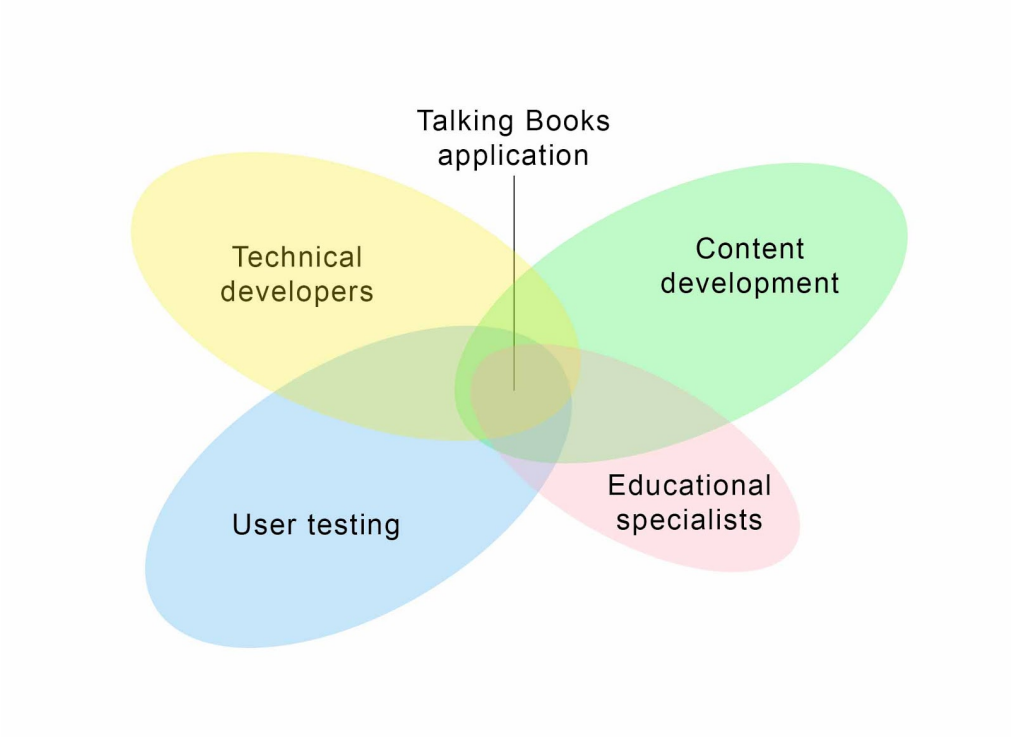

Figure 10: Collaborative areas contributing to the *Talking Books* application

3.5 Project Reflection and Analysis

Reflection is a recurrent stage in design development cycle where progress is analysed on its significance and contribution to the project. When working closely on certain aspects of the development, it is possible to lose the holistic perspective of the project. The reflection process enables the researcher and contributors to consider and analyse their contributions to the project. Reymen and Hammer (2002) recognise that “reflection on a design process can contribute to a steeper learning curve of designers, to a smoother design process, and to an improved product being designed” (p. 888). Reflection occurs sporadically and at timed intervals throughout the development timeline, where knowledge attained through the research and practice is reflected on in relation to the research aims and focus. The user testing stage in the development cycle provides data and important opportunities for reflection. This form of reflection uses valuable feedback from the user’s perspective that must be considered and used to inform the future iterations of the project.

3.6 Evaluation, Critical Analysis and Synthesis

The evaluation stage of the research project sought to attain an understanding of the project’s success in meeting its original aims, its significance, what problems occurred so that they can

be rectified, and what contribution the research has made in its field. The evaluation stage enables further reflection of the project and assists in identifying future developments and directions. The results from the user tests contributed data to the evaluation stage, as did peer and expert feedback. Evaluation has been a recurring process throughout the development as it is critical to the UCD process. It provides an opportunity to step back from the project and critically reflect on any changes against the research focus and questions.

In the critical analysis process, the research can be broken down into key components, and these are analysed from the researcher's perspective. The application is judged on its success and original intended focus, additionally, the technical and cultural significance of the project is also analysed. The potential future steps of this project are discussed, with recommendations of research areas of further relevance.

Writing has been used in this design project not only to document practical and conceptual research developments and collate the findings of the research into one document but as part of a synthetic process, where my research and practical development is combined with contextual research and critical analysis to gain insights and contribute to a new understanding of the *Talking Books* system in an academic context.

4.0 Project Development

The *Talking Books* project has developed through a number of iterations. In this section I will outline the project's development chronologically and analyse the contributions and insights gained from each prototype. Figure 11 below shows a timeline of the *Talking Books* development indicating progress that occurred before and during the Masters degree that developed this research project. This section discusses seven prototypes in total, including two paper prototypes.

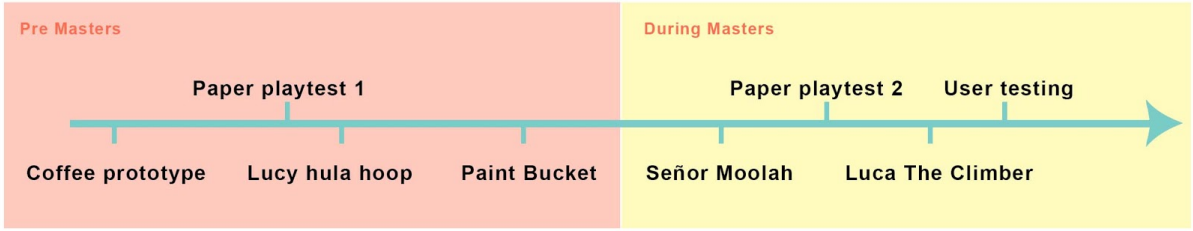

Figure 11: Timeline of the project's stages.

4.1 Practical development prior to Masters

4.1.1 Coffee Prototype, June 2013

The Coffee prototype was the first step in the technical development of the project. Basic functionality was established through this prototype.

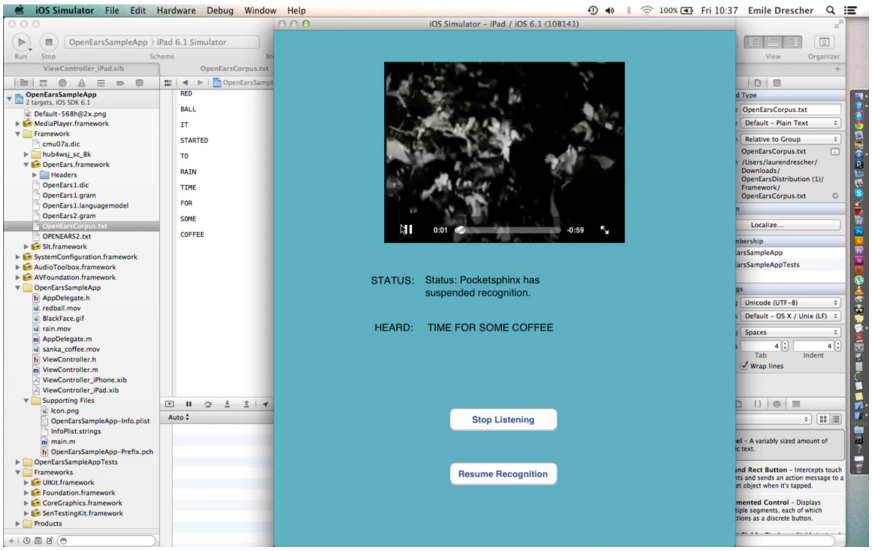

Figure 12: A screenshot of the coffee prototype.[2]

2 A video documenting this prototype can be found here: https://vimeo.com/63867398 password: talkingbooks

This prototype was my first interaction with the technical components of the project. Familiarising myself with Xcode, which is Apple's app development environment, as well as the external speech recognition framework OpenEars. The Coffee Prototype was developed by modifying the OpenEars example project. In the prototype, a video would play when the system heard a pre-defined sentence spoken. Achieving this basic voice-command functionality established an early technical foundation for this project to build on.

4.1.2 Paper playtest session 1, August 2013

A playtest was organised at a local primary school that was conducted to establish early feedback on the system's design from the intended audience, before any significant effort was invested into implementation. Paper prototyping is a technique known for usability testing a system that has a human-computer interface. It is a widely used method for designing, testing, and refining a user interface (Snyder, 2003). Two narratives were prepared for this testing session, and the feedback gained from these would inform future *Talking Books* stories. The first story was based around *Señor Moolah*, and the second story was an early version of *Luca the Climber (See sections 4.2.1 and 4.2.3 respectively)*. Through this informal feedback session I aimed to:

• Gain general feedback on the systems design and operation from the intended user group.

• Inform the project from a pedagogical standpoint through receiving teacher's feedback on the system.

• Gain early feedback on two early stories that were being developed for the application.

Outline of the paper prototype session:

The paper prototype session was developed into a classroom activity with a class of year 2 primary students aged between 6-7 years old. The exercise was a classroom activity that involved a large paper structure that resembled an iPad, two children's stories, and groups of 6 children that were split equally into readers and actors.

The readers spoke the stories narrative out loud, and the actors performed the stories narrative impromptu in sync with the reader's articulation. For example, when reading the sentence 'the dog dug at the earth looking for juicy bones", the actors jumped down to their hands and knees and began a digging motion with their hands. The teacher's assistant encouraged the readers, and followed the reading progress with their finger. After each story was finished the roles were swapped so the participants had a chance to try both reading and acting. After both stories were finished, the teachers concluded the session with an open discussion on the exercise. Questions were asked that prompted discussion around the systems functionality and the flow of the exercise.

Figure 13: Paper Prototype Session 1.

Summary of the feedback:

The children participants responded well to this exercise, and were excited to engage with the large paper prop. The children enjoyed the director/actor roles and treated the exercise as a game. The stories included were generally suitable for the children's literacy levels, with some participants requiring additional time to decipher unfamiliar words.

The unstructured questions revealed that the participants responded better to the Luca the Climber story. Participants were interested in the protagonist, asking questions about the character, and enjoyed acting out the narrative. This exercise succeeded in replicated some of the *Talking Books* functionality, but included some aspects that are not present in the software - notably the theatrical elements to this exercise. Feedback indicated that some participants were too shy to act or read aloud in front of their peers.

4.1.3 Lucy Hula Hoop, October 2013

The Lucy Hula Hoop prototype introduced an animated sequence and a user interface to the system. The animation sequence was comprised of a character called Lucy who the reader could instruct to pick up a hula hoop and play with it. The animation was repeatable and would finish at the starting frame ready to start again. An icon was included in the user interface that displayed an ear/mouth image to communicate the program's status - whether the system was currently listening to the reader, or waiting to hear them speak.

