Production and Perception of Vowels in New Zealand Popular Music

Andy Gibson

A thesis submitted to
Auckland University of Technology
in fulfilment of the requirements for the degree of
Master of Philosophy (MPhil)

2010

Institute of Culture, Discourse and Communication

Primary Supervisor: Allan Bell
Second Supervisor: Jennifer Hay

Table of Contents

List of Figures

List of Tables

Attestation of Authorship

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

Signed:

Date:

Acknowledgements

Firstly, I would like to thank Dylan Storey, Andrew Keoghan, and John Guy Howell for making this project possible. You have all been incredibly generous in allowing me to scrutinise your vowels!

To Allan Bell, thank you for your ongoing faith in me, constant encouragement and challenging debate, as well as for giving me the opportunity to do the thesis. To Jen Hay, thank you for infusing research with excitement and for having an impeccable ability to ignite that passion for learning in others, including me. I would also like to thank John Bitchener for additional supervision and all of my ICDC colleagues past and present for providing such a sociable and supportive working environment.

To my parents, Tony and Heather, thank you for the ongoing encouragement and support over so many years, and in relation to so many of my pursuits. To the rest of my family, thank you for your love. Extra thanks to Karen for guiding me past several stuck points during the writing of this thesis. My biggest appreciation goes to Aaron, for being beside me throughout this journey, for insight and strength, for knowing me.

Intellectual Property Rights

This thesis contains audio recordings, and reprinted lyrics of songs by the three artists who participated in the vowel production study. The authors reserve all rights to the lyrics, music, and sound recordings of their songs. These materials should not be reproduced without permission. Written permission to include these materials in the print and online copies of this thesis was obtained from the copyright holders for all of the songs.

Ethics Approval

Ethics approval for the production and perception experiments was granted by the Auckland University of Technology Ethics Committee, for the vowel production study on 7 November 2008 (AUTEC reference number 08/218), and for the perception experiment on 26 March 2008 (AUTEC reference number 08/41).

Abstract

An acoustic comparison of sung and spoken vowels for three New Zealand singers investigates the phonetics of pronunciation in popular music. The singers recited the lyrics to their songs and recordings of their sung vocals were also obtained, creating a dataset of paired sung and recited words. Interviews with the singers were conducted so that the pronunciation used in reciting could be compared with a more conversational style. Eight vowels were analysed in these three conditions: DRESS, TRAP, THOUGHT, LOT, START, GOOSE, GOAT and PRICE. As well as providing data for phonetic analysis, the interviews elicited information about the singers' musical influences, and investigated the singers' stances towards the use of New Zealand English (NZE) in singing.

The results of the comparison of singing and speech reflect the singers' various stances to some extent. Overall, however, there are strikingly few cases where pairs of sung and spoken vowels have similar pronunciations. The predominance of 'American' vowels in the singing of all three participants, despite stated intentions to use New Zealand forms, suggests that the American-influenced singing style is the default in this context. This finding contrasts with early research on singing pronunciation in popular music, which described the use of American pronunciation in pop music as an act of identity which involved effort and awareness (Trudgill, 1983). The results presented here support the claims of more recent studies which suggest, conversely, that it is the use of non-American accent features which requires a wilful act of identity (Beal, 2009; O'Hanlon, 2006).

An important consideration in the interpretation of vowel differences between singing and speech is the role played by the act of singing itself. It has been argued that there may be a general preference for increased sonority in singing (Morrissey, 2008) which would lead to the use of more open vowel sounds. This issue is explored and some evidence is found for a sonority-related effect. However, singing inherent effects like this can only explain a portion of the variability between singing and speaking. Most of the differences between singing and speech appear to be caused by social and stylistic motivations.

To investigate why American-influenced pronunciation might be the default in the singing of pop music, a perception experiment was conducted to examine the phenomenon from the perspective of the listener. Participants were played words from a

continuum that ranged between *bed* and *bad*, and they responded by circling whichever word they heard on a response sheet. The perception of ambiguous tokens was found to differ significantly according to whether or not the words were expected to be spoken or sung. These results are discussed with reference to exemplar theories of speech perception, arguing that the differences between singing and speech arise due to context-specific activation of phonetically detailed memories. This perspective can also be applied to the processes which underlie the *production* of vowels in sung contexts. Singers draw on their memories of popular music when they sing. Their use of American pronunciation in singing is therefore the result of the fact that a majority of their memories of pop singing involve American-influenced phonetic forms.

Chapter 1 Introduction

> There can be no doubt that singers are modifying their linguistic behaviour for the purposes of singing … An interesting question, therefore, is: why do singers modify their pronunciation in this way? (Trudgill, 1983, p. 143)

The study of pronunciation in the singing of popular music has attracted an increasing amount of interest over the last thirty years. The object of analysis changes rapidly, as do the theoretical and methodological approaches applied in its investigation. The statement made in the above quote is still highly relevant, with a large proportion of the world's popular singers using different pronunciation styles in their singing than they would use in their speech. For many singers who do not *speak* English at all, it is an entirely different language which is used for the purpose of singing pop songs. The dominant variety used in this context can be labelled an 'American-influenced' style. The question in the quote above is also still relevant, and remains open for debate. This question may be bigger than it appears; in order to understand why singers modify their pronunciation for the purposes of singing, we must also at least partially understand why any person modifies their linguistic behaviour for any purpose – and that is quite a question indeed.

Thirty years after the first publication of Trudgill's study of British pop singers (as Trudgill, 1980), the perspectives he presented continue to be influential in regard to a range of issues on this topic, particularly his discussion of how a singer's pronunciation style is constructed through acts of identity (Le Page & Tabouret-Keller, 1985). However, one aspect of Trudgill's analysis has been challenged in recent studies of popular singing, and it is the central goal of this thesis to examine this issue in some detail. The debate is about whether singers are *trying* to imitate Americans or not. The way Trudgill framed this issue is quite different to more recent discussions, such as Beal's (2009) analysis of the singing style of British band, Arctic Monkeys:

1. …singers attempt to imitate what they consider to be an American accent…

   > British pop singers are attempting to modify their pronunciation in the direction of that of a particular group with which they wish to identify – from time to time (i.e. when they are singing). This group, moreover, can clearly, if somewhat loosely, be characterized as 'Americans'(Trudgill, 1983, p. 144)

2. [American features are] indexed in this context as "mainstream pop" rather than simply "American". In practice, any British artist (or indeed amateur performer) singing a mainstream pop song will employ these features as a matter of course without any conscious act of identity's taking place. (Beal, 2009, p. 229)

The difference may seem subtle, but it has large implications for the way we interpret pop singers' identity goals. Are they attempting to imitate Americans, or do the American features come about automatically, 'as a matter of course'? It is possible that each of the above positions was correct at its time of writing, since there have been innumerable changes in the world of popular music in the period between the two articles. It is not my intention to determine whether Trudgill's analysis of his data was right or wrong on this matter. Rather, I will consider the present situation in the NZ context to try and tease apart the various mechanisms which affect pronunciation in popular singing, with the goal of shedding light on the question I have framed here. That question can be stated in colloquial terms as follows: are singers 'putting on' an American accent when they sing, or do they actually have trouble 'taking it off' in the context of pop music?

This thesis examines the phonetics of vowels in singing and speech, both in the pronunciation used by singers, and in the way their listenership perceives those vowels in musical and non-musical contexts. The main hypothesis is that the American-influenced accent is so established in popular music that it is the most automatic pronunciation style to use when singing pop songs. When singers wish to project local identities, or anti-mainstream stances, they need to apply effort in order to avoid using the American-influenced norm. Several questions will be explored to examine this phenomenon:

- What are the phonetics of the differences between singing and speech?
- Do differences occur across a range of variables, or only on those which are salient?
- What effect does the act of singing, as compared to speaking, have on vowel pronunciation?
- How do singers' opinions about pronunciation in singing relate to their own singing style?
- What social meanings might be invoked by singing accents?
- How do memories of previously heard popular songs affect pronunciation in singing?

The contents of this thesis will be outlined below, showing how these questions will be examined through the presentation of two empirical studies. Before presenting this

outline, however, it is necessary at this early stage for me to position myself in relation to the research project and acknowledge my subjectivity on this topic.

1.1. Contextualising the Researcher

This project was motivated by my own experience as a New Zealand singer/songwriter. On the positive side, this means I bring some insights and intuitions which a non-musician or non-local researcher might not have had in approaching the subject. On the downside, it means I may have ideas about singing accents which feel like common-sense to me, but which have actually arisen out of idiosyncratic aspects of my experience. To provide a context for the reader, I will briefly describe my experience of pronunciation in singing.

I grew up listening largely to music of American, British and New Zealand origin. As a child, I sang along to songs, transcribed their lyrics, and presumably emulated the singing accents used in the recordings. On reflection, the majority of the pronunciation styles on these recordings conformed largely to American norms, with this trend including singers from New Zealand. At an early age, I began writing my own songs.

As an 18 year-old (notably coinciding with my first exposure to sociolinguistics), I began to notice that I used American-sounding features when singing my own original songs. I was unhappy with this because I felt that it was inauthentic for me to sing personal songs in a voice which was dialectally distinct from my speaking voice. As a result I began a long, and to some extent ongoing, process of retraining my singing voice to be more similar to my speaking voice. Listening back over the years of recordings, this was a very gradual process which came and went according to the recording context, song style, and place of residence; but gradually I developed a relatively New Zealand English (NZE) sounding singing accent, and this has been noted (not always favourably) in reviews of my music.

My experience was that there were certain moments, points of decision, at which a specific word in a song lyric would bring pronunciation options to my attention. This began with a very marked choice between /a/ and /æ/ in words like *can't*. As my training in phonetics went on, I began to notice more subtle features, and found myself faced with decisions like 'should I use lip-rounding on the word *gone*' or 'should I raise the vowel in *edge* to match my spoken pronunciation'. A kind of conscious mapping process seemed necessary – close attention to speech and then approximation of the spoken phonetics in a sung form.

The combination of sociolinguistics and singing has therefore led me through what is perhaps an unusually exaggerated process of reflexivity. It would be easy to extrapolate out from my very individual experience to form conclusions about other New Zealand (NZ) singers, and this is a danger which I have tried to be aware of in designing this project. It is therefore important that I prioritise the data collected over my own intuitions, though these will no doubt have an effect on my interpretation of that data.

1.2. Thesis Outline

The examination of pronunciation styles in singing begins, in chapter 2, with a review of literature related to the sociolinguistics of style, discussing the emergence of trends which focus on the way speakers use language to produce and reproduce social meaning. I will briefly touch on studies of vocal technique before moving on to look in detail at studies of pronunciation in popular singing, which will be presented in a loosely chronological fashion, with the studies also broken up by region, moving from British, to Australian, and then New Zealand singers. These studies provide valuable data analyses and insightful explanations; they are also limited in several ways. These limitations will be discussed in order to motivate the methodology of the first of the two empirical studies presented in this thesis, a vowel production study comparing the phonetics of sung and spoken vowels.

Chapters 3 to 5 present the vowel production study, a comparison of vowels from the singing and reciting of original songs by three New Zealand singers. This comparison extends existing research in three ways. Firstly, it uses sung vocals extracted from their recorded context to allow for acoustic analysis. Secondly, it compares sung and recited versions of the songs, using matched environments to provide direct comparisons, isolating the effect of singing from other linguistic influences. Thirdly, the analysis includes several vowels which have not been included in prior studies, notably I have not focused solely on variables which are stereotypically associated with American or British English. The design and methodology of the vowel production study is described in chapter 3, followed in chapter 4 by the presentation of the results. The results are then discussed in chapter 5.

While the production study describes singers' usage of different vowel qualities to construct context appropriate styles, chapter 6 takes a different perspective. Following recent trends in sociophonetics, differences between singing and speech are examined from the perspective of the listener. To do this, I present a small study which examines

the way ordinary non-singers perceive vowels in musical and non-musical contexts. The experiment investigates how participants categorise an ambiguous word according to whether it is expected to be sung or spoken. To contextualise this perception experiment, chapter 6 also provides an introduction to theories of speech perception, focusing particularly on *exemplar* approaches, which argue that phonetic details are stored as episodic memories, indexed with social and contextual associations.

The final chapter revisits the production results in light of the perspectives from the speech perception experiment, drawing together the results of the two empirical studies to form conclusions about the processes involved in the production and perception of vowels in singing.

Chapter 2 Popular Singing as Stylistic Variation

In this chapter, a range of literature relevant to the study of pronunciation in singing will be reviewed. The observation that non-American singers tend to use American-influenced pronunciation when they sing can be considered an example of stylistic variation. The chapter will therefore begin by reviewing sociolinguistic approaches to style in section 2.1. Section 2.2 briefly introduces some aspects of vocal technique and considers physiological and acoustic differences between singing and speech. A longer section (2.3) follows, which is devoted to the studies that have analysed pronunciation styles in popular music singing. This section will raise issues directly pertinent to the empirical studies which follow in subsequent chapters. It should be noted here that this chapter does not review all of the literature to be discussed in this thesis. A review of the literature on speech perception is reserved until chapter 6, at which point it will be presented in order to contextualise the perception experiment described in the second half of that chapter.

At this point, a few caveats need to be made regarding the scope of topics covered below. There are several important areas of research that would provide useful perspectives which have been left out to keep the scale of the thesis at an appropriate level. Foremost among these is the work in popular music studies which considers the effect of market forces and the processes of globalisation on the production and consumption of popular music. Some of this work, particularly that which investigates hip-hop, has dealt directly with sociolinguistic questions (e.g. Pennycook, 2003). While I acknowledge its relevance and its importance, this literature is not reviewed below.

Similar limitations of scope are applied to insights from social theory. For example, Giddens' theory of structuration (Giddens, 1984) and his view of the construction of self in late-modernity (Giddens, 1991) could provide useful tools for a discussion of pronunciation style in singing, by highlighting the tension between structures of established norms and speaker agentivity. Many themes from social theory have worked their way into the sociolinguistic literature and will be raised in that context, but a review of original texts on these issues will not be presented.

While addressing limitations relating to scope, a few points on my use of terminology should be made. One of a number of simplifications made is that singing and speaking are referred to more dichotomously than is desirable. Singing and speech are not two

categorically distinct modes; a range of more or less sung and more or less spoken activities exist. Liturgical chanting, horse-race commentary (Kuiper & Austin, 1990), and rap could all be considered to blur any clear division line between singing and speech. Another terminological simplification is that the term 'pop music' is used in a very broad sense to refer to a wide range of non-classical singing styles, not solely to the more specific meaning of pop as a commercially-oriented music genre. My description of pop music in this way therefore glosses over the massive diversity of music styles which exists. A related point is that I will refer to 'American accented' singing as being the norm. It should be emphasised that it is not the only model available, different music styles are associated with different pronunciation styles, and a wide range of variability does exist in the pronunciation of pop singing.