Lucy's animations were created following a traditional cel animation technique. Each animation frame was illustrated on paper using a light box that produced an onion-skin effect[3]. This is a time intensive process, and my inexperience of animation lead to an increased workload. Little preliminary planning was done around the flow of the animation and I found it difficult to understand the relationship between the animation's frame rate and the amount of movement between each frame. As a result I overworked and illustrated too many frames. Additionally the

[3] Onion-skin effect is a technique used in animation where frames are made transparent to see the previous frames, which the animator references to create the new frame.

detailed stipple shading style proved to be a time consuming and a regrettable decision when tasked with illustrating the amount of frames required.

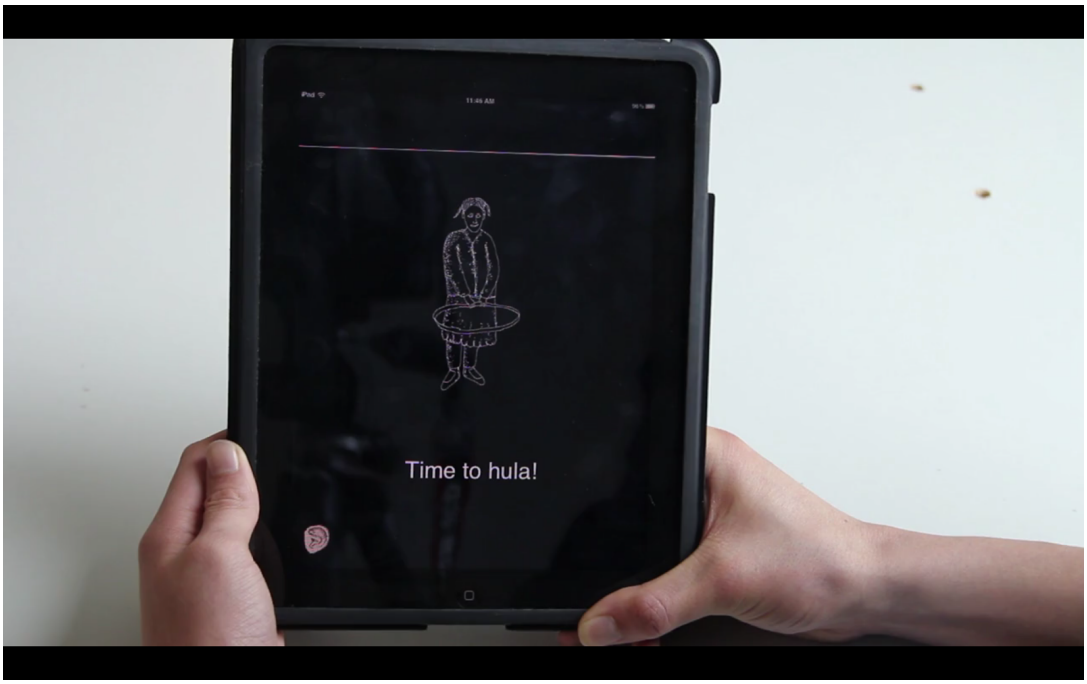

Figure 14: Lucy Hula Hoop prototype on the iPad.

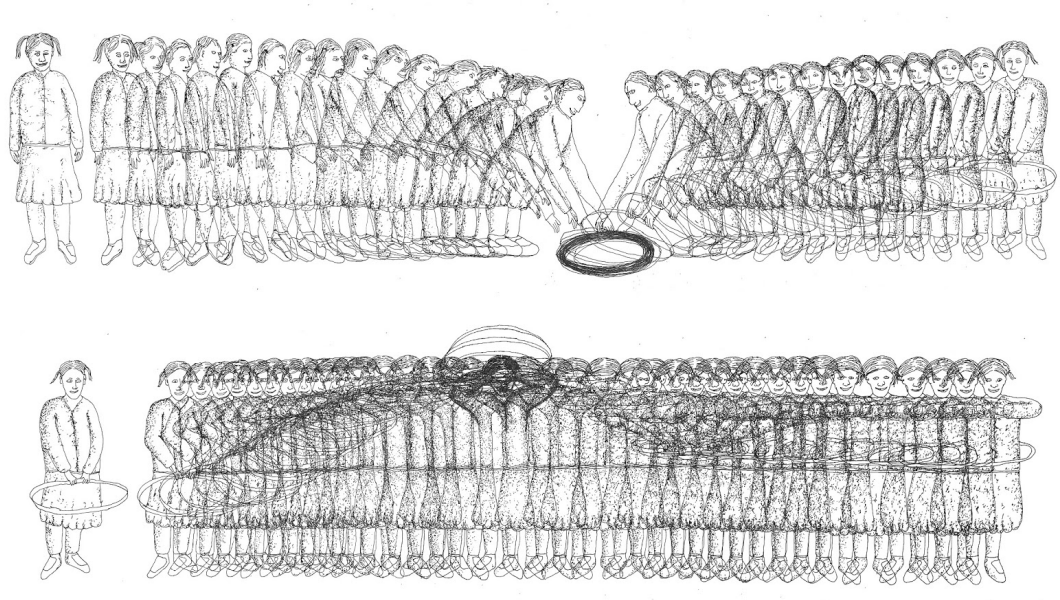

Figure 15: Animation breakdown of the Lucy Hula Hoop sequence.

The speech recognition in the Lucy prototype was basic and created through editing the OpenEars example project. To compensate for the speech recognition's low accuracy rate, this prototype was programmed to listen for a single word in the sentence, rather than the whole sentence to advance the scene. The word that the system was listening for in the sentence was

unknown to the user. This functionality simplified the speech recognition's task as it required hearing a single word, instead of the entire sentence to advance the media.

Media playback considerations

As the iPad has the capability of displaying engaging multimedia, I wanted to avoid static images within the *Talking Books* stories. This was achieved by displaying looped animations when the system is in a standby status, waiting for the reader's input. These standby animations consisted of short repeatable animations that gave the story a sense of responsiveness and liveliness.[iv]

Queuing video files formed the media playback in the Lucy prototype with the standby and action sequences loaded as separate video files. When the reader read the correct words out loud, the system would progress the narrative by loading the movie player with the suitable video. This worked and the user was able to progress through the entirety of Lucy's animated sequence, but a blank screen was visible for a short duration when a new video file was loaded into the movie player. This blank screen disrupted the flow of the narrative and animations and the result of a lapse in time required to load and buffer the new video.

Through the development of the Lucy prototype, my technical understanding of the system was improved and an early user interface was developed. The creation of the animation was a chance to experiment with some animation techniques and develop my understanding of the media's relationship with the system flow. I received criticism relating to the lack of colour in this prototype.

4.1.4 Paint bucket, December 2013

By examining how other forms of technology handle media, a solution to the video buffering issue that was present in the Lucy prototype was developed. Instead of separating all of the waiting and action sequences into individual movies and queuing them, it is more efficient to combine these sequences into one video and manipulate the playback of one video. This works in a similar fashion to how DVD players and chapter functionality, where the user can skip to a certain chapter in a movie.

The paint bucket animation was developed from a screen recording. The purpose of this prototype was to create a simple looped animation demonstrate the functionality of the system clearly. The animation sequence was repeatable and changed the colour of squares when the user “yellow square turns red”, “red square turns purple”, “purple square turns yellow”.

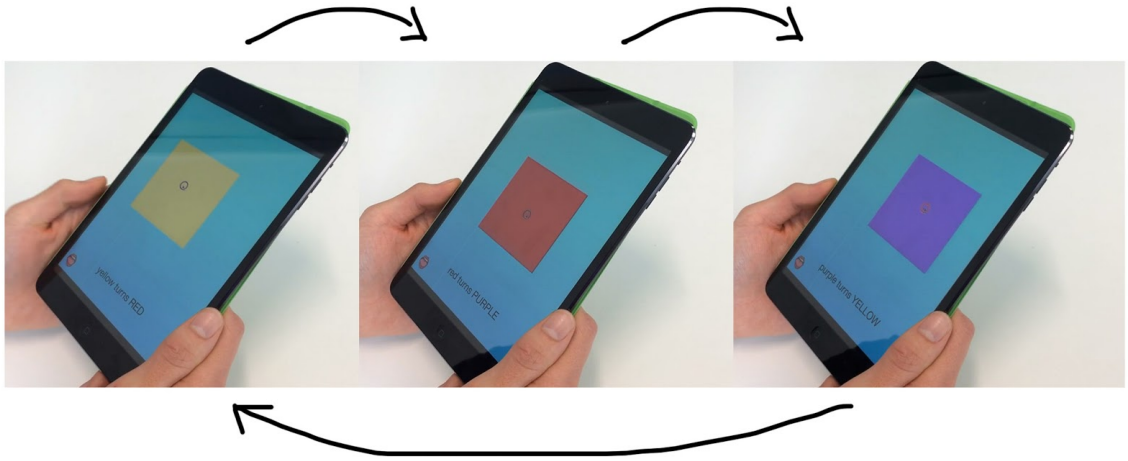

Figure 16: The paint bucket prototype showing the flow of the animation.

In this prototype it became apparent that the visual icons were not successful in communicating the system’s status to the users. The ear icon intended to convey that the system was actively listening to the user, and the mouth icon, that the system was waiting for the user to speak. However through feedback it became clear that these two visual symbols were ambiguous and proved to be confusing for the users.

4.2.0 Practical development during Masters

4.2.1 Señor Moolah, April 2014

The Señor Moolah prototype was an attempt to progress from simple looped animation sequences, to the first children's story for the application. I wanted to investigate how narrative would perform on this system, and how the target audience would react. This created the opportunity for me to develop a narrative and visual imagery to support the story. I started by reviewing imagery and themes from personal visual diaries and looking at existing children's media for inspiration. By compiling illustrations together a narrative was developed around these ideas. It included an eccentric protagonist and a story based around him and his pet snake. Below is the story's summary.

Señor Moolah and Anna, his Anaconda.

Señor Moolah is a fancy Spanish man who lives with his pet snake Anna. He is quite a strange man, who is very hairy and wealthy but lives a solitary life besides from Anna - Moolah's beloved large pet anaconda. Moolah lives at the top of a hill with gates surrounding his property, but children from the village below love to pester Moolah as he is so easily upset. These children are the bane of Moolah's existence! Every day Moolah takes Anna out for a ride around his property and occasionally catches a child who is swiftly eaten by the ravenous Anna - Moolah considers this his grounds keeping.

The day swiftly approaches where Anna has eaten all of the children, and Moolah is finally able to relax, able to comb all of his body hair in peace. But Anna's ominous appetite is growing, yearning for her normal diet of children. With no children left to appease her hunger, she turns to the unsuspecting Moolah - swallowing him whole. Moolah finds himself in an unfortunate predicament - stuck inside the belly of Anna, in close confinement with all of the children who he once despised. They are forced to work together, developing an ingenious escape plan from the belly of the unruly anaconda.

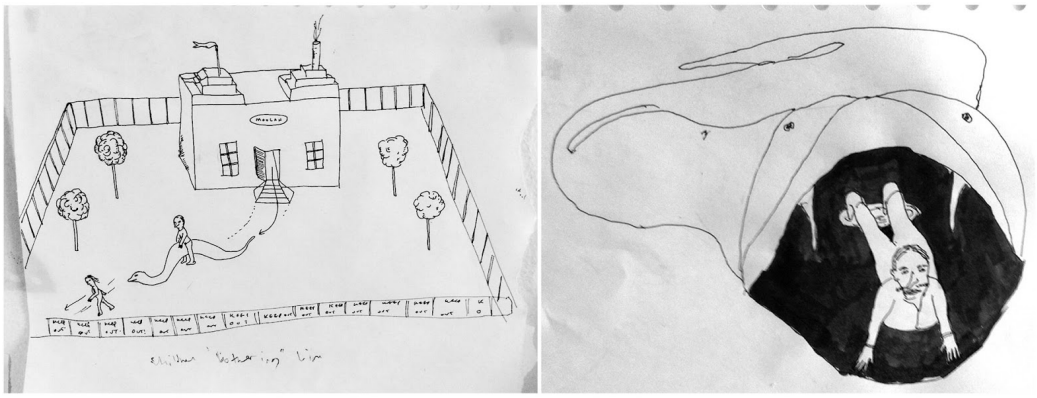

Figure 17: Early sketches of the Señor Moolah story

This story was being developed as a pilot story to showcase the application's functionality. Due to concerns about the appropriateness of the story by peers, this narrative was not developed into an animated storybook. However Señor Moolah was developed into a short animation sequence through the collaboration of two digital design students from the Bachelor of Design programme at AUT University. Using my illustrations as reference, a seamlessly loopable animation of Moolah grooming himself was developed. The digital design students constructed a digital puppet of Moolah, creating a basic skeleton that enabled basic mobility and 2D movement of joints for the animation. This animation sequence was based around Moolah's grooming routine. The reader could direct Moolah to comb his fancy moustache, then his hairy arms and finally his leg hair. After this sequence the reader could decide what part of Moolah's body needed more brushing. If the system heard the word 'Leg' then the animation sequence of Moolah brushing his leg would play. The same applies for his moustache and arm combing sequence. Touch points were embedded onto Moolah's body parts, so the user could touch Moolah's head to play the moustache combing sequence and torso for his arm, and legs for his leg hair combing sequence.

Moolah was the first prototype where sound was added to the animation. Sound effects were played during the action sequences in the video, and a confirmation noise gave the user positive feedback when the system heard the correct words spoken. Adding sound to an animation increases the animation's impact, adding depth to the characters and animation sequences.

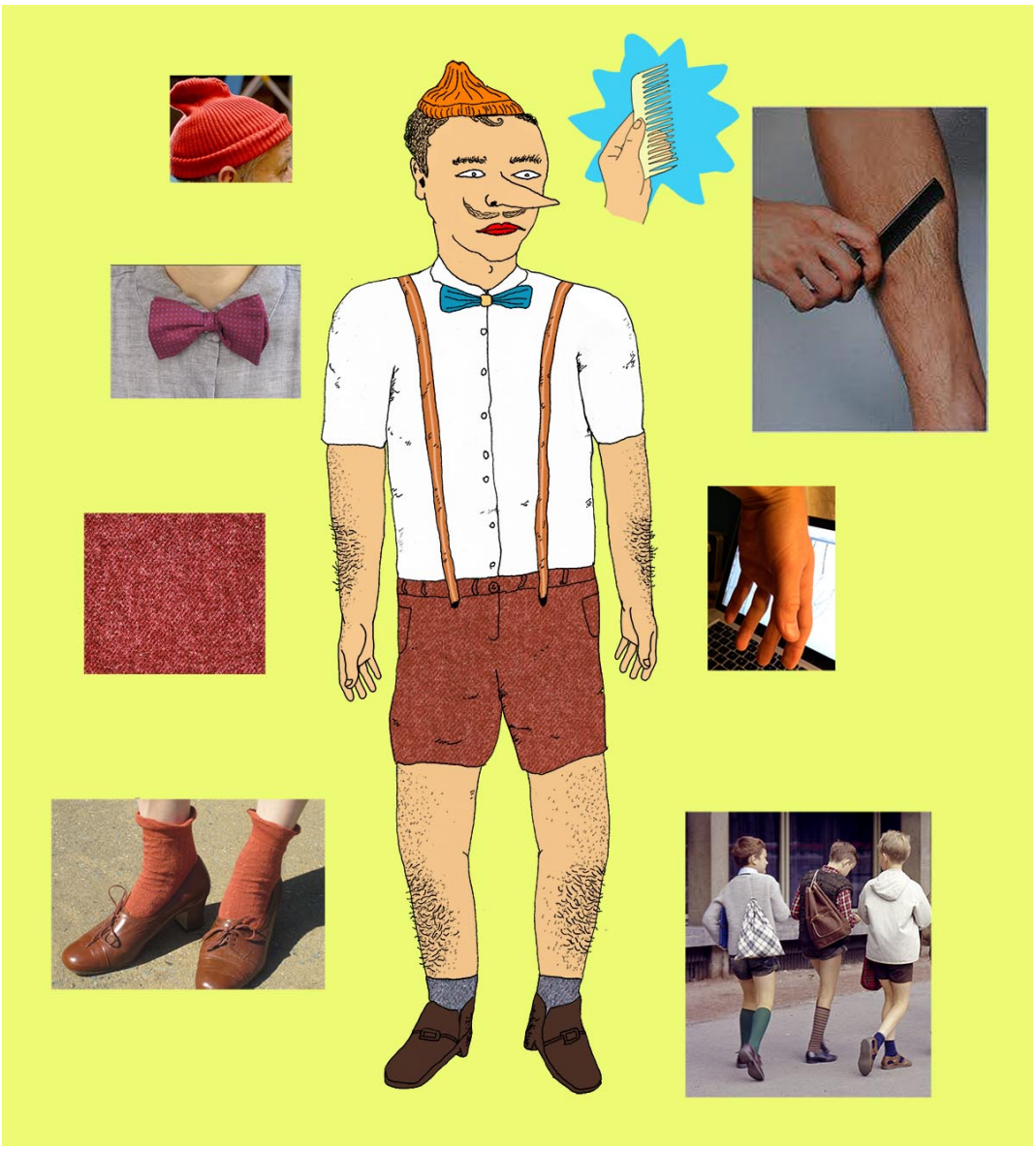

Figure 18: Señor Moolah character development with visual references.

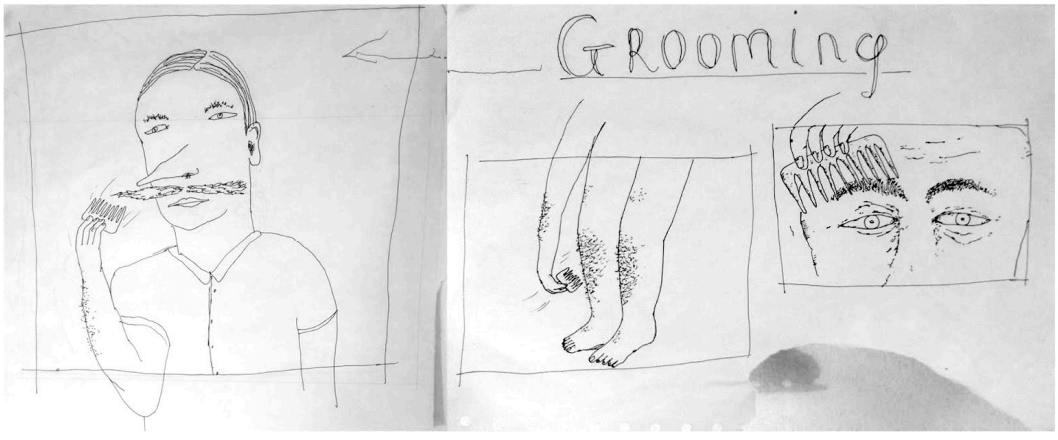

Figure 19: Early sketches of Señor Moolah's grooming sequence.

Figure 20: Animation created by the digital design students[4].

Through the testing of this prototype a usability concern related to the duration of the waiting sequences surfaced. The waiting sequences are short loops that play when the system is listening for the user to read a sentence. When the system hears the sentence spoken, it plays a confirmation noise and when the waiting sequence reaches the end, instead of looping again, it progresses onto the action sequence. Some of the looping waiting sequences were too long in this prototype. So when the user spoke the correct sentence, the time required to finish the waiting sequence was too long which created a delayed effect in the media playback. To counter this, the waiting sequences' maximum duration was set to 2 seconds long, as 2 seconds should provide ample time for an interesting waiting loop, while mitigating the delayed effect for the user.

[4] *The video can be viewed online here: https://vimeo.com/98836309*, password: talkingbooks

4.2.2 Paper Prototype session 2, July 2014

The paper prototype session was held at AUT with a group of my peers, who re-examined the systems functionality, and experimented with narrative conventions. I wanted to understand how the narrator's perspective affects the user's engagement with the story.

A cardboard prototype was constructed that involved three groups of people for operation. Firstly, the reader who represented the user was presented with a story that they had to read clearly out loud. Next, the actor/s performed the stories narrative in sync with the user's reading speed. The final person was the technician who progressed the narrative at the user's reading speed and indicated the reader's current word by signalling with their spare hand.

Figure 21: The second paper prototype session conducted with my peers.

Two stories were prepared for this paper prototype - the initial story was told from a first person perspective, and the second story from a third person perspective. As the software progresses the story as the user reads the narrative out loud, I wanted to investigate if the perspective in which the narrative was read affected the reader's engagement with the story. I speculated that there was potential for the reader to feel connected to the story if he/she related to the protagonist of the story, as they are directly involved in advancing the stories narrative through their voice. For example when the user read "I walked up to Steven and crushed him with my little pinky" they see would see the actor carrying out this action, in sync with the reader's speech. What dynamic would this create between the reader, the story and the protagonist?

This cardboard testing generated a number of interesting questions around media and narrative related to this project. A pertinent point was raised about the lack of reader choice when the narrative was set in first person. The first person perspective forces a direct relationship between the reader and the protagonist of the story. However there was no opportunity for the reader to influence the story as the narrative was pre-defined and set in a linear fashion. If the reader does not relate to the protagonist or agree with their actions, they may develop a negative reaction to the story and underlying system. The reader may feel uncomfortable and forced down a path that they would not take if they had the choice.

This problem is called the 'Narrators Paradox', where there is a conflict between the narrator and the player's control (Louchart & Aylett, 2003). Interactive narrative is an alternative and dynamic form of narrative that has been seen in video games such as *The Stanley Parable* (Galactic Cafe, 2013), and novels such as the *Goosebumps (Stine, 1992)*. Interactive narratives provide the reader with a feeling of control over the story, as they are able to make decisions at points in the story, determining the flow and the outcome of the story.

Including interactive narrative structures into future *Talking Books* stories is possible on the versatile iPad platform. With interactive narratives developed, more research into the relationship that narrator perspectives has on user experience should be conducted as it appears to be a fertile ground for user engagement.