2.1. Style in Sociolinguistics

This section describes a range of approaches that have been taken to stylistic variation in sociolinguistics beginning with the *ethnography of speaking* and moving on to variationist approaches including *audience design*, and including the influence of social psychology in the form of *accommodation theory*. I will then signal the turn towards social constructionist approaches to language variation, with their emphasis on language use as construction of social meaning. The extent to which each of these theories is relevant to the study of singing pronunciation will be considered at several points as the section progresses.

2.1.1. The contexts of language use

Early sociolinguistics emerged as a reaction to the dominance of generative linguistics, which ignored issues of identity and variability in speech, focusing on the rules of language as abstracted away from its contextualised usage. The reaction to the Chomskyan paradigm was to focus on speech as it is used by real people in real contexts. Three perspectives on the study of situated language use are presented in this section, chosen on the grounds that they provide useful insights for the present thesis. My aim is not to give a full overview of this literature, but rather to briefly introduce a few relevant points from each perspective, and to acknowledge the influential role that this early work has played in the development of sociolinguistic theory.

Hymes: Ethnography of speaking

Hymes' (1968) *ethnography of speaking* paradigm was central to the development of the study of language in situated contexts. Being positioned as much in the realm of anthropology as linguistics, ethnography of speaking broke down *speech events* into contextual factors including the participants (the sender and the receiver), the setting or situation, the topic, and the channel. The channel of a speech event refers to the distinction, for example, between speaking and writing. Singing and speaking can also be said to use different channels. The factors of speech events are also considered in terms of their motives or purpose, having related functions, such as expressive, referential, metalinguistic, or poetic.

As well as outlining a method for describing the components of speech events, Hymes also discussed the importance of *linguistic routines*, which are "recurrent sequences of verbal behaviour, whether conventional or idiosyncratic" and which make up a "vast portion of verbal behaviour" (Hymes, 1968, p. 126). Linguistic routines can become automatic, being produced without special attention or effort; they could be described as being carried out with what Giddens (1984, p. xxiii) calls *practical consciousness*, as opposed to *discursive consciousness*. Practical consciousness "consists of all the things which actors know tacitly about how to 'go on' in the contexts of social life without being able to give them direct discursive expression".

When explaining the pronunciation differences that occur between singing and speech, the notion of linguistic routines will be useful, and it will emerge in various guises throughout this thesis, for example in section 2.3.1, where singing accents are described as being determined by *vocal habits*.[1]

Halliday: Functions of language use

Another branch of research which emphasised the importance of language as it is used appeared in the influential work of Halliday (e.g. 1978), who emphasised the fact that language is used in order to carry out functions. He introduced a useful distinction between *dialect*, speech variation which relates to the identity of language users, and *register*, variation which relates to different uses of language, which vary according to

[1] I do not use Bourdieu's (1991) term *habitus* because it would require a more thorough treatment of issues in social theory, however my use of the term habit is in some sense a reference to Bourdieu's discussion of the power of social structure as manifested in the individual.

context. The distinction between registers can be related to differing *modes of discourse*. "The question underlying the concept of the mode of discourse is, what function is language being used for, what is its specific role in the goings-on to which it is contributing?" (Halliday, 1978, p. 223).

Language as used in pop singing and language as used in face-to-face talk are different modes of discourse. Language serves a different set of functions in these two modes. As Trudgill (1983, p. 159) states:

> Pop music is a field where language is especially socially symbolic, and typically low in communicative function, high on the phatic and self-expressive.

Some other functional differences between singing and speaking are described by Rampton (2006, p. 110). In singing, the grammar and semantics are fixed and determined prior to performance, whereas in talk, language is "flexibly formulated … to refer to some feature of the situation at hand". Singing also sets up very different participant roles, it "neither demands a response from co-present individual(s), nor positions them as addressees".

If singing and speech are different modes of discourse, then the varieties of language used in singing and speech could be described as different registers. This provides a way to quickly distinguish between sung and spoken language varieties; however, it does not conform well to social constructionist principles, which describe variation in terms of language processes rather than language objects (this perspective will be introduced below in section 2.1.3). While register may be an overly static notion, Halliday's focus on the functions of language is important, and should be kept in mind as we consider the factors influencing popular singing pronunciation.

Goffman: Context as a frame

Another important concept which deals with the contexts of language use is Goffman's (1974) notion of *framing*, which describes the way people can emphasise or de-emphasise different aspects of their identity according to the localised needs of a given context. Frame analysis has at its heart the idea that a person's experience is organised around the perpetual question "'What is it that's going on here?'" (Goffman, 1974, p. 8). This organising of experience sets up frames of reference which determine the shades of meaning ascribed to various aspects of a given scene. Most relevant to the discussion of singing accents is *generic framing*, which sets "meaning parameters

around talk in relation to what contextual type or genre of talk … is understood by participants to be currently on-going and relevant" (Coupland, 2007, p. 113).

Goffman discusses pop singers as an example of the way different elements of self are projected in different frames. Aspects of a singer's identity, such as their nationality, may become less relevant when they take on the role of pop singer, which is a kind of "not-self" (Goffman, 1974, p. 535), because the singer is taking on a persona which is unique to this frame. The projection of this persona leads to a "dissociation … between the figure that is projected and the human engine which animates it" (p. 573). However, Goffman also argues that audiences are attracted to singers who can reduce this dissociation, representing what seems to be a sincere representation of the singer's 'true self'. Compared to the ethnography of speaking and functional approaches which describe the contexts of language, framing describes the way that context brings forth different meanings.

Summary: The contexts of singing and speaking

There are a range of grounds upon which language use in singing and speech can be said to differ. An emphasis on context provides several ways to think about these differences:

- Singing and speech occur in characteristically different speech events that are associated with linguistic routines and habits;
- They are different modes of discourse, using language to achieve a different set of functions in each mode;
- They occur in different frames, in each of which the organisation of meaning differs, resulting in the highlighting of different aspects of identity.

The approaches outlined in this section share several qualities. They emphasise speech *as it is used*, rather than in an abstracted form. Speech is seen as embedded in a rich contextual space which determines the meaning of an utterance. This means there is plenty of scope for theorising style. What the above research paradigms lack, however, is in-depth analysis of the language itself, relying more on illustration than on rigorous description of the ways in which language can vary. This can be contrasted with the development of variationist sociolinguistics, initiated by Labov, which focused on the orderly quantitative patterns of variation which exist for specific linguistic variables.

2.1.2. Style in variationist sociolinguistics

Labov: Style as attention to speech

Labov's pioneering work in variationist sociolinguistics focused mainly on variation *between* speakers, as a way to gain insights into sound changes. He also documented stylistic variation in the speakers he studied, though this variation was not a central concern of the research paradigm. As variationist sociolinguistics has developed, however, stylistic variation has been increasingly recognised as an important phenomenon deserving attention in its own right.

Labov's study of the stratification of speech patterns in New York City used sophisticated methodologies to show that speech varies according to social categories such as gender, age, ethnicity and social class. Despite viewing social forces as affecting language "in the living present" (Labov, 1972, p. 3), most of Labov's early work focused on large scale patterns of language use by people grouped according to static social categories.[2]

As a part of this research, Labov developed an approach to style shifting, finding that not only does language vary across social groups, but also within individuals according to the various tasks involved in a sociolinguistic interview. In general, more formal tasks elicited more standard speech. Labov argued that it is the degree of attention paid to speech which drives stylistic variation, with speakers being more self-conscious when reading, for example, than when story-telling, and trying to present a higher class persona in these tasks.

Labov's 'attention to speech' explanation for style shifting can be applied to New Zealanders singing pop music, though it takes a rather different form. I will argue that it is the American singing accent which is most automatic and requires least attention. Instead of the sort of class anxiety said to account for style shifting towards a standard accent, New Zealand singers may experience 'authenticity anxiety', and may shift towards more NZ-like variants when attention paid to singing is high. This is similar to the scenario described by Prince (1987, 1988) in her discussion of an Israeli singer's

[2] This was not the case however in Labov's earliest major study, on the island of Martha's Vineyard (Labov, 1972), in which speakers varied most dramatically when categorised according to a *stance* rather than a static identity trait. Those who had a positive stance to the island and planned to stay there used more of the islands' local variants than those who were more mainland oriented, and had plans to leave the island.

efforts to use a local variety of Yiddish in her singing, which will be discussed in section 2.3.4.

Labov's work has been criticised by several authors (e.g., Bell, 1984; Coupland, 1980) for being uni-dimensional in its conception of stylistic variation, with the causality of style shifting being explained only in terms of attention paid to speech. The first major alternative to Labov's approach to style came in the form of the *audience design* framework (Bell, 1984) which claims that people style-shift according to who they are speaking to, rather than how much attention they are paying to their speech.

Bell: Audience design – style orients to people

Audience design has become a highly influential framework for dealing with style. It argues that speakers modify their language primarily in response to their audience. The main theoretical point is that style orients to *people* rather than to functions or mechanisms. The theory claims that stylistic variation is derived from variation which already exists as social variation, and that style shifts derive meaning from the association of a variety to the social group which uses it. Bell (1984) studied a range of sociolinguistic variables in the speech of radio news broadcasters. He analysed the speech of individuals reading the same news, in the same studio, on two different New Zealand radio channels, which had different audiences. He found that the newsreaders used more standard forms on the station which broadcast to a higher class national audience and more colloquial variants on the community station. He argued that these changes were due to the differing audience, not attention to speech, because everything about the context was the same in the two conditions except for the audience.

The framework distinguishes between *audience design* and *referee design*. In referee design, the speaker shifts their speech so that it is more similar to some group of people the speaker wishes to identify with. Another important distinction set out in the framework distinguishes *responsive* shifts and *initiative* shifts. In a responsive shift, the speaker is shifting their language in response to a situation. These shifts reinforce conventional ways of speaking, and are more likely to involve audience design than referee design. In initiative shifts, on the other hand, the speaker shifts their language in order to change the situation, which can be achieved through referee design. In Bell (1984), referee design was seen as secondary to audience design, however in a revision of the theory, the responsive (audience design) and the initiative (referee design) are seen as "two complementary and coexistent dimensions of style, which operate simultaneously in all speech events" (Bell, 2001, p. 165). At the same time that style

was getting more attention in sociolinguistics, Giles and colleagues (e.g. Giles & Powesland, 1975) were exploring stylistic variation from the perspective of social psychology.

Giles: Accommodation theory – identity motivations

The approach to style taken by Giles and colleagues (e.g. Giles, Coupland, & Coupland, 1991; Giles & Powesland, 1975), *communication accommodation theory* (initially *speech accommodation theory*), described style in terms of identity motivations. Speakers are seen as negotiating their identity on an ongoing basis by aligning themselves with groups they esteem, and distancing themselves from groups they wish to be distinct from. Accommodation theory proposed that speakers can do this by either *converging* with or *diverging* from the speech style of the person with whom they are speaking. In a linguistic sense, this means that their speech becomes more similar to or more distinct from the speech of their interlocutor, while in social terms this reducing or increasing of distance actually represents a desire to reduce or increase *social distance*. By accommodating to one's speech partner, speakers attempt to make themselves more desirable or likeable. These ideas are not dissimilar to those developed in audience design. Accommodation theory emphasises interaction, focusing on processes rather than states, actions more than objects. Language is viewed as one of the practices through which people can enact their identity, and negotiate boundaries.

The idea that language is used for the purposes of identity construction is central to Le Page and Tabouret-Keller's *acts of identity* framework, which states that "the individual creates for himself the patterns of his linguistic behaviour so as to resemble those of the group or groups with which from time to time he wishes to be identified, or so as to be unlike those from whom he wishes to be distinguished" (Le Page & Tabouret-Keller, 1985, p. 181). Le Page's work will be discussed in more detail in section 2.3.1, where it is invoked by Trudgill (1983) as the best way to account for the use of American pronunciation in popular singing. The acts of identity framework emphasised the speaker's agency in using language as a resource for meaning making in situated contexts. This perspective resonates with social constructionist approaches to style which have become increasingly prominent in recent sociolinguistics.

2.1.3. Speakers as social actors, creating style

Eckert (e.g. 2000) extended variationist approaches to language style, through ethnographic study of communities of practice (Wenger, 1998). While maintaining the

phonetic detail and the quantitative focus of earlier variationism, Eckert's work avoids the reduction of a speaker's social identity to age, gender, social class and ethnicity. The parameters of social identity considered analytically relevant are those that are relevant to the speakers themselves. Eckert appeals to the idea that language is used in an active way to construct identity from moment to moment, from context to context. Speakers can become aware of the meaning attached to a given form and then use it for effect. In this way, different variants come to be related to different *stances*, which relate to either momentary or enduring aspects of a speaker's identity, often in relation to certain attitudes. The ethnographic approach allows a multi-dimensional view of the meaning-making enacted through stylistic variation. This work is in line with the principles of social constructionism, which are also argued for by Coupland.

Coupland (e.g. 2007) emphasises the creative role of speakers in style, that they do not merely respond to the situation, but that they are constantly shaping and creating the situation through strategic use of language style. People do *identity work* using language to create and recreate their multiple identities. Identity is not seen as a person's belonging to certain social categories such as sex and ethnicity, but as being a complex combination of identity traits which are differently enacted at different times. Irvine (e.g. 2001) developed the notion of identity work, arguing that stylistic variation relates to speakers' construction of distinctiveness through the differentiating of language style from outgroups.

These perspectives view identity as a process, not an entity. Even responsive style accomplishes identity goals; it may be viewed as an attempt to fit in and is described as 'relational' by Coupland, whereas initiative style is more often a case of negotiating identity at the boundaries between self and other. Thus, Coupland's approach to style is oriented more to the speaker than to the audience. Schilling-Estes (2004, cited in Coupland, 2007, p. 80) labels the approach as speaker design, contrasting it with Bell's audience design. Dynamic views of identity construction bring with them an emphasis on performativity in language. Coupland refers to *stylisation* as the knowing and artful display of identities, which may include the enacting of personae not normally associated with the speaker.

In an earlier study, Coupland (1985) analysed the speech of a Cardiff-based radio host, highlighting the performative nature of stylistic variation, and the way important identity work can be done using single tokens of salient variables. The DJ was found to use what Coupland calls 'phono-opportunities' to present phonetic variants that were

rich in social meaning, presenting a local Welsh identity. The title of his show 'Hark, hark, the lark' provided opportunities to use the highly salient fronted START vowel.[3] Coupland finds many instances of performative language in the show, where the announcer is essentially 'putting on' his own local accent for effect, a case of what could be described as *ingroup referee design* (Bell, 1984).

While the above study focused on the styling of the speaker's local identity, a similar theoretical perspective can be found in Rampton's work on *crossing* and *styling the other* (Rampton, 1995, 1999), which focuses on instances where people take on a speech style that is not ordinarily associated with their own persona. The idea of crossing is akin to a case of initiative referee design, in Bell's terms, but focuses particularly on the way language can be used to subvert convention. Rampton's work takes a social constructionist stance, "focussing on a range of ways in which people use language and dialect in discursive practice to appropriate, explore, reproduce or challenge influential images and stereotypes" (Rampton, 1999, p. 421).