This paper prototype session investigated the underlying functionality of the system and explored some story conventions of the media. Through this testing session, a clearer understanding of how the system would operate with users, and encouraging feedback was received from the participants around the systems functionality.

4.2.3 Luca the Climber

Luca the Climber is the latest story developed for the *Talking Books* project. The narrative was developed in a similar manner to Señor Moolah's story, but extra consideration was taken to ensure the story was appropriate for the targeted audience of 6-8 year old children. Advances were made in the content and technical aspects of the application in this prototype. These are detailed in the following section.

Technical Development

A team of three computer science students joined the development team for the Luca the Climber prototype. These students came from programming backgrounds and applied their expertise, providing a critical and technical perspective to the project. Some significant technical advances made in the Luca the Climber prototype are considered in this section.

Story Loading

Originally the stories were hardcoded into the system, where the stories properties were defined within the main application. This is not a recommended way of loading media, as it required editing the main source code of the application when adding or changing a story, which can easily lead to other problems. The computer science team created a method that read the story information from an external text file. This enabled easier and more efficient story loading, as well as cleaning and simplifying the story-loading process. Figure 22 shows Luca the Climber's story text file, which defines the video file name, scene timings, scene type, and narrative of the story.

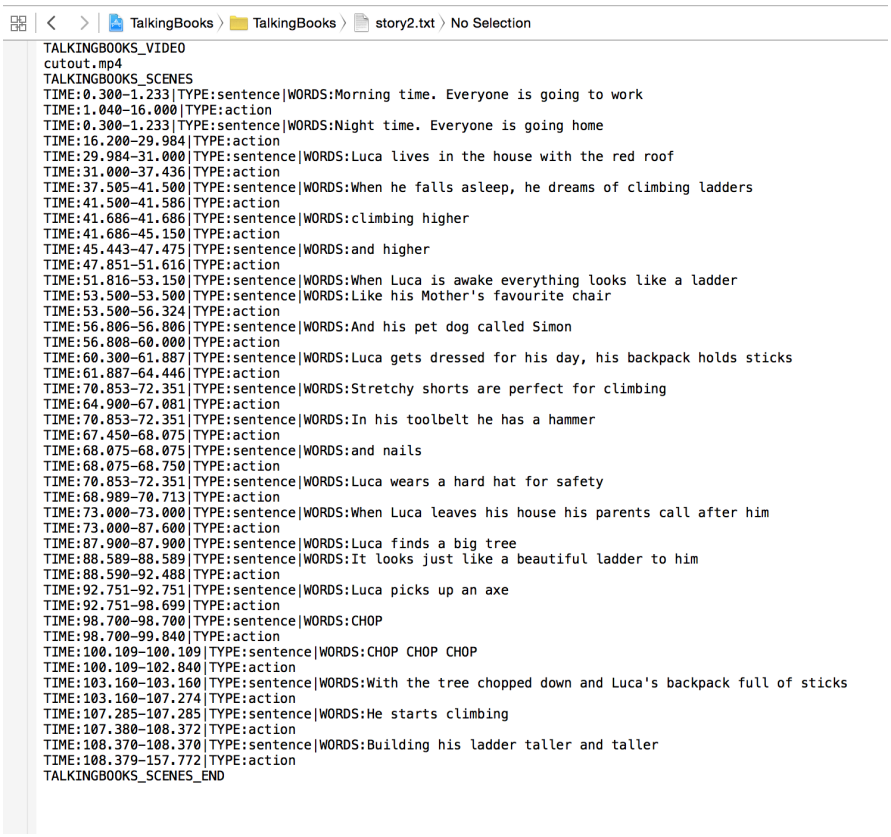

Figure 22: Luca the Climber's story text file.

RapidEars + Rejecto

The OpenEars speech recognition engine is a free and open source library that offers additional plugins that, users can purchase to improve the speech recognition's accuracy. Two plugins were purchased during for the Luca the Climber prototype – RapidEars and Rejecto. They improved the speech recognition performance.[5]

Word Highlighting

[5] Rejecto: This plugin prevents the speech recognition engine from recognising words that are not part of the defined vocabulary. Often during testing, foreign words are interpreted for words in the vocabulary - this plugin reduces the likeliness of this happening.

RapidEars: This plugin offered real-time speech recognition processing. This capability was essential for *Talking Books'* intended functionality. Real time speech recognition enabled improved interactivity, and real time feedback for the user during operation.

After implementing the RapidEars plugin, the word highlighting functionality was developed. Word highlighting was a significant user interface development that indicated which word the system was currently listening for, and provided real time feedback to the user during operation. Additionally, when a reader spoke a word in the sentence that was not the current word, this word would glow momentarily, giving the user feedback that the system heard the word spoken. Figure 23 shows the four states of the narrative; the spoken words are faded out yellow, the current word is a flashing bright yellow, the future words are matt black, and the spoken future words glow white then fade to black.

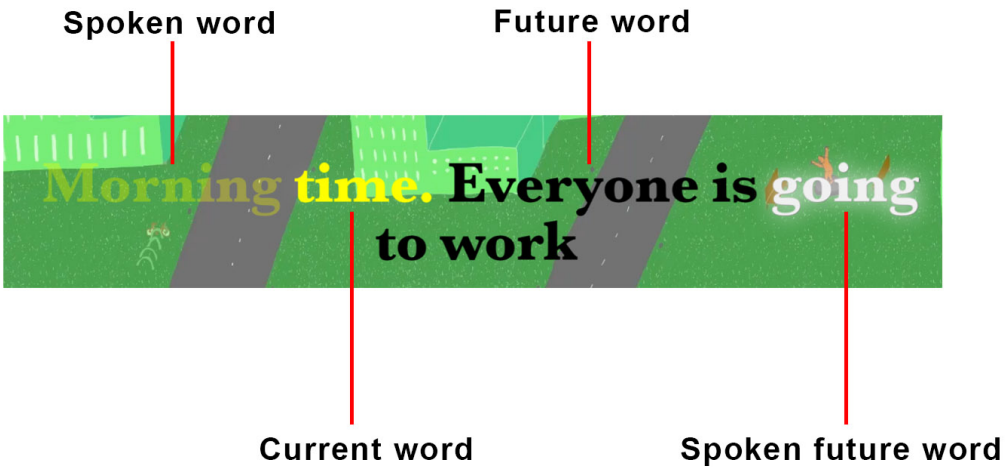

Figure 23: Narrative display showing word highlighting.

Touch to hear words

A popular feature in many children's digital storybooks is touching a word to hear its pronunciation. This provides support to the user if they should get stuck on an unfamiliar word. This function was programmed to automatically assign sound files to their corresponding words, avoiding hard coding each sound file to their corresponding word. Each word in the narrative was recorded and saved as individual sound files that were automatically assigned to the words in the user interface.

Audio waveform

The addition of an audio waveform that visualised the microphones input volume was added. This display communicated to the user that the device was actively listening to them, replacing the earlier ear/mouth displays. This visualiser was found as an open source project on *GitHub* and was similar in appearance to the visual interface of Apple's speech recognition program SIRI. It displayed an oscillating sine wave, which would distort relative to the input decibel. This was tested but did not fit with the visual aesthetic of the application, so was removed.

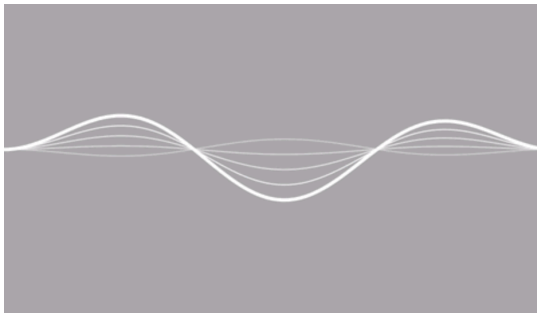

Figure 24: The SIRI waveform view

Updating dictionary each scene:

Through brainstorming methods to improve the speech recognition performance, we analysed how the speech recognition was operating. The words in the speech recognition dictionary corresponded to the narrative in the story file. Every unique word in the narrative was added to the possible word input, and the Rejecto plugin ignored any word that did not match this array of words. This meant the vocabulary was looking for some 100 possible word inputs each scene. However, this process could be optimised, as there were only 2-12 words visible in each scene. Instead of comparing the speech input to every word in the narrative, we could compare to only the words present in the current scene. This would reduce the possible inputs substantially, increasing the speech recognition performance.

Technically this proved difficult to implement, as it required restructuring the vocabulary loading process. The computer science team partially got this working but the results were disappointing, and it was difficult to see what impact this made on the speech recognition performance. Due to time constraints, this functionality was omitted from the final version, but would be a promising area for future development.

User Interface development:

(Weiser, 1991) recognised that the most profound technologies are those that disappear; weaving themselves into the fabric of everyday life until they are indistinguishable from it. Recent developments in mobile technologies are advancing towards ubiquitous computing; a "shift to human-centred computing, where technology is no longer a barrier, but works for us, adapting to our needs and preferences" (Becta, 2007). Features like touchscreens and speech recognition support the user intuitively interacting with their devices. Buzzwords such as 'gestures' and 'swipe' have been adapted from our anatomical vernacular and now have a place within our technological vocabulary.

The user interface was developed to facilitate an intuitive understanding of the systems operation. Unnecessary functionality and buttons were omitted from the interface design, displaying instead the essential components comprising of the story/visuals full screen, the corresponding narrative and an arrow to advance through scenes. The narrative word highlighting functionality was a main component of the user interface. It served to communicate the interaction between the speech recognition, the user's speech and the narrative (see section 4.2.3, Word Highlighting).

Directing the user's attention during the application's operation was an important consideration. During the waiting scenes, short loops of animation played and the user was presented with a narrative to read. The looped animation served as a supporting detail, with the narrative in the foreground of the user's attention. When the user finished reading the sentence out loud, the narrative disappeared and the user was presented with an action scene that progressed the story. Additionally a confirmation sound effect was used to give the user feedback that they had completed the sentence. During the standby scenes, the short looped animations directed the user's attention to the reading component, combined with the glowing words drawing attention to the user's reading progress.

Content Development

Luca the Climber's narrative was developed through exploring themes that interested me, sketching ideas and building a narrative around them. I found the ladder an alluring metaphorical, and visually enticing image which I focused the story around. Additionally, the story of Luca the Climber is a more relatable story that children could relate to, compared to the previous Señor Moolah story which was abstract and distant from children's lives. I was interested in the repetitive sequence of climbing which can be looped to create an infinite climbing animation (see Figure 25). Below is a synopsis of Luca the Climber's story.

Luca the Climber is a story about a determined young boy who is obsessed with ladders. Luca is a recluse in society seemingly absent in conversation with friends and family as he spends every minute of the day thinking about the ladder he is building. Everyday his parents are driven further up the wall by Luca sneaking out and cutting down their trees to make wood for his growing ladder. This project of Luca's does not stay unnoticed, and soon it is a growing attraction in his small town, with people far and wide coming to see Luca hard at work.

One day while Luca is far up the ladder, so high up that the birds no longer keep him company, he looks down at the world and sees a giant crack in the earth hurtling towards him. Too high to get down in time, he holds on for his dear life and gets shaken to the bone. As the tremors reduce he makes his way down the ladder to discover the tremendous shaking was an earthquake. There is a big commotion next to the river that separates the two sides of Boobaloo city so Luca investigates. He discovers the sole bridge connecting the two sides has washed away leaving everyone very upset. Loved ones desperately yell across the river at each other and Luca's family dog barks at him from the other side.

Luca develops an ingenious plan and runs back to his house to fetch his oversized ladder. He sacrifices his hard work and lays his ladder across the river, connecting the two sides, and helps everyone return home to their loved ones. Later the mayor of the town is so impressed by the structural integrity of Luca's ladder that he appoints Luca the lead engineer and designer for the replacement bridge. Luca gratefully accepts the challenge and builds a beautiful ladder bridge to serve the people of Boobaloo city.

Animation style and influences:

The animations for Luca the Climber were created digitally using a drawing tablet in Adobe Photoshop. Adobe Photoshop has a timeline feature where short animations can be created by ordering the frames on different layers and assigning duration intervals between the frames. The digital process was far more efficient compared to the previous hand drawn Lucy hula-hoop animation. It simplified the animation development, cutting out many cumbersome steps and saving a substantial amount of time. Working digitally, mistakes are easier to correct, frames are easier to compare with one another, and the animation progress is easily analysed through previews. However there is a distinct and intrinsic difference between traditional illustration and digital Illustration. In traditional illustration, ink bleeds into paper differently to the way pixels are digitally generated on a screen, giving the animation a distinctive visual aesthetic and overall feel.

Figure 25: A loopable sequence of Luca climbing.

The animation frames were created in *Adobe Photoshop*, and the sequences were composited into *Adobe After Effects. After Effects* was also used in some frame development to create a morphing effect in the chair-ladder and dog-ladder sequences. This morphing effect was created by blending together two images and animating the pixel difference. This can be seen in

Figure 29, which shows key frames from the chair morphing into a ladder chair, and the dog morphing into a ladder dog sequences.

The animation style in the Luca the Climber story is very different to the previous Lucy Hula Hoop animation. The rough stipple pen style used in the Lucy animation was replaced by flat blocks of colour. These flat blocks of colour are easier to manipulate and animate, as the animator is working with shapes instead of multiple lines. The aesthetic was influenced by surrounding visual media that contributed to the look and feel of Luca the Climber's animation. Figure 26 shows some examples of influential animations.

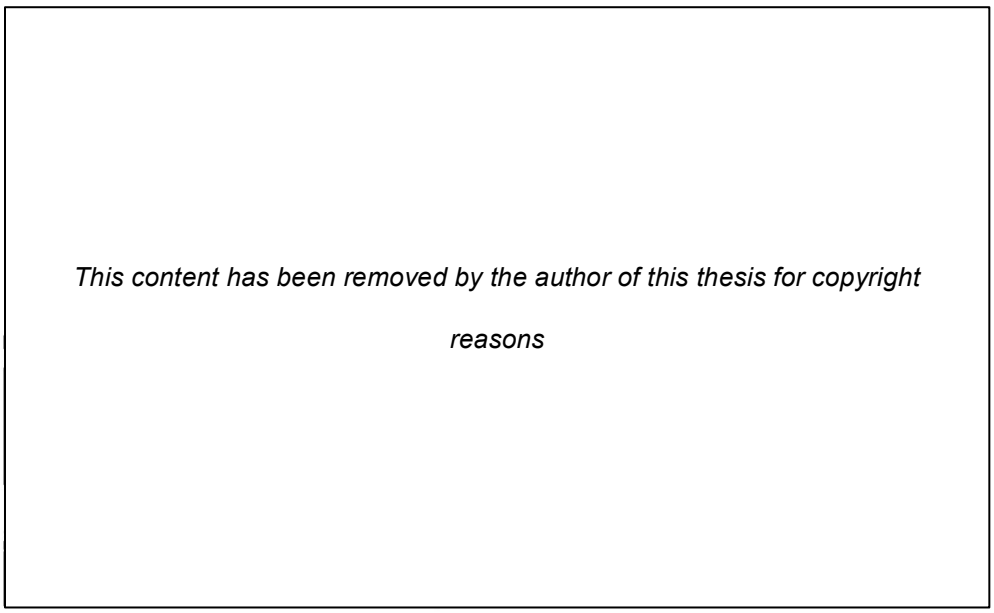

Figure 26: Animation influences

Clockwise from top left: The Fableists "Epic Thread" by The Mill, Small Circuits by Grace Helmer, Somewhere by Nicolas Mérnard, Not About Us by Michael Frei.[6]

[6] These influential animations were mainly 2D animations with flat pastel-like colours. The Fableist's "Epic Thread" (The Mill, 2014) animation is an exception as it included shadows, giving the animation a sense of depth.

• The Fableist's video had a clean and appealing style that is attractive to children.

• Grace Helmer's *Small Circuits* (Helmer, 2014) was composed of oil paintings. Each frame painted with beautiful tones. Helmer's colour palette and simple animations provided a visually appealing short video.

An important consideration during the design of the animation style is the relationship that the illustration style has with the animation workload. When illustrating a single frame, the more detail in this frame substantially increases the overall animation workload. Simplifying the animation into blocks of colour was an aesthetic and practical consideration, reducing the overall animation workload.

***Animation technique*:**

Initially the visuals elements in the story were sketched out using pen and paper. These sketches evolved into a storyboard that established the chronological order of the visuals. This storyboard was then developed into an animatic[7] that developed the timings of the shots and animations, as well as overlaying the narrated audio over the story. With the animatic created the story was split into scenes and started developing the animations within these scenes. The animation process for Luca the Climber was similar to the earlier Lucy animation, except created digitally on a computer.

Animation frames were created in Adobe Photoshop by drawing basic outlines, then filling these outlines with colour. Frames were easily duplicated and edited to create subtle movement. Visual references were sourced from books and the Internet that were used to support the animation development. Figure 27 shows the animation development of Luca rolling over sequence.

• Nicolas Mérnard's *Somewhere* (Mérnard, 2015) uses textures and offset registration to create an interesting visual aesthetic. Bold flat colours create vivid imagery within this animation.

• Not About Us (Frei, 2013) is an abstract black and white animation created with traditional cel techniques. Frei uses strong contrasting colours and imagery with a rough illustrative style, using a crinkled paper texture as a background giving the animation a tactile feel.

[7] The animatic video can be seen here: https://vimeo.com/129949112, password: talkingbooks

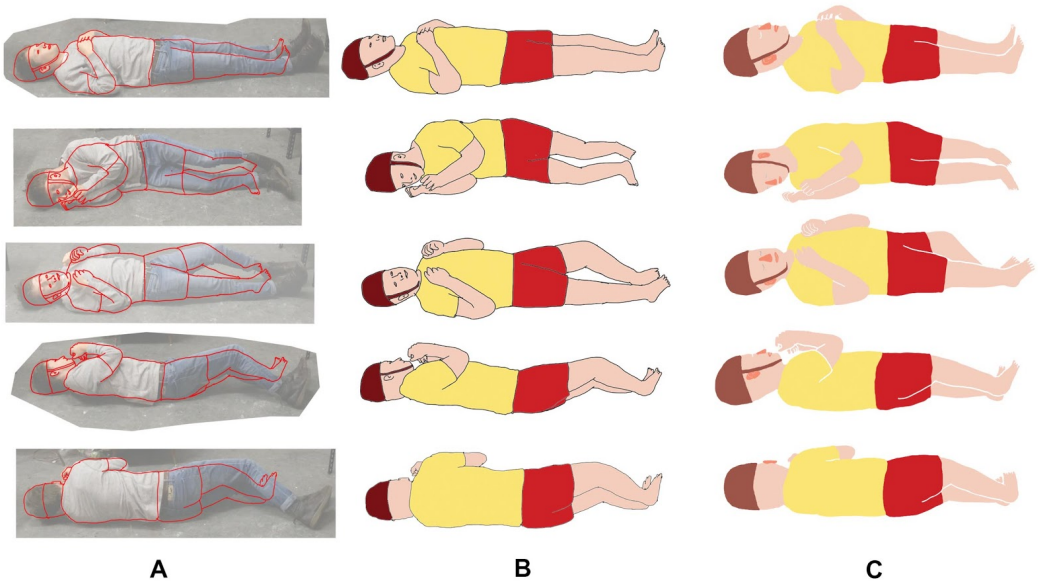

Figure 27: The animation development of Luca rolling sequence.

Within Figure 27, A is the initial step where the reference image was created by filming myself replicating the desired animation. Key frames from the video were selected and imported into Adobe Photoshop. Next these key frames were rotoscoped[8] and the legs were shortened to align to Luca's intended age. In step B, the outline was filled in with blocks of colour and the outline was changed from red, to a black line. In this step it became apparent that Luca's proportions were not accurate. In Step C Luca's body proportion was corrected and the black outlines were removed.

[8] An animation technique where frames are traced using video footage as a reference.