Language forms can be said to index social meanings through their association with certain perceived categories of speaker. The ideas of indexicality are rooted in the semiotic theory of Peirce, which classifies the various ways in which a signifier can refer to its referent (Peirce, Hartshorne, & Weiss, 1935). An *index* does not resemble the referent, but refers to it through association – smoke is an index of fire, for example. Eckert (2008, p. 453) argues that "the meanings of variables are not precise or fixed but rather constitute a field of potential meanings – an *indexical field"*. This is a promising approach because it acknowledges inherently the multi-dimensional and dynamic properties of language style.

The most fully realised theory of indexicality in sociolinguistics is expounded in Silverstein (2003), though it is more accessible as applied elsewhere (for example, Johnstone & Kiesling, 2008; also Ochs, 1992, provides a useful treatment of indexicality in language). One interesting aspect of Silverstein's theory, relating to the distinction between responsive and initiative style, is the description of how the use of a certain variant can be both "appropriate-to" the situated context, responding to it, and "effective-in" the context, initiating new meanings. The context includes metapragmatic awareness, which links ideologies about people to linguistic variants.

[3] Throughout this thesis, vowels are identified using the labels for lexical sets outlined by Wells (1982).

A useful way to theorise language styles which have attracted a high level of awareness is Agha's (2003, p. 231) *enregisterment*, the "processes through which a linguistic repertoire becomes differentiable within a language as a socially recognized register of forms". The American accent can be viewed as holding this sort of status in popular music, as an enregistered symbol of the activity of singing pop. One of the strengths of Agha's account of enregistered voices, as discussed by Eckert, is that registers are seen as being "in a continual process of production and reproduction" rather than as "a static collocation of features associated with a specific setting" (Eckert, 2008, p. 456) – while a register is an object, enregisterment is a process.

Social constructionist approaches to language are highly relevant to the interests of this thesis, since popular music is a site of overt identity construction. Recordings of popular music are rich in stylised semiotic material, carefully honed to present artists in a certain way. They can therefore act as sites of quite conscious levels of performativity. With their emphasis on the agency of the speaker, social constructionist approaches make examples of performative language highly relevant to the study of style.

In the broadest sense, all language is performed. Goffman (1959, p. 15) defined a performance as "all the activity of a given participant on a given occasion which serves to influence in any way any of the other participants". A narrower definition, though, is to treat performance as that which is in some sense *staged*, that is, occurring in a clearly defined and temporally bounded context, with clearly distinguished performer and audience roles.

Coupland uses the term *high performance* to separate this narrower definition from *mundane performance*, while in Gibson & Bell (fc) we have used the term *staged performance*. This type of overt performance is a fruitful site for studying the knowing use of language to play with and subvert existing indexical relationships. In our study of an animated comedy (Gibson & Bell, fc), for example, there are highly stylised portrayals of enregistered varieties. These performances vary in their footing from earnest autobiographical performances to caricature, but the construction of voices is always knowingly loaded with social meaning.

2.1.4. The common ground between theories of style

Theories of style are relevant to the study of pronunciation in singing because they attempt to account for context-dependent changes in individuals' speech patterns. Individuals create very different styles in speech and singing, and the social meanings

of those styles are contingent on the frame of the speech (or singing) event. As we move on to consider issues more directly related to music in the following sections, the perspectives from sociolinguistics should be kept in mind.

All of the perspectives on style that have been discussed above have in common the theme that people are able to change their language from one situation to the next. The different approaches disagree on the way we should describe and explain this variation, and which factors are primary in stylistic changes. Anthropological approaches focus on having a detailed understanding of the context of speech events. Social psychology focuses on speakers' motivations in regard to identity goals. Variationist approaches use quantitative methods to establish the big picture of variation within a structuralist perspective. Social constructionist approaches argue that style is not merely a response to external forces, but an attempt to change the situation at hand.

Methodological changes have involved a movement from categories established by the analyst, to categories which emerge from the perspectives of the speakers under analysis, to studies which analyse the social meaning of variables in richly contextualised accounts of individuals. Another change in methodology, therefore, has been the move from the sole use of large scale quantitative studies to the inclusion of individual tokens as informative sites of social meaning (Bell, 1999; Coupland, 1985; Schilling-Estes, 1998).

While the situated meaning of performative tokens is lost in quantitative analyses, there is also a danger in the qualitative approach. The analyst's interpretation of the social meaning of individual tokens in their context of use is difficult to verify. Qualitative studies often lack a systematic way to show that the interpretations of meaning are not caused by analytical bias and that other analysts would be likely to arrive at similar conclusions. These obstacles can be overcome by using a mixture of quantitative and qualitative methods whereby the more general correlations between language and social categories provide the justification for the interpretations of situated meaning (see, e.g. Schilling-Estes, 2004).

Another methodological difference between approaches is the way in which context is described. The ethnography of speaking approach uses a checklist of contextual factors; Labov appeals to attention to speech; Bell to the audience; Giles, Bell, Le Page and subsequently Coupland all highlighted the importance of internal identity motivations. This range of agendas serves to highlight the massive scope of the phenomenon that is

style, and also shows how difficult it is to pin it down with a singular, cohesive framework.

In the next section I focus on research which is more about music than language, and in the subsequent section I will turn to the body of research which is most directly relevant to this study – the intersection of music and language that occurs when people sing. The study of that topic will draw upon and weave together the information laid out in this first half of the chapter, including the following insights from studies of voice.

2.2. Vocal Style and Vocal Technique

This section begins with a review of Potter's treatment of the development of vocal technique in the 20th century, which provides a bridge from the social focus above to the more technical focus of the second part of this section which deals with the specificities of voice in its embodied form.

2.2.1. A history of vocal style

In the book *Vocal Authority*, Potter (1998) presents a comprehensive overview of the historical development of singing styles from early classical music through to 20th century popular music. He contrasts classical singing with popular singing on several dimensions. In the first half of the 20th Century, jazz, blues, and then rock music brought with them "a return to singing as carrier of text, a vehicle for the articulation of meanings" (p. 189). This change contrasts with classical singing, where the text is often subservient to vocal virtuosity, the most obvious example being the use of languages which neither singer nor audience understand.

Developments in pop singing styles are described by Potter with reference to three influential artists. Firstly, Frank Sinatra represents a transition point between classical singing and later pop singing, using a vocal style that included attributes of classical singing technique such as a stable, lowered larynx position. Elvis Presley, in the 1950s, marked a move towards a singing style more similar in technique to speaking than earlier styles had been. The next development in vocal technique was influenced by a shift in the role pop singers were expected to play in singing. From the early 1960s, coinciding with the increasing popularity of the Beatles, singers began to be seen as *creators* of song, rather than *interpreters* of song. This orientation already existed, for example in blues, but Potter argues that the Beatles were the first major pop group to

emphasise themselves as writers in this way. This shift in perspective had the result of further decreasing the distance between singing and speaking styles.

A primary reason for the development of the 'speech-related technique', which placed emphasis on both the meaning of words, and on the identity of the singer, was the use of microphones. Rather than having to project their voices over an accompanying orchestra, singers were able to use higher larynx positions without vocal damage, and use voice qualities such as breathy voice and creaky voice that would not be used often in classical singing. These changes in technique lead to a singing style which is somewhat closer to speech than classical singing.

This comparison of technique between classical styles and pop music raises the question more generally about the notion of distance between singing and speaking. There are vocal styles in music which are much closer to speech than is pop singing, the obvious example being rap. Potter claims that rap is not unique in its blurring of the line between singing and speaking. For example, early baroque recitatives stylised the pitch contours of speech to produce music directly from text. The important point to take from this is that singing and speech are not two distinct categories, but positions along a continuum of vocal styles. There are singy styles of speech, such as the sermonising of an Evangelical minister, and there are speaky styles of singing, like rap. There are also styles which blur the lines to the extent where it is difficult to even categorise the vocal form as one mode or the other, as in some chanting traditions.

These differences are highly relevant to any theory of pronunciation in singing, because the more singing and speech diverge from each other in their overall spectral qualities, the less likely vowel variants may be to carry the same social meaning in singing as they would in speech. This claim will be raised again in the discussion of the results of the comparison of singing and speech, and will be further developed in chapter 6, where I will turn to theories of speech perception and representation. For now, let us look more closely at the claim that popular singing is closer to speech than is classical singing. This has been examined empirically in relation to many aspects of technique through the work of Sundberg and colleagues.

2.2.2. The acoustics of singing: Insights from voice studies

In a study of pronunciation in singing, it is important to recognise the possible effect of factors which are unique to singing, alongside the sociolinguistic and psycholinguistic influences which would normally be considered in sociolinguistic studies. This is

necessary because it would be a mistake to assume that the common sense of sociolinguistics in regard to 'linguistic factors' will extend appropriately to singing or be sufficient. I will call these sorts of factors *singing inherent effects*. For example, Wray (1999), in examining the way a choir learns to pronounce an unfamiliar vowel, argues that singers need to be able to produce a given phoneme with a wide range of phonetic realisations in order to achieve maximum resonance and maintain healthy production. The demands of singing may lead to much greater variability in formant values than is required in speech. The most thorough treatment of the kinds of factors which may be relevant to this study is presented in Sundberg's (1987) book, *The Science of Singing*, along with many of his previous and subsequent empirical studies.

This research paradigm has focused on comparisons between singing and speaking, and between singing in different styles. In a series of studies on country singers, several facets of the singers' voices were found to be similar between singing and speaking, such as the overall resonant structure of the voice in terms of long-term-average spectra (Cleveland, Sundberg, & Stone, 2001), subglottal pressure (Cleveland, Stone, Sundberg, & Iwarsson, 1997) and respiratory function (Hoit, Jenks, Watson, & Cleveland, 1996). On the basis of the similarities found in this series of studies, country singing was contrasted with operatic singing. Opera singing was found to be different from speech for all of the variables studied.

Stone, Cleveland, Sundberg, & Prokop (2003) had a classically-trained professional female singer perform the American national anthem in both an operatic and a Broadway style. There were large differences between the styles, with the opera style tending to have both lower F1 and F2 than the Broadway style. This wholesale difference in formant values was not vowel specific, suggesting that it was more about the overall configuration of the vocal tract than an effect of vowel placement. In particular, classical singing is associated with a lower and more fixed larynx position. "Larynx height variation is a feature of untrained singing and singing in different styles" (Howard, 2009, p. 155). While larynx variation depends on pitch for untrained singers, changes in pitch have other effects too, for both trained and untrained alike. Higher pitches are associated with lowering of tongue dorsum, decreasing of lip-rounding and increased jaw opening, particularly in cases where the main acoustic correlate of pitch (f0) approaches the first formant frequency (F1) of a vowel. All of these articulatory changes have the effect of raising F1 (Sundberg, 2009).

Several studies have illustrated this tendency for singers to open the jaw more as they sing higher in order to keep the F1 above the pitch (Austin, 2007; Sundberg & Skoog, 1997). This increases the amplitude of the note without increasing vocal effort. Austin found that novice singers were similar to trained singers in this regard and suggests that "the most obvious explanation is that the need to open the mouth more for high notes is somehow "natural" and does not have to be learned or taught" (Austin, 2007, p. 78). If this is the case, then we expect higher notes to have more open vowels. This has important implications in a sociolinguistic study of singing as these singing inherent constraints are based not only on linguistic factors but also on musical and physiological factors.

Other studies in voice research have looked at long term averaged formant values of singers of pop (Borch & Sundberg, 1997) and country (Burns, 1986) in comparison to opera singers. They found evidence that opera singers have a band of high amplitude energy around 3kHz, referred to as the *singer's formant*, while the non-operatic singers do not. The singer's formant makes an operatic voice audible over an orchestra. This is not required for a singer of popular music with a microphone and amplification.

The relevance of these studies is that they demonstrate many kinds of acoustic variability which are not sociolinguistically motivated, with differences in F1 and F2 arising from vocal tract settings associated with different singing techniques. In the comparison of singing and speech in this thesis, then, care should be taken in construing sociolinguistic meaning from vowel formant differences between the two modes, particularly in cases where these differences occur evenly across several vowels.

One study from the researchers discussed in this section deserves to be considered in some detail, as it is the one most directly relevant to the topic of this thesis. Stone, Cleveland and Sundberg (1999) performed an acoustic analysis of speech and singing by five professional country singers from the south of the USA. The participants sang the American national anthem and a song of their choice, and then recited the words to both these songs. It was found that the formant values, on a vowel by vowel basis, were not significantly different between their speech and singing. This study differs from those discussed above in that it deals with vowels individually rather than looking at long-term averages of formant values. The methodology used by Stone at al. (1999) is the only example I have found in the literature which performs a direct acoustic comparison of the vowels in sung and recited versions of a text. This method will be adapted in the present thesis to compare singing and speech in the New Zealand setting.

In terms of identity processes, the result of Stone et al.'s research is especially interesting. It contrasts with almost all of the examinations of singing pronunciation discussed below in 2.3, which focus on artists outside of America who have different pronunciation patterns in singing than they would in their speech. If the pronunciation model for country music is a southern American accent, then there is no conflict between personal identity and genre appropriateness for Stone et al.'s singers. That is, they already 'own' the normative accent for this singing style and there is no motivation for them to adjust their accent when singing. For a country singer anywhere else in the world, on the other hand, a conflict between their spoken voice and the normative country voice is likely to exist. There is a tension between their local identity, and the global cultural phenomenon that is country music.

2.3. Sociolinguistics of Singing

In this section, I review previous sociolinguistic studies of pronunciation in popular music singing. Sections 2.3.1 - 2.3.3 cover this research by region, beginning with the pioneering study by Trudgill on British singers then moving on to studies in the Australian and New Zealand contexts. In section 2.3.4, I will consider two studies which looked at singing in Yiddish and Hebrew, and which both discuss their results with reference to salience. Section 2.4 will then critically overview the research on singing pronunciation, and will motivate the methodology for a comparison of singing and speech.

2.3.1. Pronunciation in British popular music

Trudgill 1983: Acts of conflicting identity

The first sociolinguistic analysis of popular music singing pronunciation was Trudgill's (1983, originally published in 1980) study of British pop/rock[4] and punk bands in the 1960s and 1970s. It was found that the British singers used a large proportion of 'American' variants in their singing. This proportion, however, declined after about 1964 for bands such as the Beatles and the Rolling Stones. In the late 1970s, punk bands emerged, and introduced features associated with urban, working-class, British youth.

[4] Throughout this thesis, music genre labels should not be taken as absolute. The classification of music styles is a process which changes quickly; existing labels shift their meanings, and new labels are constantly invented.

Performers in this genre used these overt markers of local identity alongside a number of the more established American features.

Bell's original conception of audience and referee design (1984) discussed Trudgill's (1983) study, claiming that both the British singers and their audiences shared 'Americans' as an outgroup referee to be referenced in popular singing. He stated that the return to local pronunciation in punk music was a "peculiarly inverted initiative design", whereby the singers were breaking with an "institutionalised" form of referee design, back to audience design (Bell, 1984:195).