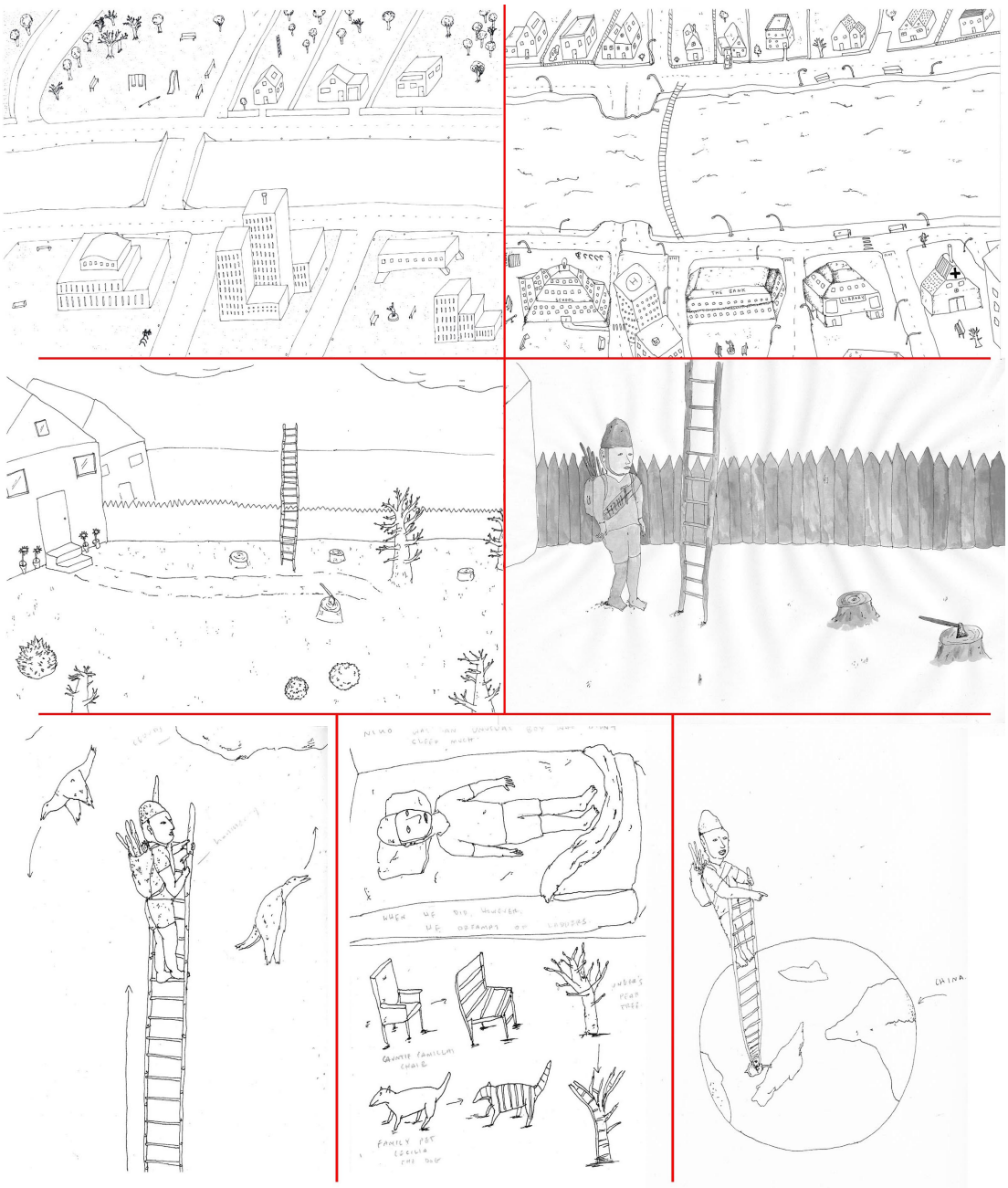

Figure 28: Early sketches and illustrations of Luca the climber.

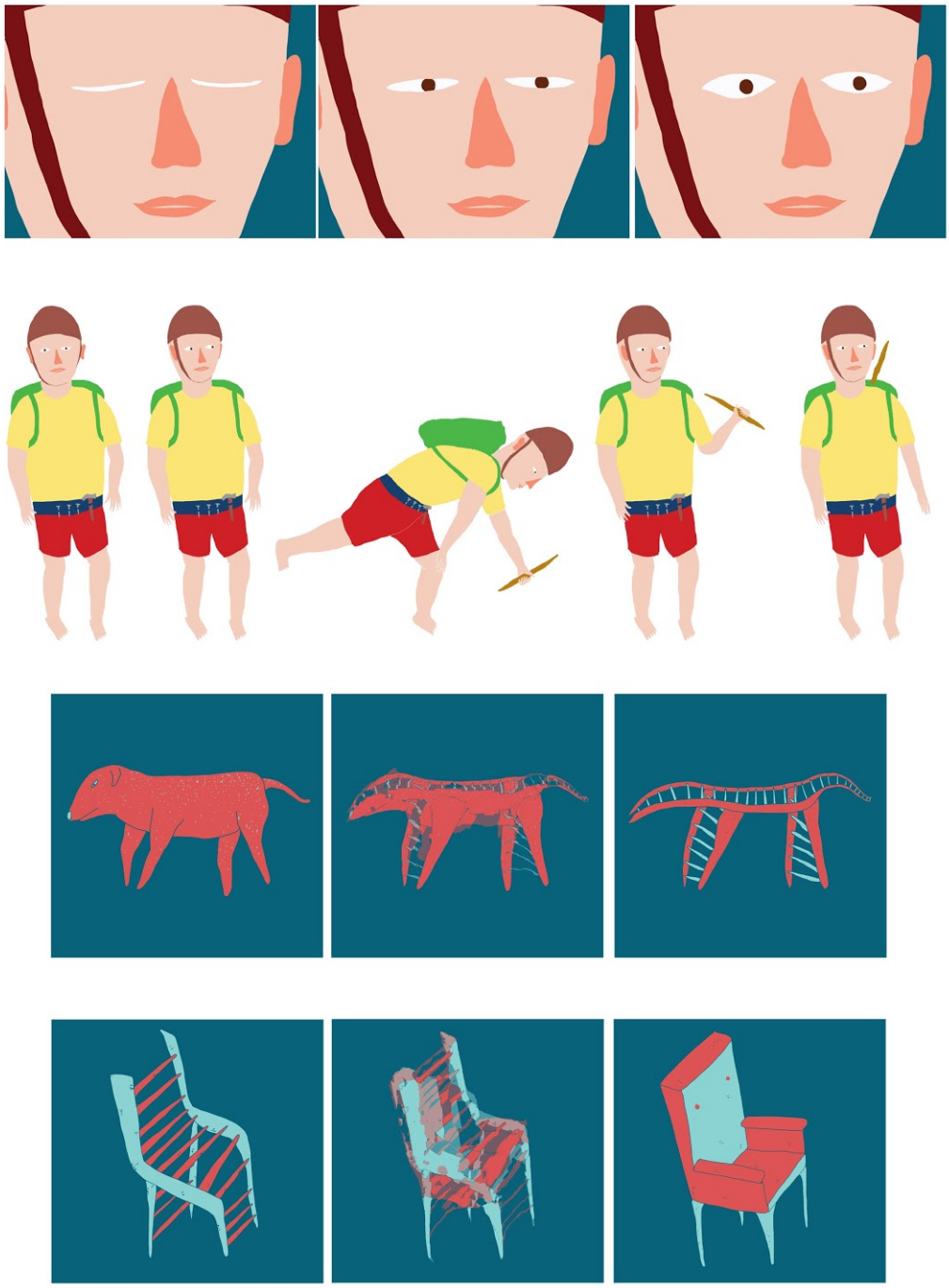

Figure 29: Animated stills from Luca the Climbers animation.

Sound design development:

The sound that accompanied the Luca the Climber animation was developed through collaboration with sound artist Alex Guthrie. Alex's expertise in audio engineering, specifically electronic production and field recordings was sought for the project.

During the sound development an important consideration was *when* the sound could accompany the application. As the application manipulates the animation playback, creating loops (see appendix IV), and the main functionality uses the inbuilt microphone, sound could only be played during specific intervals in the animation. If sounds were playing when the microphone was active, these sounds would compete with the reader's voice and cause an adverse effect on the speech recognition performance.

Conclusion of Luca the Climber Prototype

The animation of the Luca the Climber's story was half finished due to time constraints. The animation proved to be a very time intensive process as I was producing all of the visual content independently. Given the time deadlines and other aspects of this project I decided not to rush the second half of the story, but instead, finish the story at a dramatic moment as a cliff-hanger. This did not have a negative effect on the prototype application, as the animation is long enough to experience how the system operated.

5.0. Testing and Results

Testing was conducted to gain feedback from the targeted user group during the development of the application. This feedback was used to inform the design process, to gain an understanding of how the software performed with its users, and how the users responded to the software. I used the following questions to guide the testing:

• How does the speech recognition perform with the targeted user group?
• Do the participants understand the system's operation?
• How do the participants respond to the story content? Specifically - animation, sound and narrative?
• Is the vocabulary at an adequate level for the intended users?
• Are the supportive features functional?

Subjects

Two schools were visited, with each school providing 10 grade-one students aged between 5 and 7 years old. The participants were included through a voluntary sign up process with a parental information and consent forms sent home with the children. The two schools that hosted the testing sessions were Kadimah School and Birkdale North Primary School.

Kadimah School is an inner city primary school providing education for children in preschool up to year 8 in a contemporary Jewish environment. Kadimah is a decile 10 school with a sizeable community of South African children. In the testing at Kadimah, there were 6 English-speaking children, 3 bilingual children and 1 trilingual child. Two of the bilingual children spoke South African and English, with the remainder bilingual participant speaking Vietnamese and English. The trilingual participant spoke English, Italian and Chinese.
Birkdale North Primary School is a multicultural public school with a decile rating of 4. Birkdale North has a popular bilingual French program and half of the children in the testing were native French speakers, with English as their second language. 5 participants were of French descent and the remaining New Zealanders.

Process

In both of the testing sessions I was assisted by a colleague to manage the task of observational note taking and supporting the participants during testing. Catherine Ellis, who teaches young children and is familiar with the research project, assisted me for the testing session at Kadimah College. At Birkdale North Primary my research partner Sam Joe assisted me.

Pairs of participants aged 5-7 years old were taken into a room to user-test the app. Within the room, an audio recording device recorded the playtest. After introducing ourselves to the participants, we asked a few preliminary questions around iPads and their experience with storytelling apps on tablet devices. Next we introduced the *Talking Books* app, giving a short demonstration of the functionality behind the software's operation. We observed the user testing through observational note taking as well as recording the user testing with an external microphone. The software testing took roughly 10 minutes, leaving 5 minutes for some unstructured indicative questions based around user experience and the systems flow. These questions included:

• What did you think about the application?

• What do you think about speaking to advance the story?

• How well did you think it worked?

• What part did you like the most?

• What didn’t you like about it?

Results and discussion

The testing revealed the weak performance that the speech recognition had with the target audience. This speech recognition was an essential aspect of the application and its performance proved to be a detrimental factor affecting user experience. Although the application had been robustly tested internally prior to user testing, the speech recognition was not versatile and did not perform with the intended younger users.

Table 2 shows the speech recognition scores and meanings, which were used in the results table in Figure 31 and 32.

Score	Meaning
0	None of the speakers input was understood.
1	0 - 10% of the speaker's input was understood.
2	11-20% of the speaker's input was understood.
3	21-30% of the speaker's input was understood.
4	31-40% of the speaker's input was understood.
5	41-50% of the speaker's input was understood.
6	51-60% of the speaker's input was understood.
7	61-70% of the speaker's input was understood.
8	71-80% of the speaker's input was understood.
9	81-90% of the speaker's input was understood.
10	91-99% of the speaker's input was understood.

Table 1: Speech recognition scores and meanings.