Trudgill interprets the use of American features in pop singing with reference to the theory of linguistic modification developed by Le Page (see Le Page & Tabouret-Keller, 1985). As mentioned earlier, the acts of identity framework states that speakers' linguistic behaviour is modified to resemble groups with which the speaker wishes to identify. These modifications are constrained by i) the speaker's ability to identify the model group, ii) the speaker's access to the speech of the model group, and enough analytical ability to work out the rules of their speech behaviour, iii) the strength of motivation to identify with a number of different groups which may have conflicting speech models, and iv) the speaker's ability to modify speech behaviour, which lessens with age.

Trudgill argues that in the case of British singing pronunciation, the first two of these constraints play an active role in singing pronunciation. Firstly, British singers identify 'Americans' as their model, thus attempting to create a single linguistic model from what is, in reality, a highly variable set of linguistic codes. Trudgill argues that even if the second constraint (adequate exposure to the model group, and ability to analyse the accent) was met, this attempt to homogenize American Englishes into a single accent would be highly variable and inconsistent. He goes on to argue that this is indeed the case in British singing. Using the example of non-prevocalic /r/, he states that British singers are actually quite unsuccessful in accurate reproduction of American English singing accents, citing instances of 'hyper-American' /r/, such as in Cliff Richard's "I'll be a/r/ bachelor boy".

The third of Le Page's constraints on linguistic modification, that different models may be in conflict, is insightful and relevant to the general discussion of identity construction in language. When a singer is motivated to simultaneously identify with multiple esteemed groups that have different and conflicting speech models, a mixing of codes may result. For example, the punk singer who exhibits a mixture of American and

British features may be identifying simultaneously with certain esteemed rock influences while defying normative forces by asserting their localised, working class identity.

Trudgill concludes that from the mid-1960s onwards, British singers were "trying *less hard* to sound like Americans" (p. 154, emphasis in original). This is deemed to be at least partly due to a reduction of America's cultural domination in popular music, alongside the developing British pop music identity, forged by the large-scale success of bands such as the Beatles. This process could be viewed as a change in what sort of singing styles were appropriate. The Beatles made it more appropriate to use British features in singing, and these variants took on different social meanings as a result.

Many of the factors outlined in Le Page's model are highly relevant and will be applied in this thesis, such as the amount of exposure a person has to the model's linguistic code. I am interested in investigating whether there is a distinction between those variables which have achieved some level of salience for the singer and those which have not. For example, hyper-American /r/ could occur if there was both a desire to resemble the American model *and* a stereotype that 'Americans use a lot of /r/'.

Trudgill notes that the main flaw in Le Page's theory in explaining modification of pronunciation in singing is that it does not "explain why *particular* … consonantal, vocalic or other variants are retained, rejected or selected, and not others" (Trudgill, 1983, p.159-160). In order to explain why particular variables behave as they do, we need to look at the social meanings of a variant for a given singer at a given moment, their audience, their identity goals at that moment, and the level of metapragmatic awareness they have for that variable at that moment. We also need to acknowledge the role of the vocal habits that may arise subconsciously through repeated exposure to a certain vowel realisation in a certain restricted context.

Because there is limited research on the sociolinguistics of pop singing, Trudgill's study has been considered to be the bedrock of this area. While it is a careful and insightful study, it has several limitations which have often been overlooked in the research which has followed it. Firstly, several claims are made on the basis of either intuition or a couple of examples. For instance, the claim that singers are largely unsuccessful in emulating American singing accents is not fully substantiated, being largely backed up with anecdotes. This style of analysis has been repeated to a large extent in the studies by Simpson, Morrissey and Beal introduced below.

A second limitation of Trudgill's study is that those claims which *are* based on quantitative analysis use impressionistic, categorical coding of the data. This is, of course, understandable since there is no way to do acoustic analysis when the voice appears alongside the instrumentation in a recorded song. But it does disadvantage the study of pop singing as compared to other variationist work, particularly when it comes to studying vowels.

Thirdly, we cannot treat singing like we would treat speech. Trudgill does not take the sorts of issues raised above in section 2.2 into consideration. This is a flaw which has been remedied in some more recent studies, as we shall see below. Since these factors are not those we would normally consider in a sociolinguistic analysis, it is important to be vigilant in determining the kinds of singing-inherent parameters that might be relevant.

A final limitation of Trudgill's study is that all of the features chosen for analysis were chosen on the grounds that they are "stereotypically associated by the British with American pronunciation" (p. 144). The British features discussed in relation to punk music are also ones which have a high level of salience for speakers of British English. This choice of features assumes that salience of sociolinguistic variables is of importance, but does not investigate its role directly. There is no mention of how non-salient variables are pronounced in singing.

Simpson 1999: Mixed codes, and the importance of mode of discourse

Simpson (1999) provides the first substantial follow-up to Trudgill (1983). Aiming to develop a longitudinal extension to Trudgill's study, Simpson provides a series of qualitative analyses of stretches of pop, rock and Britpop songs by British artists spanning the 1980s and 1990s. Simpson (1999) uses the term *USA-5* to describe the set of variables which had been studied by Trudgill and which seemed to be most stereotypically associated with American accented singing. These variables are:

- intervocalic /t/ – flapped in American English (AmE), aspirated in British English (BritE);
- non-prevocalic /r/ – AmE is at least partially /r/-ful, most BritE is /r/-less;
- BATH – AmE [æ], Southern BritE [ɑ], Northern BritE [a];
- PRICE – AmE monophthongal, BritE diphthongal;

- LOT – AmE unrounded and open, and merged with THOUGHT, BritE rounded, close and distinct from THOUGHT.

By the early 1980s, the punk movement was in decline and Simpson found that artists tended to mix USA-5 features with British ones, but now with prestige RP type variants in place of the working class variants used in punk. He argues that this shift reflected the tastes of an aging group of record company executives, who acted as 'cultural intermediaries'. A comparison of recordings from the late 1970s and the early 1980s by punk group the Stranglers reveals a progression from working class to RP features. The singer changed his pronunciation and the band changed their image along with the changing market.

Van Morrison also changed his singing pronunciation to coincide with a changing image. Songs recorded in the late 1960s used mainly American pronunciation, while his 1988 album *Celtic Heartbeat*, which emphasises his Irish roots, is less American. The song 'Sense of Wonder' includes a spoken passage with extremely localized Belfast Irish features. On the one hand this is a clear statement of local identity, but it raises another issue. As Simpson puts it, "it would sound odd to hear a performer mimic a spoken variety other than his or her own during a voice-over sequence" (p. 359). This brings us back to the important notion that singing and speech are not dichotomous modes; there is a range between the very sung and the not at all sung. The pronunciation styles which are appropriate to singing are not appropriate to speaking, even in the middle of a song. This is especially so in the case where an artist is aiming at a projected 'authentic self' kind of identity.[5]

Simpson's study differs from Trudgill's in its focus on the changing cultural patterns in the music industry as essential to a theory of singing accent. He quotes leading popular music theorist, Simon Frith, stating that "…pop musicians of whatever type are acting according to *conventions*" (Frith, 1988, p. 4; quoted in Simpson, 1999, p. 344). Simpson refers to singing accents in terms of *vocal habits*, which are wrapped up in those conventions. A singer who wishes to fit in to a certain category or emulate certain

[5] I use the term authentic in its sense as something sincere or non-artificial, with positive connotations, relating to being 'real' (something like Coupland's description of personal authenticity, Authentic Language 5; Coupland, 2003; Labov, 1972). This is different to other meanings of the term in sociolinguistics, referring to vernacular speech that is not 'tainted' with self-consciousness (for discussion, see Eckert, 2003).

influences is pulled towards whatever singing style happens to be conventional for that musical tradition.

While American-influenced singing accents have been dominant in the history of recorded pop, several music movements (such as British punk in the 1970s and reggae music in and outside of Jamaica) have rejected the use of non-local pronunciation norms. As each of these subgenres gains prominence, injecting non-American accented singing into the market, the associations between American pronunciation and popular music are weakened slightly. However, a new set of associations is formed each time, linking the new accent style to the relevant music genre. In the case of punk, Southern England working-class pronunciations in singing have come to be associated with the punk genre and its associated ethos.

Simpson states that "the USA model … no longer has the resonances it once had … Americanisms … are epiphenomena rather than true indices of sociolinguistic motivation … whereas the USA-model might seem alive and well in many musical arenas, its associations and resonances … have altered inexorably over the years." (Simpson, 1999, pp. 363-364)

Morrissey 2008: Mixed codes continue, and the importance of sonority

A recent article by Morrissey (2008) extends the work of Simpson, taking the same approach of using short examples from a range of artists and analysing variables of note, focusing particularly on those variables which are deemed to be emblematic of AmE, as well as various British features. Morrissey theorises singing style in terms of Bell's (1984, 2001) audience design framework, stating that the use of AmE in pop singing is a case of outgroup referee design, where Americans are esteemed by both singers and their audiences as an agreed referee for singing pop music. Morrissey uses the term *reference style* to refer to the collection of (usually AmE) variants which are used by singers for the purposes of referee design in pop singing. As for the origins of this reference style, Morrissey states that

> Blues, blues-derived rock and pop, and country-based singer-songwriter material are obviously rooted in American popular culture and generally associated with an American reference style. Song material of this kind, for which the expected American reference style is not used, will, like marked forms in language, either sound odd or have a specific effect on the audience. (Morrissey, 2008, p. 199)

As an example of the markedness which occurs when singers diverge from the accepted norm, Morrissey analyses the Small Faces'[6] song 'Sunday Afternoon' which was performed in a Cockney style, unlike any of their other music. The song became a hit but the band was resentful of its success because they saw it as a 'novelty single' (Wikipedia, n.d.), which suggests that the use of Cockney was indeed intended to be either comical or ironic.

While the use of Cockney may have been associated with novelty and comedy, RP style features had very different connotations. Morrissey gives examples of the use of Standard British English features in 1970s progressive/art rock bands such as Pink Floyd. Their music was marketed as being more sophisticated than a 3-min pop song; it was 'serious' music with literary lyrics and theatrical elements. Interestingly, these connotations led to instances of American artists using a British reference style in their singing. For example, American band Jefferson Airplane evokes the kind of 'serious art' connotations associated with Pink Floyd, using RP variants. These examples all support the idea that by the 1970s, AmE was not the only reference style available. Different music styles have different cultural heritages, and different accent norms. While this thesis will not investigate the role of music genre in depth, it is important to recognise that it is a central force in the formation of singing styles. For example, while AmE has been dominant in pop music, RP is the normative pronunciation style in the singing of choral music.

Morrissey's article provides the first clear statement of the potential importance of singing inherent effects. He states that singing is inherently different to speech, and that songs are subject to rules which are not governed by sociolinguistic factors; certain speech sounds "lend themselves better to singing than others" (p. 211). In particular, vowels with greater sonority are argued to be easier to sing. He shows that for British singers that use both BritE and AmE features, the AmE features tend to be more sonorous than their British counterparts. This is an insightful observation, and one that I will explore further when comparing spoken and sung vowels in chapter 4.

Feld et al. 2004: An example of divergence in singing pronunciation

> What do you do if you love country but you hate rednecks? If you love the music, but you cæin't stæind the sæiund? (Feld, Fox, Porcello, & Samuels, 2004, p. 339)

[6] A late 1960s rock and roll band from East London.

While genre norms are crucial to singing style, there are cases where the social meanings of a singer's local speech environment are of such significant importance to identity construction that they will override the norms of singing. Feld et al. (2004) give examples of Southern American accent features that are used to index "country-ness" in singing, both throughout and outside of America. Apache Indians in San Carlos are reported to avoid these country pronunciation features because of their association with "rednecks". Feld et al. describe the way singers in this community refused to pronounce words like *plan* as "[plæiən]". These "twanging" variants do not just mark country-ness, they are also strongly associated with an outgroup that the people in San Carlos do not want to be identified with. In cases such as this, singers are able to "layer varying attitudes within a single utterance … making the sense of their words richer and more complex. Through such practices, Apache country singers are able to layer their attitudes about 'rednecks' through the way they sing their music" (Feld et al., 2004, p. 340). The next section deals with another example of a singer rejecting the dominant singing accent, choosing instead to stylise a local speech variety.

Beal 2009: A fully non-American singing accent

> When I first heard these reports, my natural instinct was to think, "Lads from Sheffield singing in Sheffield accents: so what?" But I soon became aware that such a phenomenon *is* unusual (Beal, 2009, p. 224, original emphasis)

In the accounts discussed so far, singers who use non-AmE features tend to do so only for certain variables, alongside a range of AmE features. Beal (2009) provides an example of a fully non-American singing accent. In the music of Sheffield band Arctic Monkeys, Beal found that the singer Alex Turner uses many local variants, including some features usually associated with older Sheffield speakers such as [eɪ] for PRICE. These 'old-fashioned' variants are said to index localness rather than old-ness, however, because they co-occur with features (such as TH-fronting) which are associated more generally with British youth. Turner therefore constructs a stylised local place identity in his singing, deploying combinations of linguistic resources to project both localness and youthfulness.

One of the most striking indexicalities conveyed by this singing accent, however, is not youthfulness or Sheffield-ness, but rather an anti-mainstream stance. Singing in their 'own accent' is part of Arctic Monkeys' defiance of the mainstream, which is also manifested in their refusal to turn up to receive accolades at high-profile award ceremonies, and their reluctance to talk to the media.

The pronunciation style used by Alex Turner is not a case of "trying less hard to sound American", as Trudgill (1983, p. 153) described the changes in the Beatles' singing:

> far from using their "own" accents in an unself-conscious and natural way, Arctic Monkeys are in fact performing their Sheffield identity in a very knowing and sophisticated manner. (Beal, 2009, p. 224)

The ideology behind this 'knowing' performance relates to authenticity. "Arctic Monkeys make it very clear that pretending to be American when you are in fact from the North of England is both 'fake' and conformist" (p. 225). This highly confrontational discourse is not dissimilar to that of punk music in the late 1970s. Arctic Monkeys are 'indie', and like punk bands, their anti-mainstream stance is an important part of their identity.

Beal appeals to Coupland's (2007) discussion of authenticity to draw a distinction between the local pronunciation style used by Arctic Monkeys, and that which is used by folk musicians in the British Isles, who also use local features in their singing. For folk musicians, singing in their own accent is *historically authentic*, building on local traditions. For Arctic Monkeys, however, the use of a local singing accent is *ontologically authentic*, "projecting the 'ordinary' speech of Sheffield as opposed to the "fake" mid-Atlantic accent of mainstream pop … In the context of British popular music, regional pronunciations index authenticity and independence, whereas the 'American' features outlined by Trudgill (1983) index conformity to the pop mainstream" (Beal, 2009, p. 238). To be ontologically authentic is to 'keep it real', to use the phrase associated with the literature on identity construction in hip-hop (Alim, 2002).

The reason I quote Beal's article at length is because the claims it makes are particularly pertinent to the research questions of this thesis. Are there New Zealand artists who also project a non-conformist stance through the use of NZE features? Or could the Northern British features used by Arctic Monkeys become enregistered in their own right as the sound of indie music and thus be replicated by other artists elsewhere, drawing on the anti-mainstream stance while leaving behind the Sheffield-ness? This is what Beal argues happened to the American accent in pop music. At some point, the USA-5 features stopped indexing 'American' and began to index primarily 'mainstream pop'. It was this shifting of the indexical field in the context of the pop music frame which has led AmE-influenced singing accents to be normative and habitual, being employed "as a matter of course without any conscious act of identity's taking place" (p. 229).

The number of fully localised singing accents in Britain seems to be increasing (examples include Lily Allen, Kate Nash, The Kooks, The Streets). Despite this increase in non-AmE singing styles, a cursory analysis of popular British artists in the music charts reveals that this trend is still the exception and not the norm. 'Own accent singing' therefore attracts the attention of music consumers, most likely because of its novelty. A blog post entitled "Singing in your accent? Annoying?" begins with the following:

> It seems recently that every new hip artist or band has to sing in their own accent. Although this does work sometimes, is it getting kind of grating? (Milkshake, 2007, September 18)

The blogger suggests that while the Arctic Monkeys' singing accent "works for them", there are other British artists whose BritE singing accent either "ruins the flow of the song" or is "bloody annoying". A quick review of the opinions of music fans in the blogosphere shows that while the opinions vary in regard to own accent singing, the artists who break with the AmE-influenced norms tend to provoke a 'love it or hate it' type of response.

A detailed analysis of listeners' attitudes could provide a useful perspective on the study of singing accents, especially in light of the perspective that normative accents may actually be more about stances to the mainstream than identification with America, which brings quite a different (though related) set of ideological discourses into play. The opinions of music consumers are more than just reflections on existing recordings, they are constructive discursive practices which produce and reproduce the ideologies upon which pop singers draw, in their construction of their singing style. Coupland, in his commentary on Beal's article, acknowledges the importance of this issue, stating that "With mediated acts of identity, analysts need to take their place alongside many other interested parties and among a potential welter of interpretive voices" (Coupland, 2009, p. 288).

2.3.2. Pronunciation in Australian popular music

The existing literature on singing accents has largely focused on British artists, but there are also a few studies of singing pronunciation in Australia and New Zealand, which provide a more specifically relevant background to the empirical work of this thesis.

O'Hanlon 2006: More commercial genres are more American sounding

O'Hanlon (2006) reports on an impressionistic analysis of 60 Australian songs focussing on five variables (non-prevocalic /r/, LOT, BATH,[7] GOAT, PRICE). 30 of these were hip-hop songs, and the other 30 were a mixture of pop, rock, alternative, and punk songs. O'Hanlon's aim was to test the claim that Australian hip-hop artists *do not* conform to American pronunciation norms, while artists in other genres do. Each token was categorised as Australian or American, and the results showed a clear distinction between hip-hop and the other genres for all variables, with hip-hop using the least AmE features. It was also found that the rock, alternative, and punk songs had less AmE variants than pop songs.

Non-prevocalic /r/, for example, was realised 24 percent of the time in pop songs, compared to 10 percent in other non-hip-hop songs and 2 percent in hip-hop. BATH was realised as [æ] categorically in pop songs, 62 percent of the time in other non-hip-hop genres and 11 percent of the time in hip-hop. It is of note that there were some hip-hop songs with singing as well as rapping, and that "these songs were not accompanied by any increase in American pronunciation".

The results are discussed in relation to Le Page's theory of linguistic modification, as it was used in Trudgill (1983), with additional insights from studies of hip-hop, which suggest that unlike other genres of music, an important part of being appropriate to the genre is to present a localised authentic self. As Alim (2002) puts it, there is an emphasis in hip-hop on "keepin' it real". At this point, O'Hanlon's argument takes an insightful departure from the theory put forward by Trudgill. Trudgill talked about the decreasing use of AmE variants as being caused by the British artists 'trying less hard' to sound American, but O'Hanlon states that

> there is some sort of active suppression of American accents in the Australian hip-hop performances, with the occasional *slip up* (p. 202, emphasis in original)

This is a very important distinction. Trudgill's position treats American accented singing as initiative, active and knowing, whereas O'Hanlon's claim is that American accented singing is automatic, normative, and something which requires active suppression to be avoided, a similar argumentation to that put forward by Beal (2009).

[7] Note that O'Hanlon's study excluded words from the DANCE lexical set because some varieties of Australian English (AusE) use [æ] in these words, which is treated as the American variant for words in the BATH lexical set.

It is my intuition that the latter position is also more plausible in the New Zealand setting. To sing in a non-American accent requires either a genre-related precedent (as in the British punk in the 1970s, or contemporary Australian hip-hop), or specific attention.

One of the strengths of O'Hanlon's work is its emphasis on the role of musical genre in singing pronunciation, elegantly demonstrating that the most commercially-orientated pop music follows the American pronunciation norms most closely, while in a music tradition which esteems loyalty to locality, Australian hip-hop subverts the American norms.

Dominello 2008: Hyper-local performance in hip-hop

This strong representation of local identity through pronunciation in rap[8] is explored further in Dominello's (2008) study. Rather than comparing local Australian English (AusE) variants to AmE variants, he considers the relevant comparison to be between broad and standard variants of AusE. The study builds on Alim's (2002) finding of higher rates of copula deletion (a feature of AAVE) in USA hip-hop artists' rapping than in their casual speech.

Dominello's study analysed GOAT (comparing broad AusE [ʌʊ] to standard AusE [oʊ]) and PRICE ([ɔɪ] for broad and [aɪ] for standard) for two Australian hip-hop acts, comparing the rap from a recorded song to the speech in a radio interview. In both variables, and for both artists, significantly more broad variants of AusE were found in the song than in the interviews. Australian hip-hop recordings, in eschewing an American model, may actually use hyper-performed Australian English forms. This is a very interesting finding, and one which is not uncommon in occurrences of stylised, staged performance (see Coupland, 2007; Gibson & Bell, fc). This was only a small study however, with several limitations. Aside from the categorical, impressionistic coding of the data, this study could also be questioned on the grounds that a radio interview is likely to produce more formal speech – a fairer comparison of the artists' speaking style might be obtained by recording the members of the hip-hop groups interacting amongst themselves.

[8] Rap refers specifically to the vocal form used in hip-hop music.

Attitudes to the use of AusE in singing vary, and the debate is hotly contested both in hip-hop and in other genres. Australian country singer Greg Champion, says

> I think it is quite important for our national identity and for the advancement of Australian culture that the Australian accent is paramount in Australian country music. (Smith, 2005, pp. 115-116)

In contrast, Aboriginal hip-hop artist Wire MC gives a different perspective:

> As for the whole Aussie accent thing, man, I have a struggle going on with that one personally … having white boys come up to me and saying "you know, maybe you should rap a bit more Aussie". And I'm like "What?! Are you trying to colonize me again dude?! Stop it. Stop it". (Pennycook & Mitchell, 2009, p. 37)

2.3.3. Pronunciation in New Zealand popular music

In the New Zealand setting, the work by Coddington (2003, 2004) provides the only comprehensive analysis of pronunciation in local popular music. The first, smaller, study looks at a single artist's linguistic style between several settings of speech and singing, and the second explores the pronunciation of a wider range of variables for eight artists, along with attitudes to singing accents. These studies serve as a very useful background to my own research.

Coddington 2003: Singing pronunciation in different settings

Coddington (2003) discusses the pronunciation (of non-prevocalic /r/ and LOT) by singer-songwriter Anika Moa across a range of sung and spoken contexts: two albums; two live performances (both sung and spoken sections were analysed); a formal interview; and a relaxed conversation with friends (including Coddington herself). Anika's pronunciation differed not only between singing and speaking, but also between the singing styles of her first and second albums.

The first album was recorded in New York with American musicians, sound engineers and producers. At the time of release, Anika openly acknowledged some of her main influences as being American artists. In interviews about the recording of the second album, however, she claimed that she was not influenced by American artists. This second album was recorded in New Zealand with NZ personnel. The album emphasised her Māori identity, including two songs in Māori, and a song on the topic of the Māori land wars. In the first album, 12 percent of potential non-prevocalic /r/s were realised. While this is a relatively low proportion, a clear shift is seen between albums, with less than 1 percent of potential /r/s realised on the second album. Realisation of LOT as [a], on the other hand, did not exhibit this marked decrease between the two albums, with 63

percent being realised as [a] on the first album, compared to 57 percent on the second. These findings can be explained in terms of several differences between non-prevocalic /r/ and LOT.

Anika Moa stated in discussions with Coddington that she was consciously trying to represent herself as non-American, non-commercial, and staunch Māori in her second album. She managed to eradicate non-prevocalic /r/ from her singing accent, but did not manage to produce LOT as NZE [ɒ]. As Coddington insightfully points out, "for Anika at least, the point is not that it is "appropriate" to sound American but that it is "inappropriate" to sound 'too New Zealand' in mainstream pop music" (p.9). Another explanation is that Anika does not index open, unrounded LOT as American, and thus does not have an aversion to it. Note also the difference in sonority, with the NZE LOT vowel being raised and rounded as compared to the AmE variant. As Morrissey (2008) claimed, in cases where an American variant is much more sonorant than a local variant, the use of the local variant will seem marked.

Coddington 2004: Singers' perspectives on their singing accents

Coddington (2004) expanded upon the earlier study by including eight singers, across three musical genres and for five phonetic variables: non-prevocalic /r/, LOT, BATH,[9] GOAT and PRICE (all of which will also be studied in the present thesis) – the same set of variables as those analysed in O'Hanlon (2006). The analysis of twelve albums by these artists was impressionistic, with each token being categorised as either a NZ variant or a 'non-NZ' (which essentially translated to AmE) variant, with a third category for ambiguous or neutral tokens. Interviews were conducted with each of the singers before data analysis, focusing particularly on issues of ethnic and national identity, attitudes to America, and attitudes to NZE in general and more specifically in singing.

Coddington discusses the interview responses with reference to the diminishing of New Zealanders' cultural cringe (Bayard, 1995) around the sound of their own accent. Coddington found no negative views to NZE in her interviews. She states:

> There are no explicitly negative comments regarding NZE accents in singing and some performers even explicitly state a preference for it. No performers express a

[9] A range of lexical sets are represented with the same [a] vowel in NZE, including BATH, START, PALM and DANCE. This will be mentioned again in Chapter 3.

> preference for singing in AmE (or any other foreign accent) and some are overtly negative about AmE accents in singing. (Coddington, 2004, p. 62)

As with the earlier study on Anika Moa, there were clear links between singers' responses in the interviews and their singing accents. In general, those who claimed to be 'anti-AmE in singing' in their interviews had less American variants and more NZ variants, and those who acknowledged an American influence on their music had more AmE realisations. However, this alignment occurred for some variables more than others, on a speaker-dependent basis. The overall pattern was that singers were more accurate in their self-descriptions for those variables which they used as examples, the variables for which they displayed some meta-pragmatic awareness.

The variables non-prevocalic /r/ and BATH were more salient overall, being volunteered as an example of the difference between AmE and NZE ways of singing by most of the singers. LOT was the least often discussed. This is particularly interesting given the large phonetic difference between the AmE and NZE versions of this vowel, the latter being raised, retracted and rounded in contrast to AmE variants.

Tina Pihema, lead singer of the punk group the Coolies, contrasted their band with others who sing "really American" and stated, "we just try to be ourselves and that's the New Zealand way" (Coddington, 2004, p. 36). The example she gave was that she uses [a] not [æ] in words like *dance*. This proved to be the case, with zero occurrences of [æ] in the BATH variable. However, in other variables, she used AmE variants – in LOT, for example, 67% of tokens were classified as 'non-NZ'. This singer was the only one in the sample to have a majority of her GOAT tokens classified as NZ, though she did not mention this variable in her interview. Another example of this pattern comes with Brooke Fraser's claim, "I don't sound like an American and I don't roll my R's" (p. 36). While she did produce a handful of non-prevocalic /r/s (just under 3%), this was very low compared to her high rates of non-NZ LOT (52%) and non-NZ BATH (58%), which aligns more with her comment that her music in itself does not have a New Zealand sound.

Certain variables align with the singers' self-descriptions while others do not. This may be because singers have conscious awareness of, and thus greater control over, salient variables than non-salient ones. The salience of individual variables can vary from person to person, but there is also consensus; certain variables achieve salience and carry similar meaning for all of the singers. Several of the artists in Coddington's study said that they were aiming for neutrality in their singing accent, to avoid any features

seen as marked with regional meaning. Bic Runga labels this neutrality as a "tidy 'international standard' singing accent" (Coddington, 2004, p. 35). This is an interesting description, and perhaps quite an insightful one. The 'international standard singing accent' which Bic Runga is referring to is most probably something similar to what I have been labelling the 'American-influenced singing accent'. Bic's description suggests that this singing style does not necessarily index American-ness per se, since it is used in pop music from around the world.

Another reason why certain variables seem to be more amenable to an NZE accent than others relates to singing inherent effects, the kind of constraining factors which would not matter in speech, but which may well matter in singing. As mentioned above, Morrissey (2008) argued that sonority is the key singing inherent factor, and that singers will prefer more sonorant variants. In the case of LOT, for example, both the raising and lip-rounding associated with the NZE vowel would be disfavoured in singing on these grounds, giving American LOT as [a] an advantage over NZE [ɒ].

Drager 2003: American as automatic in singing, but difficult in speech

Drager (2003) looked at New Zealand band the Brunettes, which has two singers. The male singer, Jonathan Bree, sings consistently in an American style while the female singer Heather Mansfield uses some NZE variants in her singing. Drager began her research with the hypothesis that American accented singing requires a conscious shift away from speech. She found that the singers claimed it was an unconscious process, even though they were aware of the outcome, they did not think about the accent itself. Heather claimed to sing in a NZ accent live, while using an American accent on recordings, though Drager found this to be false in a live show. It seems that a perception of 'singing in a NZ accent' can be driven by the use of the NZE form of just a few variables.

Another interesting finding in Drager's study, comes from a novel method. She asked Jonathan to read a word list in an American accent. His attempt contained some interesting features, including an inability to distinguish NEAR and SQUARE (which are merged in NZE) and a hyper-American /r/ in the word *lawn*. The fascinating thing about the inability to differentiate NEAR and SQUARE in a word list is that these vowels are distinguished in his singing (as evidenced for example in the open nucleus of *hair* in the song 'Her Hairagami Set'). This suggests that the use of AmE in singing is not based on 'knowing the rules' of the accent (part of Le Page's second constraint to linguistic

modification, discussed above), but on context-specific vocal habits. I will argue in chapter 6 that these context-specific vocal habits arise from the prevalence of exposure to American accents in the context of pop singing.

Gibson 2005: AAVE in NZ hip-hop

In a small-scale study of the pronunciation of non-prevocalic /r/ in three NZ Hip-hop artists, I (Gibson, 2005) found that /r/ was realized near-categorically in words from the NURSE lexical set, and used only very infrequently in any other environments. At that time, I attributed this to be a mixing of the global and local as has occurred in other non-American forms of hip-hop. In order to further examine this, I have recently analysed a few songs by American hip-hop artists Dr. Dre and A Tribe Called Quest and found the exact same pattern, with non-prevocalic /r/ occurring consistently after NURSE and not in other environments. The New Zealand rappers reported on in my 2005 article were actually conforming accurately with the American hip-hop model for this variable. This partial rhoticity may be specific to hip-hop, as neither Coddington (2004) nor O'Hanlon (2006) found any effect of surrounding environment on rates of /r/ in their pop, alternative and punk singers.