Birkdale North Primary School

Number	Background and age	Speech recognition accuracy	Comments from participants	Notes
1	**A**. 6 year old boy, English second language, native french speaker. **B**. 7 year old girl, English second language, native french speaker.	**A**. 2/10 **B.** 2/10	"Actually you might like giving us something harder then that, it's too easy".	**A.** was a confident reader with a detectable French accent. He got frustrated when the application didn't understand him and raising his voice to be understood. Enjoyed the mom yelling unintelligibly at Luca, instead of words, they said it was funny.
2	**C.** 6 year old boy, English second language, native french speaker. **D.** 6 year old girl, English second language, native french speaker.	**C.** 3/10 **D.** 1/10	-	Speech recognition worked well for A, his voice was suited to the story. Gaelle was quiet with hands obscuring her face, no words were registered when she spoke. Xavier repeating words a lot, voice loud and clear. Both children were quite shy and did not make any comments.
3	**E.** 7 year old boy, English second language, native french speaker. **F.** 7 year old girl, NZ Born, native english speaker.	**E**. 2/10 **F**. 1/10	-	The girl was quiet and shy, and the system did not understand anything that she said. After saying it once and it not progressing she stopped. Touching the screen/words a lot. **E** raises his voice in frustration. When **E** came across a word he didn't know he touched the word to hear it's pronunciation.
4	**G**. 5 year old girl, NZ born, native English speaker **H**. 7 year old boy, NZ born, native English speaker	**G**. 1/10 **H**. 1/10	"There it is! there's the one with the red roof and the brown walls!" engaging with the story. Favourite part was the ladder chair.	Mid-way through explaining it, **G** began reading with good elocution. when she got to the end of the sentence she paused and said hmm as it didn't work.
5	**I**. 7 year old girl, native english speaker, NZ born **J**. 6 year old girl, native english speaker, NZ born	**I**. 2/10 **J**. 1/10	"Sounds like someone is using their mouth to make the noises"	When having trouble reading a word the other participant says 'you can just tap on it'. Thereafter they used this feature often. Lots of giggling especially around sounds.

Figure 30: Results tables outlining the Birkdale North testing session.

Kadimah School

Number	Background and age	Speech recognition accuracy	Comments	Notes
1	**A.** 6 years old boy, native english speaker, NZ born. **B.** 6 years old boy, native english speaker, NZ born, parents speak mostly Vietnamese.	**A.** 1/10 **B.** 1/10	'*that doesn't make sense*' difficulty following the narrative. *'that looks like Mr. Bean'.*	Most words were familiar, some words such as '*toolbelt*' and '*stretchy*' caused difficulty. Lots of giggling with the car noises.
2	**C**. 6 years old boy, NZ born, parents speak South African. **D**. 6 years old boy, NZ Born, native english speaker.	**C**.1/10 **D**.1/10	"I like it because it plays a video".	Speech recognition did not work, **C** quickly discovered the hidden button to advance scenes/ **C**. noticed that the system was not working for his voice and made frustrated sounds when it didn't work.
3	**E**. 5 years old girl, NZ Born, native english speaker. **F**. 5 years old boy, NZ Born, native english speaker.	**E**. 1/10 **F**. 1/10	Currently **E.** was reading chapter books.	Touching the screen/words a lot. understood the function of the flashing words, repeats the word 'his' many times, clear frustration in her voice as the application didn't understand her. giggling at the parents yelling scene
4	**G**. 6 years old boy, NZ born, native English speaker. **H**. 6 years old girl, NZ born, native English speaker.	**G**. 1/10 **H**. 1/10	-	unconfident readers, speaking with hesitation and slowly sounding out words. raises voice when the application did not progress.
5	**I**. 7 year old boy, NZ born, parents speak Chinese and Italian. **J**. 6 year old boy, NZ born, parents speak South African.	**I**. 2/10 **J**. 1/10	Do you think it's listening to you? - *"No, I think it's deaf."* What do you think you have to do? *"scream/ shout?"*	Raises voice when his reading was not recognised. Visibly frustrated when it didn't work. Reads Tintin, and Captain Underpants.

Figure 31: Results tables outlining the Kadimah testing session.

Once the technical difficulties around speech recognition became apparent, additional support was provided to the participants, including providing positive feedback and prompting the participants through the story. A hidden button was then used to manually override the speech recognition and advance the scenes. This intervention was justified as we considered if no additional support was provided, the application would not have been usable preventing any useful feedback being gained from the testing.

The speech recognition difficulties created confusion around the software's operation as the instructions that were explained to the participants conflicted with the applications actual performance. The iPad was described to some participants as having ears and listening to the readers during the story. When the application failed to register the reader's voices, some of the participant's natural reactions were to speak louder to the device. Participant **J** from the Kadimah testing said "the application was deaf" when the speech recognition failed to understand his voice.

There was a noticable relationship between the speech recognition performance and the gender of the participants. The speech recognition had poorer results when interpreting female voices, with their naturally higher pitched voices. Participant C from Birkdale North Primary School had the best speech recognition results and was a boy with the deepest voice in this set of testing.

The participants responded well to the sound aspects of the story, specifically the car noises that were made by recording mouth noises. There was lots of giggling specifically with participant A and B from Kadimah School during these scenes. The participants gave generally positive feedback around the visual elements of the story, complementing the colours and engaging with the story and animations.

Critical Analysis:

Many of the participants became annoyed when the speech recognition failed to work with their voices. They were prompted to read slowly and clearly so the system could understand them. However, even with perfect elocution the system failed to understand their voices. When the

system failed to register words that the users considered easy and were confident with speaking, the users lost motivation and became visibly frustrated (see participant E in Figure 32).

As the application is being developed as a potential learning tool for children developing their reading skills, the system must be versatile, robust and accurate. If the system fails to work for all potential users, the system may have a negative, rather than beneficial effect on children's literacy development. For example, in the case where a child correctly reads a word out loud and the system fails to recognize it, the child may question their own pronunciation, assuming that it must be wrong.

This problem lies in the technical aspects of the project, specifically, the speech recognition engine. The user group will not train their word elocution to match the speech recognition's phonetic interpretation, instead the speech recognition engine needs to be versatile to accompany the entire range of the user's accents and unique voices. As speech recognition relies on a clear audio input, good enunciation is essential for the application to understand the user. If the reader is mumbling or talking away from the microphone, the speech recognition performance is affected.

Technical Summary and Future Technical Developments:

The technical difficulties that were encountered during the latest testing sessions were not anticipated. Although the system had been robustly tested with our internal team and with friends, little testing had been conducted with our intended users. The best speech recognition performance was achieved with my voice, able to achieve a recognition score of 9-10 (see Table 2). This was likely a result of inadvertently altering my pronunciation to align to what worked best with the speech recognition engine, through the testing stages of the project.

Future technical developments would see the speech recognition optimised for the intended child user group. Replacing the adult acoustical models with child-trained models greatly increases speech recognition performance (Elenius & Blomberg, 2005). There are pre-existing child trained language models exist such as *The OGI Kid's Speech Corpus and Recognizers* (Shobaki, Hosom, & Cole, 2000), but these models are not specific to our testing demographic as they were recorded using US English. Ideally to achieve the best results, the intended user demographic would create a new language model that was trained using the storybook corpus. Additionally, further mathematical techniques to optimise speech recognition for children have been developed, and often the best results use a combination of these techniques (Shivakumar, Potaminaos, Lee, and Narayanan, 2014).

Content Summary:

Luca the Climber and previous story prototypes created during the development Talking Books were played in the form of rendered video files. The video files were structured in a specific way, which allowed the system to play through them. The application manipulated video playback during operation, however as the video was pre-defined, it lacked flexible interaction. Future *Talking Books* stories could incorporate procedural animation. Procedural animation (see section 2.4.2) allows for dynamic content interaction, where users can manipulate visual elements and have more control over media playback. Utilising this animation technique may enable a more immersive experience as it empowers users to affect and personalise their experience of the story.

Additionally, interactive narrative can be used to enhance user immersion into the story by empowering the user to influence the story through their actions. This storytelling technique

(discussed in section 4.2.1) aligns well to the digital platform and should be considered for future stories, as it potentially improves user engagement with the story.

6.0 Conclusion

In this research, a literature and contextual review explored existing media, and the surrounding areas related to this project. This section aimed to locate and inform the project, while supporting the research rationale. The Methodology section explained how the research was approached and conducted. The Project Development section outlined the practical developments that were coordinated through the creation of iterations of prototypes. The significance of each prototype was acknowledged and used to inform the future developments. The Results section detailed the user testing sessions, and followed up with a discussion. Through user testing the application, encouraging feedback was gained but the underlying technical speech recognition performance proved to be poor with the intended target users - children aged 6-8 years old.

The system's proposed functionality can be demonstrated when using the app with adults; however the main challenge this project currently faces is achieving satisfactory results with the intended child audience. Improving this technical problem is a priority for future developments, with some techniques outlined in the previous *Technical Summary* section. Additionally, future content developments have been outlined in the previous *Content Summary* section, and include:

• Creating stories using procedural animation to increase user interaction.
• Incorporating the iPad's interactive features into future story narratives.
• Embedding interactive narrative structures into stories to increase user engagement.

Following this further development, the research indicates that the *Talking Books* system will be useful to the next generation of early readers, publishing high quality children's material that is experienced though a novel interactive system. While not fully realised in this research, the development to date indicates that this platform will encourage collaboration between storytellers, animators, programmers and artists to create future stories for this new genre of children's storytelling.

App design and development is a complex and interdisciplinary process that is generally confined within commercial realms. There is a limited amount of literature examining this process from the designer's perspective. This work has explored the creation of an educational app and looks to add to the body of knowledge within this specialised field.

References

• Adams, M. J. (2011). Technology for developing children's language and literacy: Bringing speech recognition to the classroom. In New York, NY: The Joan Ganz Cooney Center at Sesame Workshop.

• Apple Insider (2013). IDC sees tablet shipments eclipsing notebooks this year, all PCs in 2015. Apple Insider retrieved from http://appleinsider.com/articles/13/05/28/idc-sees-tablet-shipments-eclipsing-notebooks-this-year-all-pcs-in-2015

• American Academy of Pediatrics. (n.d) Media and Children, retrieved 11.05.15 from https://www.aap.org/en-us/advocacy-and-policy/aap-health-initiatives/pages/media-and-children.aspx

• Chaitin, J (2003) Narratives and Story-telling, Retrieved from http://www.beyondintractability.org/essay/narratives

• Colwell, E (1980): Storytelling Great Britain, The Bodley Head Ltd p.8

• Common Sense Media (2014) Jack and the Beanstalk by Nosy Crow, Retrieved from https://www.commonsensemedia.org/app-reviews/jack-and-the-beanstalk-by-nosy-crow

• Common Sense Media, & Rideout, V. (2011). Zero to eight: Children's media use in America. Common Sense Media.

• Council on Early Childhood (2014) Literacy Promotion: An Essential Component of Primary Care Pediatric Practice, peds.2014-1384;

• Becta, D, L (2007) Ubiquitous Computing. Emerging Technologies for Learning.

• Diehl J, Thüs H, (undated) Mobile Application Development, Retrieved from http://hci.rwth-aachen.de/tiki-download_wiki_attachment.php?attId=1218

• Elenius D, & Blomberg, M. (2005) INTERSPEECH, Adaptation and normalization experiments in speech recognition for 4 to 8 year old children, pp. 2749–2752.
• Frei, M (2013) Not About Us [Video File] Retrieved fromhttps://vimeo.com/51535113

• Gybels, G. (2010) Understanding Speech Recognition, Retrieved from http://www.guidogybels.eu/docs/Understanding%20Speech%20Recognition.pdf

• Galactic Cafe. (2013) The Stanley Parable [Video Game]. Retrieved from http://steam.steampowered.com

• Galloway, J. (2009) Harnessing Technology for Every Child Matters and Personalised Learning. New York: David Fulton Pub.