2.3.4. Singing pronunciation in other languages

Before concluding this chapter, I will briefly mention three studies which looked at singing pronunciation in languages other than English. The latter two studies have in common a focus on the role of salience as being crucial to the artist's ability to control their singing accent.

Bell 1999: Indexicalities transcending language borders

Bell (1999) analysed pronunciation in the singing performances of the Māori song 'Pōkarekare Ana', as it was used in a TV advertisement for Air New Zealand. He found that despite the fact that the song is sung in the Māori language, phonetic variants semiotically referencing several varieties of English from around the world were used in the delivery of the Māori words. This case study is a fascinating example of the way language can evoke social meaning, even across language barriers. For example, two lines of the song are performed by an African American blues singer in a New Orleans club. AAVE is referenced by several pronunciation features including the

monophthongising of /ai/ in *mai* to /a:/. In another section of the ad, a Pākehā[10] man is busking with a guitar in the London underground. His pronunciation is characteristic of the "down-home, anglicized pronunciation of the Māori language that an ordinary Pakeha would use" (Bell, 1999, p. 534). It is of particular interest that the actor in this segment was able to deliver a convincing Pākehā accent while singing in *Māori*, especially when compared to the examples above of New Zealand singers, none of whom were able to adopt this accent when singing in *English*.

Prince 1987, 1988: Maintaining a stigmatised dialect

Prince (1987, 1988) described an Israeli singer, Sarah Gorby, who was successful in her desire to maintain her minority dialect (Bessarabian) of Yiddish in her singing style, despite the dialect being stigmatised, and despite the lack of prevalent singing models in this dialect. Notably, Gorby was more successful at using the consciously desired Bessarabian phonological variants in open class words than in closed class words.[11] This is explained as being related to the higher level of salience drawn to open class items. For a change which is consciously controlled, the desired variants should occur more often in words which are salient to the speaker. Prince states that "the most-consciously aimed-at target is reached more successfully in open-class items, the items to which speakers are better able to attend" (Prince, 1987, p. 110). Labov (2001) refers to Prince's study as a case where attention paid to speech is clearly the cause of the style shifting. This seems to be the case, but it should be emphasised that it is the interaction of attention to speech with desire to project a certain identity which characterises the situation. The style shift is in the opposite direction when compared to Labov's work, where attention to speech causes participants to shift towards more standard variants. Gorby uses more stigmatised variants when attention to speech increases.

Yaeger-Dror 1991: Effects of cognitive salience

Yaeger-Dror (1991) presents an insightful study of a collection of recordings of singing and interviews with Hebrew speakers in Israel of different backgrounds. A strength of this study is the way it shows elegantly that "different song genres have different dialect targets" (p. 312). Not only do different genres have different pronunciation styles, but those styles do not map cleanly onto the speech styles of the corresponding singers. She

---

[10] New Zealander of European descent.

[11] Open and closed class words are also commonly referred to as lexical and function words, respectively.

also found that singers' pronunciation varies according to the cognitive salience of a given token. Words with more salience will be more in line with the singer's conscious identity goals than less salient items.

Cognitive salience was hypothesised to be greater on open class words, infrequent words, beginnings of words, and beginnings of lines. Also, the occurrence of one variable that carries a high degree of social meaning is hypothesised to increase the salience of surrounding variables. Both Yaeger-Dror and Prince's studies raise interesting arguments about the role of attention and salience in singing, which will provide useful insights for my discussion of differences between singing and speech in the New Zealand setting.

2.4. Summary and Critique

Patterns emerge in reviewing all of the literature on pronunciation in popular singing. Though several theoretical stances have been taken by different authors, they tend to revolve around notions of linguistic modification for identity purposes. There are conflicting views on whether AmE in singing is a wilful act of identity, or an automated default, though there is building evidence for the latter. There are no cases in this literature of non-American singers using the exact same dialect in speech and singing. Singing accents employ a mix of American and local features in many cases, and in cases where singers project their local identity, a stylised and sometimes 'hyper' version of a local dialect is used. This suggests that when people avoid using the normative AmE variety, it is not a simple case of 'trying less hard' to sound American, it requires active stylisation, or crossing. Recalling the terms introduced by Silverstein (2003), in singing pop music "appropriateness-to" the context may involve singing with AmE features, while "effectiveness-in" might involve the use of local variants, shifting the indexical field.

There are two main shortcomings of the existing studies of singing pronunciation which I will mention in order to motivate the empirical work to be discussed in the next few chapters. The first relates to the restricted range of variables that have been analysed, and the second is that the studies all have in common the use of impressionistic techniques for analysing vowels.

Limited range of variables

The studies discussed above have focused on a fairly limited range of variables, often centring their analysis around the USA-5 features, which "are stereotypically associated

… with American pronunciation" (Trudgill, 1983, p. 144). The NZ and Australian studies continued to focus on the USA-5, with the notable addition of the GOAT vowel, which appears to have some local relevance in both NZ and Australian singing. The AmE variant of GOAT is described as rounded and retracted, transcribed as [ou], while the NZ variant is transcribed by Coddington as [ʊʉ], and said to have an opener nucleus and fronter offglide than either the British or AmE equivalents.

It is interesting that most studies mentioned have looked at roughly this same set of variables, those which have some salience as Americanisms. While they acknowledge this as a methodological choice, they do not justify it with any rationale. Presumably, Trudgill and Simpson, for example, do not think that British singers shift to American variants of non-salient variables. In the New Zealand scenario, this would mean using New Zealand pronunciation on those vowels which do not have an American stereotyped equivalent. This seems unlikely to me, and is something which will be tested in the vowel production study.

In my comparison of singing and speaking, I will include several vowels which have never been studied before in the literature cited, such as DRESS, TRAP and GOOSE. It is my hypothesis that variation between singing and speaking should occur across all variables, if AmE is the normative style requiring least attention to singing. When a singer has a desire to *avoid* AmE in singing, perhaps to project an 'authentic persona', they will be most successful at styling a non-AmE variety when attention to singing is heightened.

Impressionistic description

Another shortcoming of the approach taken in the studies reviewed is that they all use impressionistic techniques for analysing variables. This is not too problematic for consonants, but can lead to analyst bias in the description of vowels. Particularly problematic is the use of quantitative methodologies based on binary categorisation of vowel tokens as 'American' or 'local'. To mitigate the issues this causes, Coddington (2004) included a category for ambiguous or neutral tokens. One in six tokens fell into this category, suggesting that the NZ versus non-NZ categorisation was concealing a lot of variability and doing a lot of 'rounding' on the phonetics of tokens.

Thomas (2002:189) argues that "sociolinguists too often do not examine closely the phonetic details of the variables they study". The PRICE vowel, for example, is often coded as either diphthongal or monophthongal even though actual realisations of PRICE

vary in a gradient way involving “factors such as duration and steady-state structure that require instrumental measurement” (p. 189). Of course, these authors all have a very valid defence in that it is impossible to use acoustic analysis on vowels which occur in the context of a fully produced song. Acoustic analysis requires the collection of voice only recordings, extracted from their accompaniment. My comparison of singing and speaking does just that, building on the methodology used by Stone et al. (1999) that was discussed above.

Chapter 3 Vowel Production Study: Design and Methodology

This chapter outlines the methodology and background for the acoustic comparison of spoken and sung vowels which follows in chapter 4. After introducing the basic design of the study, I will discuss the way the three singers were chosen, and describe the conditions under which each of the recordings was made. After providing some details about the themes discussed in the interviews, a biography of each of the singers is given, including their responses to the interview questions. This contextual information will be important when analysing each singer's pronunciation. The chapter will conclude with a discussion of the processes involved in the acoustic analysis of the vowels.

3.1. Design of the Study

The lack of existing research into singing accents using in-depth phonetic analysis encouraged me to conduct an acoustic comparison of singing and speaking. This perspective means sacrificing the range of artists and songs under investigation for the sake of a more detailed analysis across a larger number of variables. Following the methodology used in Stone, Cleveland, & Sundberg (1999), recited versions of songs were compared to their sung counterparts. This methodology allows for pairwise comparisons that hold constant the effects of linguistic environment and word frequency. I intended to analyse one song from each of three singers, though I ended up with extra songs for one singer (the reasons for which will be described below). In addition to the comparison of singing and reciting of song lyrics, an interview was conducted with each of the singers.

Limiting the scale of the study to just a few songs has several benefits. Firstly, it allows the extraction and hand-checking of formant values for each vowel, as opposed to making impressionistic categorical judgments. Secondly, eight vowels were analysed rather than just a few. As mentioned at the end of the last chapter, previous studies have tended to focus on variables which are considered to be 'Americanisms'. The eight vowels analysed in this study allow the testing of the hypothesis that modification towards the American norm happens across a range of variables, not just those that have special salience. Thirdly, the use of a smaller sample means that the analysis can examine individual instances and investigate contextualised meanings. This has been

done in addition to looking at the broader tendencies across averaged data, which can sometimes miss important information when considered in isolation.

3.2. Choice of Participants

There were several decisions to be made about which singers to approach to be involved in the project. Firstly, I decided to approach singers with whom I already had some level of friendship. As mentioned in the introduction to the thesis, I myself am a singer. There was therefore a degree of inherent bias to the process of choosing participants. Rather than ignoring this fact, it is worth taking a moment to outline some of my reasons for choosing acquaintances to participate in the study. The main reason was that a high level of trust is required in order to ask a singer to extract the vocal files from a recorded song and allow them to be scrutinised by a sociolinguist. Involving friends meant that any potential awkwardness around this process was mitigated. The ease and practicality that came with choosing participants from my own social network should also be acknowledged as part of my decision process. It should be noted that for the singing data, we are in the methodologically unusual situation of having no issues with the 'observer's paradox'. The recordings were already made before this study was conceived, and are therefore free from the interviewer effects often grappled with in sociolinguistic studies.[1] The reciting and interview data, however, are vulnerable to such effects.

The second decision made about choice of participants was to invite only male singers, in order to make the acoustic analysis more reliable. When dealing with singing, as mentioned above in section 2.2.2, singers may change articulation in cases where the pitch (f0) gets higher than the first formant of the vowel. This happens more for females than for males, given that they sing with higher average pitch than men, but only slightly higher formant values. The other pitfall of a higher f0 is the potential for the automatic formant tracker to misrepresent f0 as F1. The likelihood of this problem is reduced by using male singers.

A third decision about which singers to invite as participants was that the singers chosen needed to perform songs which they themselves had written. There are already a lot of

[1] However, this turned out not to be the case for the song 'Wintersun', which was recorded in the experiment (see below).

complex variables involved in this analysis, and introducing questions about the effects of covering of another person's song would add unnecessary complication.

Finally, the musical genre of the artists needed to be considered. As we will see below, the three artists chosen each have distinct styles of music, but they can all be grouped somewhere within the guitar centred singer/songwriter type of style, distinct from punk, heavy rock, electronic, or hip-hop styles, for example. It would have been interesting to include a singer from each of three very different genres, but if that route had been taken, it would be difficult to ascertain whether the differences between singers were idiosyncratic or genre related.

Having taken all of these considerations into account, three singers were invited to participate in the research and they all agreed. They are Dylan Storey, Andrew Keoghan, and John Guy Howell. Dylan and Andrew both perform under their own names, while John is a lead singer in the group the Broken Heartbreakers. All three singers are Pākehā males in their early to mid thirties, based in Auckland, New Zealand. They are all semi-professional musicians who work day jobs as well as playing live gigs and recording their original songs. Detailed information on each of the singers is provided below, in section 3.4.

3.3. Collection of Recordings

In this section, the dataset that was obtained will be described, including details of how songs were chosen and the conditions under which recordings were made.

3.3.1. Ethical considerations

Ethical approval was obtained before approaching the participants, and the letter of approval is included in Appendix A. The appendix also includes the information sheet which each participant read before beginning the recordings, and the consent form which was signed. Each of the singers gave permission for their names to be used in the reporting of results, and subsequent permission was obtained in writing to include the song recordings in their entirety as a supplement to this thesis.

The participants were told that the purpose of the study was to "learn about the differences and similarities between singing and speaking", and in practice they also knew that I was particularly interested in pronunciation. They were not told that I hypothesised American influence, or that I was interested in cases of New Zealand-like pronunciation, nor were they told that I was particularly interested in vowel sounds. I

avoided discussing the project as much as possible until after the recordings of reciting had been made. Afterwards, during the recorded interviews, I was more open about the topic of the research. It should be noted that I had already had several conversations about singing pronunciation in the past with Dylan, so he would have had more insight into my research questions than the other two singers.

3.3.2. Choice of songs

I let each singer choose the song(s) they would like me to use in the study. I ended up analysing one song each for Dylan and Andrew, and four songs by John. There are two reasons why I collected extra songs for John. Firstly, John's song 'Wintersun' had many fewer words than Dylan and Andrew's songs, and had low token counts for several vowels that I was planning to analyse. Secondly, while Dylan and Andrew provided pre-recorded vocals of singing, John sang the vocal for 'Wintersun' live because at the outset he could not get access to his studio recordings. I was concerned that John may have sung differently under the artificial setting of the experiment. For these reasons, I went back to John a second time, and he supplied me with studio recordings of the singing for three additional songs, and I recorded him reciting the lyrics to those songs. Overall, each of John's songs had fewer words than the songs collected for the other two singers so that the total number of words across all four of John's songs is not much above that in Dylan's song. The title and word count for each of the six songs analysed is shown in Table 3.1. The lyrics of all six songs are reproduced in Appendix D.

Table 3.1: *Titles and Word Counts of Analysed Songs*

Singer	Song	Words (excluding repeated lines)	Words (total)
Dylan Storey	'Sold it all away'	199	249
Andrew Keoghan	'Gloria'	133	171
John Guy Howell	Total	258	371
	'Wintersun'	95	149
	'Mi Corazon'[2]	78	90
	'Ballad of Archibald Baxter'	46	78
	'Tell that Boy'	39	54

[2] The chorus of 'Mi Corazon' is in Spanish. When I asked John about this, he said the chorus was influenced by, and a reference to, the Clash's song 'Spanish Bombs' in which Joe Strummer uses many Spanish words throughout the song, including the chorus hook, "oh my corazón". The Spanish line of John's song is excluded from the analysis, and also from this word count.

3.3.3. Recording of singing, reciting, and interviews

The singers were all recorded either in the studio of one of the participants or in my own home. With the exception of 'Wintersun', the recordings of singing were taken from existing recordings. For these five songs, I ensured that I had just the raw vocal recording with no accompaniment, and no effects like delay or reverb. I do not have detailed information about the processes used in the creation of these pre-existing recordings. Each artist will have used different microphones, which may have a slight effect on the spectral structure of the recording, though not to the extent that it would affect the analysis. Some of the audio files received were final edits of the songs, while others were single vocal takes, without any 'comping'.[3] None of the vocals used any copying and pasting – repeated sections are genuine repetitions, not duplications of the same recording.