• Gelman, D. L. (2014). Design For Kids. Rosenfeld Media.

• Gray, S. S., Willett, D., Lu, J., Pinto, J., Maergner, P., & Bodenstab, N. (2014) Child Automatic Speech Recognition for US English: Child Interaction with Living-Room-Electronic-Devices.

• Gupta, H, Wadhwa S D. (2014) Speech Feature Extraction and Recognition Using Genetic Algorithm. International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 1, January 2014)

• Guthrie, J., & Wigfield, A. (2000). Engagement and motivation in reading. In M. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), Handbook of reading research: Vol. 3 (pp. 403-422). Mahwah, NJ: Lawrence Erlbaum.

• Hall, J (2012) E-Books could become the norm for children as sales soar, The Telegraph. retrieved from: http://www.telegraph.co.uk/news/uknews/9551136/E-books-could-become-the-norm-for-children-as-sales-soar.html

• Helmer, G (2014) Fundamentals of Neuroscience Presents: Small Circuit [Video File] Retrieved from https://vimeo.com/101425853

• Henderson, S., & Yeow, J. (2012). iPad in education: A case study of iPad adoption and use in a primary school. In System Science (HICSS), 2012 45th Hawaii International Conference on (pp. 78-87). IEEE.

• Higgins, S., Xiao, Z., & Katsipataki, M. (2012). The impact of digital technology on learning: A summary for the education endowment foundation. Durham, UK: Education Endowment Foundation and Durham University.

• House, R. (2012) The inappropriateness of ICT in early childhood education: Arguments from philosophy, pedagogy and developmental psychology. In: Suggate S and Reese E (eds) Contemporary Debates in Childhood Education and Development. Routledge: New York, pp. 105–121.

• Huggins-Daines, D., Kumar, M., Chan, A., Black, A. W., Ravishankar, M., & Rudnicky, A. (2006, May). Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices. In Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on (Vol. 1, pp. I-I). IEEE.

• Johnson, C. (2012). Reader-Animated Stories. Retrieved from http://www.twodee.org/forothers/proposals/reader_animated.pdf

• Kalman, M. (1990) Max Makes a Million, Viking Penguin

• Karsenti, T., & Fievez, A. (2013). The iPad in education: uses, benefts, and challenges – A survey of 6,057 students and 302 teachers in Quebec, Canada. Montreal, QC: CRIFPE.

• Kauffman, S. (2005).Story Elements: Which Impack Children's Reading Interests (Doctoral dissertation, Bowling Green State University).

• Knowles, M., & Malmkjær, K. (2002). Language and control in children's literature. Routledge.

• Kucirkova N, Messer D, Sheehy K and Flewitt R, S. (2013) Sharing personalised stories on iPads: A close look at one parent-child interaction. Literacy 47(3): 115–122.

• Lerner, J., & Tirole, J. (2001). The open source movement: Key research questions. European Economic Review, 45(4), 819-826.

• Louchart, S., Aylett, R. (2003). Solving the narrative paradox in VEs - lessons from RPGs. In: Intelligent Virtual Agents, 4th International Workshop p. 244-248.

• Lynch, J., & Redpath, T. (2012). 'Smart'technologies in early years literacy education: A meta-narrative of paradigmatic tensions in iPad use in an Australian preparatory classroom. Journal of Early Childhood Literacy, 1468798412453150.

• Maltepe, S (2014) The Importance of Literary Childen's Books in Preschool Teaching, retrieved 7.05.15 from http://www.tlu.ee/UserFiles/Pedagoogiline%20Seminar/International%20Week%202014/Sadet%20Maltepe%20lecture%20materials.pdf

• Mérnard, N. (2015) Somewhere [Video File] Retrieved from https://www.youtube.com/watch?v=HmVY3QuuDn8

• Merriam-Webster. (n.d.). Technology Retrieved June 22, 2015, from http://www.merriam-webster.com/dictionary/technology

• National Union of Teachers. (2013). Stereotypes Stop You Doing Stuff. Changing gender stereotypes through primary education. Retrieved from: https://www.teachers.org.uk/files/stereotypes-stop.pdf

• Neumann, S.B., Copple, C., & Bredekamp, S. (2000). Learning to read and write: Developmentally appropriate practices for young children. Washington, DC: National Association for the Education of Young Children.

• Nosy Crow, (2014). Jack and the Beanstalk, (Version 1.0.3) [Mobile application software]. Retrieved fromhttp://itunes.apple.com.

• Norvell, George W. (1958) What Boys and Girls Like to Read. Morristown, New Jersey: Silver Burdett.

• Nuance, (2015) Dragon Speech Recognition Software, Retrieved from http://www.nuance.com/dragon/index.htm

• Paradiso, J. A., Heidemann, J., & Zimmerman, T. G. (2008). Hacking is pervasive. Pervasive Computing, IEEE, 7(3), 13-15.

• PolitePix, (2015) OpenEars™: free speech recognition and speech synthesis for the iPhone Retrieved from http://www.politepix.com/openears/

• Potamianos, A., & Narayanan, S. (2003). Robust recognition of children's speech. Speech and audio processing, ieee transactions on, 11(6), 603-616.

• Protein, (2014) Dandelion (Version 2.6.0) [Mobile Application] retrieved from https://itunes.apple.com/us/app/dandelion/id521903725

• Bloomer, R. (1960) Children's Preferences and Responses as Related to Styles and Themes of Illustration

• Rannikko, P. (2011). User-centered design in agile software development. Retrieved from:http://www.statista.com/statistics/270291/popular-categories-in-the-app-store/

• Reymen, I. M. M. J., & Hammer, D. K. (2002). Structured reflection for improving design processes. In DS 30: Proceedings of DESIGN 2002, the 7th International Design Conference, Dubrovnik.

• Rubin, Z., & Kurniawan, S. (2013). Speech adventure: using speech recognition for cleft speech therapy. In Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments (p. 35). ACM.

• Saunders, E. and Stappers P. (2008) Co creation and the new landscape of design. Co Design Vol. 4, No. 1, March 2008, 5–18.

• Schaller, R, R. (1997) Moore's Law: Past, Present, and Future. IEEE Spectrum Retrieved from http://www.mae.ncsu.edu/zhu/courses/mae536/Reading/Moores_Law.pdf

• Shneiderman, B., & Ben, S. (2003). Designing the user interface. Pearson Education India.

• Schneiderman, R. (2014) Accuracy, Apps Advance Speech Recognition. IEEE Signal Processing Magazine Page 12-14, Retrieved from http://research.microsoft.com/en-US/people/deng/spm2015.pdf, 12-14

• Richards, S. (2012) The Boy and the Bears Read Aloud, (Version 1.8) [Mobile application software] Retrieved from http://itunes.apple.com.

• Shafer, G. (2003). Reflections on a democratically constructed canon. Teaching English in the Two-Year College, 31, 144-153.

•Shivakumar, P.G., Potamianos, A., Lee, S., & Narayanan, S. (2014) Improving Speech Recognition for Children using Acoustic Adaptation and Pronunciation Modeling

• Shobaki K, Hosom, J-P, & Cole, R, (2000) The OGI Kids' Speech Corpus and Recognizers, Proc. ICSLP, Beijing, China.

• Smith, S. (2004). The non-fiction reading habits of young successful boy readers: Forming connections between masculinity and reading. Literacy, 38, 10-16.

• Statista (2014a) Cumulative number of apps downloaded from the Apple App Store from July 2008 to October 2014 (in billions). Retrieved from http://www.statista.com/statistics/263794/number-of-downloads-from-the-apple-app-store/

• Statista (2014b) Most popular Apple App Store categories in March 2015, by share of available apps.

• Stine, R. L. (1992). Goosebumps [series]. Scholastic Incorporated.

• Stockar, D, V, (2006) The Importance of Literacy and Books in Children's Development. Retrieved fromhttp://www.ibby.org/718.0.html

• Strasburger, V. C., Hogan, M. J., Mulligan, D. A., Ameenuddin, N., Christakis, D. A., Cross, C., ... & Swanson, W. S. L. (2013). Children, adolescents, and the media. Pediatrics, 132(5), 958-961.

• The Mill (2014) The Fableist "Epic Thread" [Video File]. Retrieved from https://vimeo.com/106442542

• Wilson, K. (2013). A defence of story apps after a speaker at the bookseller children's book conference said that apps interfered with story Retrieved from: http://nosycrow.com/blog/not-so-appily-ever-after-disagreeing-with-a-point-of-view-expressed-at-the-bookseller-children-s-conference

• Zamani, E. D., Giaglis, G. M., & Pouloudi, A. (2013). Gravitating to" the new shiny thing": Meaning, emotions, and tablets. Technology and Society Magazine, IEEE, 32(2), 74-80.

Endnotes

[i] Society may impose gender stereotypes on children that can influence their interests. “Children are taught which colours, toys, games and books are for boys and which for girls” (National Union of Teachers, 2013, p. 3). Further, gender stereotypes are seen in children’s media with protagonists and characters aligning to traditional gender roles, supporting a gender-biased ideology. These stereotypes may have a limiting effect on children as they adopt narrow perceptions of social roles. Children may self-constrain their reading according to gender-based notions of literacy.

[ii] Significant contributions to children’s literature were developed during this period by the works of the Grimms brothers who collected and published famous children’s fairy tales such as Little Red Riding Hood, Sleeping Beauty, Rumpelstiltskin, and Hansel and Gretel. Hans Christian Anderson was another prominent author who contributed to children’s literature during this time, publishing popular stories such as The Ugly Duckling, The Little Mermaid and The Emperor’s New Clothes.

[iii] Most speech recognition systems consist of three components: an acoustical model, a phonetic dictionary and a language model (Rubin & Kurniawan, 2013). The acoustical model is created from software interpreting audio recordings with corresponding transcripts. It analyses the relationship between the audio signal and the phonemes or linguistic units that make speed and it uses software to create a statistical model of what sounds make up each word. The phonetic dictionary is a list of words with their corresponding phonetic makeup, see Table 1.

Word	Phonetic Interpretation
ELEPHANT	EH L AH F AH N T
YELLOW	Y EH L OW

A phonetic dictionary

The last component, the language model is a statistical algorithm that analyses the likelihood of the phrases being said in relation to sentence structure and language grammar. Language models also help to distinguish between similar sounding words and phrases based on contextual sentence structure.

[iv] Below is a breakdown of the media type, narrative and corresponding animation in the Lucy sequence.

Sequence Type:		Narrative:	Animation
Standby	Looped	This is Lucy	Lucy sways back and forth
Action	Single		Lucy waves her right hand
Standby	Looped	Lucy loves her hula-hoop	Lucy looks at her hula hoop
Action	Single		Lucy picks up her hula hoop
Standby	Looped	Time to Hula!	Lucy holds the hula swaying
Action	Single		Lucy uses the hula hoop