For the recording of 'Wintersun', I first recorded John playing the guitar part to the song. He then recorded the vocal afterwards, while listening to the guitar part in headphones. This amounts to a similar process to that which would have occurred for the other recordings in some respects. However, the environment of recording, being part of a linguistics experiment, was rather different to a studio setting, and may have caused John to have heightened awareness of his pronunciation. It was decided to include 'Wintersun' despite these issues, since a large portion of the analysis had already been done by the time the extra three songs were obtained. While the results for this song have been included in the overall analysis, I will be mindful of any differences between this song and John's other songs when analysing the results.

All three artists were recorded reciting the words to their songs. I asked each singer to recite the words in 'your normal conversational voice'. All reciting was done from memory with no rehearsal, and all recordings were completed in the first take. Dylan and Andrew's reciting, as well as all three interviews and both the singing and reciting of 'Wintersun', were recorded onto a PC using Protools recording software and a Rode NT1 condenser microphone. The reciting of John's additional three songs was recorded onto a Mac using Logic Studio with an AKG C4000B condenser microphone.

After each singer had recited their song, I conducted a fairly informal interview/conversation with each of them. I had an interview schedule which guided the

[3]Comping is the editing together of sections from different vocal takes into a single sound file.

conversations (see Appendix B), and while all of the planned questions were covered for each singer, I also encouraged them to talk freely, letting the conversation move to whatever topics arose. The main themes of the interviews were as follows:

- Musical influences
- Target audience and musical aspirations
- Awareness of singing accent, including specific examples where forthcoming
- The extent to which their persona as a singer reflects their normal day-to-day persona.

The interviews serve a dual purpose. Firstly, as a source of linguistic data, they allow the examination of any pronunciation differences between the reciting of songs and conversational speech. Secondly, they are a source of content, gathering information on the singers' influences and their ideas about pronunciation in singing.

3.4. Singer Profiles

Before introducing each of the singers in turn, I want to give the reader some idea about the profile each of the artists has within the NZ music scene. I have used the number of views of the artists' MySpace pages as a very rough measure of their level of exposure compared to other NZ bands. Table 3.2 shows the number of views for the three singers analysed in this study along with several other alternative and mainstream NZ artists. I chose the top three New Zealand artists from the Kiwi FM Top 10 (nationwide NZ-music-only station), the 95bFM Top 10 (Auckland student radio station), and the nationwide sales chart for singles for the week of 7 December 2009. I also chose three bands with top-selling albums in 2008 (Recording Industry Association of New Zealand, n.d.).

MySpace view counts are by no means a non-problematic measure of exposure – they are biased by the length of time an artist's page has been available and the age group of the artist's target audience, amongst other things. Nevertheless, this figure provides a simple way to get an idea of the kind of profile these artists have relative to other independent artists, and relative to major-label commercial acts. The figures show that the three singers have relatively similar levels of exposure to one another, with their page view counts sitting together at the upper end of the bands from local radio charts, but well below any of the artists that compete with international acts for a place on the sales-based charts. Each of the participating singers will now be introduced in more detail.

Table 3.2: *Number of MySpace Views for a Selection of NZ Bands, including the Participants of this Study*

Band name	Reason chosen	MySpace views (at 9 Dec 2009)
Cinema 90	Kiwi FM Top 10	1,937
Rackets	95bFM Top 10	4,130
Farmer Pimp	Kiwi FM Top 10	10,386
Bandicoot	95bFM Top 10	13,083
Andrew Keoghan	*Participant*	*13,265*
Dylan Storey	*Participant*	*21,548*
Broken Heartbreakers	*Participant*	*24,333*
Stan Walker	RIANZ singles chart	43,097
Dane Rumble	RIANZ singles chart	43,367
Sola Rosa	95bFM Top 10	67,902
Tiki Taane	RIANZ annual album chart	77,120
Opshop	RIANZ annual album chart	133,831
Gin	RIANZ singles chart	303,394
The Black Seeds	RIANZ annual album chart	407,429

3.4.1. Dylan Storey

Dylan Storey (pictured in Figure 3.1) plays 70s-influenced blues/country/rock. One reviewer described his style as follows:

> The balladeer indulges his non-conformist views through abstract folk, psychedelic country and alternative blues rock. (Einstein Music Journal, n.d.)

Dylan[4] grew up in Northland and moved to Auckland at the age of 18 to pursue his interest in music and also to study. Having studied engineering, Dylan spent two years living in Denmark where he became fluent in Danish. Upon his return to NZ he completed a Bachelor of Arts degree, studying Māori, French and linguistics. His lyrics are intellectual and philosophical, covering themes like "kiwi troops in WWII, primordial evolution, non Euclidian space, colourful politicians, reckless drivers in expensive cars and self reflective existentialism" (MySpace, n.d.-b). The lyrics sometimes situate Dylan's songs in a specifically New Zealand context. As well as having some local themes, it is of interest that Dylan occasionally uses NZ-specific words, such as *taniwha*,[5] in the most recent album. Dylan has independently released three albums, which have progressed from solo efforts recorded in a bedroom through to professionally mixed and mastered albums performed with the various incarnations of his live band. One single from the second album reached the bFM and Kiwi FM top

[4] For simplicity's sake, I will refer to the singers by their first names throughout the thesis.

[5] Māori word for a mythical monster, in fairly common use in general NZE.

10 chart and the newly released third album has received very positive reviews in several major NZ music publications. He has been on nationwide tours to support the release of these albums and has a strong reputation in the Auckland live music scene. Dylan is the driving force of his project, writing, arranging and singing all the songs. Though he is not a formally trained singer, Dylan has received some vocal coaching in recent years.

Figure 3.1: Dylan Storey

Dylan states that he was most influenced by 1970s American rock bands such as Grand Funk Railroad. He was inspired by the "raw energy that seemed present in the seventies" and placed a high degree of importance on a band's musicianship and the kinds of sounds they could achieve from their instruments.

Dylan is highly conscious of the accent he uses when he sings. This is not altogether surprising given his training in linguistics and languages, which may play an important role in his singing pronunciation. He says that there is a tension between not wanting to sound American, and finding it difficult to use New Zealand vowels in singing:

> Once you start thinking about it, and once you've got a little bit of a background in languages as I have, it starts to become painful to blatantly sing American vowels, but going the other way is quite difficult too, you have to be really conscious. For one thing – and I'm not sure if this is just perceived or actual – but it does seem easier to sing an American accent.

Dylan was particularly conscious of the GOAT vowel. In the interview, when asked whether he had thought about the accent he sings in, he said "the one that springs to mind is the /ou/, like 'so I know'". He proceeded to perform several examples of the vowel in different ways, trying out first a spoken NZ example, then a sung AmE one –

saying "it just seems far easier just to round it", referring to the sung version. The actual realisations produced in these self-conscious performances will be presented in the discussion of the GOAT results, in section 4.3.7.

While Dylan acknowledges that there are pronunciation differences between his singing and his speech, he says that he does not think there is a difference between his persona in singing and his daily conversational persona. As he discussed the reasons for the tension between American and NZ pronunciations in singing, he suggested that the style of music is an important factor – a NZ accent would "sound ridiculous" in mainstream pop, but:

> with the type of music I'm doing, it can't be pinned down into a category that easily so you can get away with having a unique spin… My voice is not that mainstream radio friendly voice at all, it's gonna be an acquired taste, it's gonna be for someone who's into more obscure music to appreciate, and so for that reason, the accent thing goes with that – you can get away with it.

His opinions on NZ accented singing are mixed, he seems to believe that it is good and right to sing with a New Zealand accent, but that it is "an acquired taste". Dylan mentions a local singer, Tono, who uses NZE in his singing, suggesting that when he first heard Tono sing he had to rationally overcome a negative initial reaction to the sung NZ accent. After the initial surprise, "you realise … that's what we *should* sound like". While Dylan feels he can 'get away with' some modifications towards a sung NZ pronunciation, he opts to stick largely with the more accepted model. When discussing the recording of vocals on his most recent album, he states:

> If I wanted to on this last album I could have spent more time trying to make sure that it was all purely New Zealand accent. But I didn't want to, I wanted somewhere in the middle. For one, I don't think NZ pop culture is ready for that, I think it's something that's gotta be eased into, and I don't think I'm quite ready for it either.

3.4.2. Andrew Keoghan

Andrew Keoghan (pictured in Figure 3.2) is an Auckland based singer/songwriter who plays acoustic pop music with a jazz influence. His solo performances are distinctive, using guitar and violin with delay effects to create an engaging live show. Andrew grew up in Dunedin and Christchurch, in the South Island of New Zealand. He attended classical singing lessons from childhood through to his early twenties, by which time he had passed the exams for Licentiate of Trinity College London which requires an extremely high level of technical competence. At this time, he sang in Puccini's opera Turandot, performed by Christchurch Operatic. Andrew's high level of classical training

provides interesting points for comparison between the three singers, particularly in relation to aspects of his technique such as larynx position (which is more likely to be low and fairly stable), and jaw opening (which is likely to be greater).

Figure 3.2: Andrew Keoghan

Aged twenty-one, Andrew moved to Wellington to work as a news reporter for TVNZ. During his five years there he sang as a jazz vocalist in various Wellington bands. Since moving to Auckland in 2006, he has shifted his musical focus to the performance of his original songs. At the time of writing, he is set to release his debut album, which includes a re-recorded version of the song being analysed in this thesis. Between his career as jazz singer and his present work with original songs, he has a lot of live experience. He plays frequently in Auckland and has been on two nationwide tours.

He states that his audience is a thinking audience in their 30s and 40s, and while at times he has tried to make his music appealing to a younger crowd, he is happy to play to whatever audience appreciates his music. Andrew's earliest influences include the late-70s disco/funk of Michael Jackson's *Off the Wall*, along with the Beatles, specifically their later albums such as *The White Album* and *Sgt. Pepper's Lonely Hearts Club Band*. After leaving school, Andrew was influenced by jazz singers such as Ella Fitzgerald and Billie Holiday, and by 'crooners' like Frank Sinatra and Bobby Darin. More recent influences include MGMT and the Presets, electronic alternative pop acts.

When I asked Andrew about his singing accent, he replied in more general terms than either Dylan or John. He said he had thought about his singing accent, but rather than referring to segmental aspects of pronunciation, he said "I'm conscious that sometimes I

sound too laid back in songs, maybe that's because I used to be a jazz singer". Referring to the song chosen for this study, he says "in the example I gave, I definitely gave it more edge than my natural voice ... it's definitely not completely natural". I did not ask him to define what he meant by 'natural', but he went on to elaborate somewhat, saying "when I listen to 'Gloria', I do think it's not quite me being me … the way I've sung it". He seems to be referring to the overall impression of the vocal – this may include musical/vocal stylistic elements as well as pronunciation features. In answer to the more specific question, "is your singing accent different from your speaking accent", he answered "I think it is – when I talk it just comes out naturally and I don't think about the way I say it".

Andrew appears to be aware that there is a difference between his singing and speaking pronunciation upon being asked directly, but he does not seem to have many preconceived opinions about it. Specifically, he does not seem concerned about whether or not this difference results in the projection of an American or NZ identity, something which was specifically raised by both Dylan and John. This could be related to the fact that he has experience singing in other genres of music, including opera, in which the idea of presenting a sincere version of self in singing is irrelevant, giving way to the overt performance of pre-determined characters.

I asked Andrew if he had noticed any trend for New Zealand artists to sing more with their speaking accent. After some hesitation, he said

> maybe it has become more acceptable to sing in a Kiwi accent and maybe that was hidden for a while for whatever reason, maybe there was that cultural cringe that there isn't so much anymore, and people are willing to just let it out.

While being understanding of others who choose to do so, Andrew expresses no desire to use NZE in his own singing. He is more conscious of the emotional energy of his vocal style than of any particular features of pronunciation.

3.4.3. John Guy Howell

John Guy Howell (pictured in Figure 3.3) is one of the lead singers of the folk-pop group the Broken Heartbreakers, who formed in 2002. John grew up in Dunedin, and then spent two years in England in the late 1990s before settling in Auckland in 2000. He states that his largest musical influences when growing up were bands associated with the local independent record label, Flying Nun. Bands on this label, such as the Clean and the Chills, developed the internationally recognised post-punk movement

referred to as the 'Dunedin sound' by creating "their own version of punk rock" (Churton, 1999, p.69, cited in Coddington, 2004). The linguistic style of the Flying Nun bands was mixed; in the Chills' song 'Pink Frost', for example, features of NZE, BritE and AmE can all be identified, though the NZE tokens may be particularly salient due to their novelty. The members of these Dunedin bands were about ten years older than John as he grew up and they were role models for him. He also cites English 1960s pop groups like the Beatles and the Kinks as musical influences. John has received no formal vocal training.

Figure 3.3: The Broken Heartbreakers, John Guy Howell at front

The Broken Heartbreakers have released two albums and have toured extensively throughout New Zealand and also in Melbourne and England. Their second album received a number of excellent reviews, the songs being described as "perfectly formed country and folk-pop songs" (Otago Daily Times, quoted on MySpace, n.d.-a).The group is based in Auckland and includes five members, centred around John Guy Howell and Rachel Bailey, who share lead singing duties fairly evenly. John wrote the lyrics and melodies to all of the songs to be analysed in this study but the band share in the authorship of the music since they are part of the process of interpreting, harmonising and arranging the songs. Three of the four songs chosen for analysis are on the band's third album, *Wintersun*, which is in the process of being released at the time of writing.

John, like Dylan, is aware of his pronunciation in singing. In response to the question "have you ever thought about the accent you sing in", he said that he had, recounting:

> I heard Martin Phillips [of the Chills] singing in what I considered to be a strange voice, and I only realised years later that it was 'cause he was singing in a Kiwi accent, and, good on him!

He says he has also noticed that Phoenix Foundation use "a reasonably broad New Zealand accent" and that Don McGlashan of the Muttonbirds has always sung "comparatively in a Kiwi accent". John esteemed all of these artists for their use of features of NZE.[6] As for his own singing, he said, "I have noticed it and become aware of it and actually made a conscious decision to sing more in my speaking voice in the last four or five years". Despite this decision, when asked whether his singing voice is still different to his speaking voice, he says:

> it is, slightly, I think less than it used to be, and it's all part of the process of finding what 'the artist' calls 'their voice' … as to whether I would want to sing completely in my speaking voice, I can't imagine that at the moment so it's very hard to say.

In one song, however, he makes a point of trying to use his spoken pronunciation style:

> … the song 'Angela' on our album, I made a point of it sounding, I had to re-record it 'cause when I listened back I didn't feel like my accent was coming through strongly enough

He said that in this particular song more than any other, he wanted to portray a NZ accent due to the lyrical content – a plea to an estranged loved one to return home from London for a visit. Unfortunately, the vocal track for this song could not be obtained for acoustic analysis, though its pronunciation will be considered impressionistically in the discussion of results in chapter 5.

When asked about examples of differences between singing and speaking styles, DANCE was the first to come to mind: "I would never have done *dance* [pronounced /dæns/], even when I was a kid". Another example comes from the story about re-recording 'Angela' to emphasise his NZ accent. Two of the words he noticed in this re-recording process were *care* and *there*, which have a raised nucleus in spoken NZE as a result of the NEAR/SQUARE merger. He tried to produce more NZ-like versions of these words

---

[6] Don McGlashan's "New Zealand-ness" may come about as much through the use of localised lyrics as through the use of NZE phonetics, singing songs about New Zealand places and landmarks ('Dominion Road', 'Harbour Bridge'). The key point about John's influences is that while they do indeed use *comparatively* NZE accents in their singing, this comparison may be attributed to a small range of New Zealand-like aspects of each singer's pronunciation style.

when re-recording the song. In the interview, John demonstrated how the first recorded version of the SQUARE vowel in *care* had a more open nucleus, while in the re-recorded version the nucleus was closer.

In regard to his singing persona as compared to his 'normal day-to-day' persona, John stated that "I would like to try to represent my normal persona, I'm aware that a lot of people like to separate the two but for me personally, I'd like for it to be representative". He sees this as creating a certain distinctiveness which may have a positive impact in attracting listeners, a view which contrasts with previous claims that singers use Americanisms in order to be acceptable to an international market (e.g. Lee, 1982). John says that the Broken Heartbreakers' New Zealand-ness may help them to market their music internationally:

> Being from a country like New Zealand, you could be perceived to be of interest because you're from a small country in the bottom of the South Pacific

The most striking aspect of the interview with John is his very positive attitude towards the use of NZE in pop singing, and his decision to try and do so himself. He also sees this as a widespread change, commenting explicitly on the growth of NZ accented singing:

> I have noticed particularly in the last 2 years really, a real increase in NZ accent-isms in popular music here, and I think it's great, and long may it continue.

3.5. Choice of Variables for Analysis

Eight vowels were chosen for acoustic analysis: DRESS, TRAP, THOUGHT, LOT, START, GOOSE, GOAT, and PRICE.[7] LOT and PRICE are part of the standard canon of variables for research in singing pronunciation, the USA-5. Previous studies of PRICE have described the AmE variant as a monophthong and the BritE, AusE, and NZE variants as diphthongs. It is of interest to investigate the realisation of PRICE with acoustic methods.[8]

---

[7] These vowel labels are based on Wells' (1982) description of the lexical sets of English. Note that the variable I describe as START includes one word from the BATH lexical set in singing/reciting, and several BATH, DANCE and PALM words in the interview data. There were not enough tokens to analyse these separately. Similarly, the THOUGHT and FORCE lexical sets will be described together with the label THOUGHT.

[8] For the purposes of clarity, I continue to use the cover-all terms 'American' and American English (AmE), despite the variability which these labels conceal. Another simplification here is that my

LOT is rounded in NZE, and more raised than AmE. THOUGHT is also included in the analysis since many dialects of AmE have a merger between LOT and THOUGHT (the COT-CAUGHT merger).[9] It will be of interest to see whether the distinction maintained between these vowels in spoken NZE is lost in singing. GOAT was chosen to further the prior analyses of this vowel in the Australasian context (Coddington, 2004; O'Hanlon, 2006) and also since it was picked out by Dylan as a salient variable.

The other vowel described as part of the USA-5 is BATH, which is unique because the distinction between AmE and other varieties is phonemic rather than sub-phonemic. "Many American speakers use TRAP for words like *half* and *grass*, but NZE … uses START" (Hay, Maclagan, & Gordon, 2008, p. 24). This variable seems to act as the shibboleth that distinguishes between American and NZE accented singing for many NZ singers (see the interview data in Coddington, 2004). It is therefore rather unfortunate that the lyrics of the songs analysed in the present study only yielded one token from these lexical sets.

The START vowel is fronter in NZE than AmE (Hay et al., 2008), so this variable provides an opportunity to investigate whether there are differences between speaking and singing in terms of the front-back dimension. GOOSE is also relatively fronted in NZE as compared to AmE, and "Increasingly it is realised with an onglide" (Hay et al., 2008, p. 24) resulting in pronunciations like [əʉ]. It is important to analyse these variables, for which the differences between NZE and AmE are primarily on the F2 dimension rather than the F1 dimension, because F1 distinctions may be complicated by singing inherent effects like a preference for sonority. We can be more certain that any F2 differences found between singing and speaking are stylistic.

DRESS and TRAP were chosen because they have raised realisations in NZE (something like [e], and [ɛ] respectively) as compared to American and British varieties, and yet this aspect of NZE does not seem to be especially salient to New Zealanders.[10] These vowels have not been discussed in previous studies on pop singing pronunciation, and

---

discussion is largely restricted to comparisons between AmE and NZE, despite the fact that comparisons to British varieties would also be relevant.

[9] Descriptions of American English treat the THOUGHT and LOT vowels singly as the CAUGHT vowel, and what I refer to as START is labelled COT. See Clopper, Pisoni, & de Jong (2005) for details of which dialects are affected by this merger.

[10] See Hay, Nolan, & Drager (2006) for a discussion of the salience of these variables, or lack thereof.

there is no evidence that the relative openness of the vowels in AmE is seen as stereotypically American. The results for these vowels will therefore provide a chance to test the hypothesis that style differences between speech and singing occurs across all variables, not just on variables which are salient 'Americanisms'.

3.6. Methodology of Data Selection

All cases of each of the eight vowels being analysed were identified in the song lyrics. It should be noted that function words were included (for example the word *and* as a potential token of TRAP, and the word *to* as a potential token of GOOSE). There are several reasons why function words and lexical words were included together in the quantitative analysis, rather than being treated separately as they might be in other sociolinguistic studies. Firstly, it should be noted that only those function words which received full, unreduced vowel realisations are included in the analysis, which reduces some of the potential problems of considering function and lexical items together. Secondly, treating all of the words together is not problematic for the paired Wilcoxon tests (see 4.1), because they hold constant any differences that do occur between function and content words. Finally, there are not sufficient tokens to separate function words from other lexical items since they make up a sizable portion of the dataset.

Once the potential tokens of interest had been identified, the recorded .wav soundfiles were analysed using Praat.[11] Textgrids[12] were created for each of the three singers in each of the three conditions: singing, reciting, and interview. An interval tier was set up in each textgrid to measure the duration of each token, and a point tier was created for each of the eight variables. The beginning and end of each vowel was marked in the duration tier. The word in which the vowel occurred was also labelled, along with the preceding or following word in cases where the vowel occurred at a word boundary. This ensured that I had a record of the phonological environment of each token. Once the word was entered on the interval tier, points for measurement were then marked on the point tiers.

[11] Praat (Boersma & Weenink, 2009) is free to download software for speech analysis.

[12] A textgrid is used to label a soundfile with either intervals or points that are being analysed. Textgrids provide a way to keep track of where tokens were measured, and can also be used to extract information about the intervals and points by using a script function in Praat.

When determining boundaries in the duration tier, visual cues from the spectrogram and the waveform were usually sufficient for marking the start and end of the vowel segment. When a vowel occurred next to another vowel or a glide, a combination of visual cues and auditory judgment were used.

The ideal point of formant measurement for monophthongs was the vowel's target point, which was determined using F1/F2 maxima/minima as appropriate to each vowel. Where the vowel had a long, target-like steady state, the measurement was taken roughly at the temporal midpoint of the vowel.[13] For the diphthongs GOAT and PRICE, I took two measurements, at steady states where possible, representing the first and second targets of the vowel. These measurements will be referred to as points *a* and *b*. The GOOSE vowel was also treated as a diphthong because it was clear that most tokens had two meaningful targets.

In singing and reciting, all available tokens were gathered. Because the token counts varied greatly between vowels, I decided that for each speaker, for each vowel, I would analyse the same number of tokens in the interview as were found in the lyrics of the song. There are two exceptions to this however. Firstly, for John, the token counts in the interview condition relate to the token counts for the song 'Wintersun'. This is because I did not analyse any further tokens from the interview after collecting John's three additional songs. Secondly, no instances of PRICE were analysed from the interview data, which was the last vowel to be analysed. These decisions were both made after preliminary analyses had been carried out, which had found that the reciting of lyrics and the speech in the interviews were very similar. At that point it was decided that further analysis of the phonetics of the interview data was not necessary.

Several tokens were excluded from the analysis either at the coding stage, or after preliminary analysis of the data. As mentioned above, vowels that were unstressed and had a schwa-like quality were not coded for analysis. The other main exclusions were on the grounds of environment. The three vowels measured as diphthongs (PRICE, GOAT and GOOSE) were excluded from the analysis if they preceded a /w/ or a non-front vowel (for example in the phrases *go on* and *swallow a*). These tokens were problematic because they assimilated strongly to the following sound, which made the second target

[13] In retrospect, it would have been preferable to take an average of each formant across a period in the middle of the vowel in order to reduce the potential impact of any measurement or formant tracking errors.

incomparable with the rest of the data. Additionally, GOOSE tokens preceded by a /j/ were excluded for similar reasons.

One issue with the decision to exclude GOOSE tokens that occurred after /j/ was that it meant excluding the word *you*, which made up a large portion of the GOOSE tokens. While *you* is not included in the main analysis, the realisations of this word will be looked at in their own right in section 4.3.6, which discusses the GOOSE results. Similarly, 16 instances of the word *gonna* were excluded from John's data (11 from singing and 5 from reciting). The word was initially, and perhaps wrongly, coded as containing a LOT vowel. However, upon closer inspection it was found to behave rather differently from other examples of LOT.[14] Once again, even though it is not included in the main dataset used for statistical tests, *gonna* will be considered in its own right when discussing LOT pairs in section 4.3.3.

In previous studies (e.g. Coddington, 2004) choruses and other repeated sections were only included in the analysis once, on the grounds that pronunciation did not change across such repeated sections. Counter to prior studies, I include all repetitions of a word. In doing so, subtle phonetic differences that would not show up in a categorical analysis may be found between two instances of a repeated word.[15] This decision has an impact on the type/token ratio, meaning that in some cases a large proportion of the measurements for a given vowel are based on the same word. While this is somewhat problematic, it was decided that the study would benefit from the greater token count which this method affords.

The first two formants of each vowel token were measured using the automatic formant tracking in Praat. The formant tracker settings instructed Praat to find five formants below 5000Hz, with a window length of 5 milliseconds. The decision about which formant tracker settings to use was made through a process of trialling different settings and taking notes on which settings seemed to capture the formants best, according to visual inspection of the spectrograms. I was open to the possibility that each singer

---

[14] Treating *gonna* as including a LOT vowel was probably overly simplistic, given the many pronunciations of this word.

[15] There was one exception to this, for the phrase *go on* in John's song 'Wintersun'. The phrase is repeated twelve times in each chorus, so only the first chorus was analysed to avoid gross weighting of John's LOT data towards this phrase.

might need a different setting, but in the end the best setting was the same for all three singers.

There were several tokens where Praat's formant tracker did not automatically detect the first two formants correctly, and this was noticed during the marking of measurement points in the textgrids. In order to manage this problem all tokens were manually checked to ensure that the tracker was picking out the first two formants and that the points were appropriately placed for each formant. The most common problem was the tracker inserting an extra formant somewhere between the true first and second formants. In these cases, the first solution was to move the measurement point to another part of the vowel where the tracking was correct, however this was only done if the formants were roughly the same at both points. If this was not possible, a manual reading of F1 and F2 was taken and noted down with the token ID number in a separate document. These values were then added to the master data spreadsheets later compiled. There were a total of 26 of these manual replacements across the whole dataset (which includes a total of 765 vowels).

A few tokens were discarded because the formants could not be robustly determined. There were 5 sung tokens, 4 recited tokens and 1 interview token removed for this reason. For some tokens, the pitch was not found by the automatic pitch tracker. In these cases, I collected the pitch manually by getting a pitch measurement from the nearest possible place to the formant readings, ensuring by ear that the pitch was the same.[16] Four tokens were removed from the interview data because the pitch could not be determined.

After the manual checking of all of formant tracking for each vowel, a script was used to extract the data from the audio signal.[17] The script outputted a text file which included the following information: the name of the soundfile (which included the singer name, the song name for John, and the condition); the number of the token; measurements of f0, F1, and F2; the duration of the vowel; and the word (with

[16] Another way to determine pitch would have been to measure the duration of the relevant pitch periods and then calculate f0 from these. This was not done, however, since several of the cases where pitch was difficult to determine were caused by glottalisation, having large gaps between the glottal pulses.

[17] This script was originally based on 'collect_pitch_data_from_files' distributed under the GNU General Public License, copyright Mietta Lennes, 2003. The script was modified by Jen Hay in February 2009 for the present project.

neighbouring words if vowel occurred at a word boundary). For the diphthongs, the output included measurements for F1a, F2a, F1b, and F2b. The next chapter presents a detailed examination of this resulting data.

Chapter 4 Vowel Production Study: Results

The three singers introduced in the last chapter provide a range of stances and attitudes towards NZ accented singing. John has a very positive attitude to NZE in singing and explicitly stated that he tries to sing in this way. Dylan shares these positive attitudes, and a concomitant negative attitude towards 'blatant' Americanisms in singing, but says that it is difficult to sing with NZE vowels. Andrew provides an interesting point of comparison, expressing no concerns about using AmE in song. As the results are presented, it is of interest to determine not only whether the singers' speech and singing differ, but also whether their pronunciation data aligns with the sentiments expressed in the interviews.

The dataset used for analysis will be described in this introductory section, followed in section 4.1 by a discussion of the decisions made regarding the statistical analysis. Sections 4.2 and 4.3 present a thorough description of the paired data – those words that were both sung and recited. While addressing this paired data, particular tokens of interest will be discussed. The full vowel space for each singer will be presented in section 4.4, based on averages of all of the data collected; tokens from the interviews will also be examined. Both of these analyses are worth pursuing – the paired design benefits from the fact that words are matched for surrounding environment and lexical effects such as word frequency, while the analysis of the full dataset benefits from a higher token count. Section 4.5 considers the differences between singing and speech in terms of pitch and duration of vowels, and provides an in-depth enquiry into the relationship between pitch and F1. The chapter concludes with a brief analysis of two consonant variables in section 4.6. To allow the reader to become acquainted with the data under analysis, audio files of the songs, in both their sung and recited form, are included on the CD attached to the thesis. Fully produced versions of the songs can also be heard, except for John's 'Ballad of Archibald Baxter', for which a final mix could not be obtained. See Appendix E for the tracklisting of this CD.

Table 4.1: *Number of Tokens Found for each Speaker for each Vowel in each Condition*

	Dylan			Andrew			John			
	Sing	Rec	Int	Sing	Rec	Int	Sing	Rec	Int	Total
DRESS	8	8	10	9	10	12	13	13	8	91
TRAP	14	10	12	11	9	15	14	15	4	104
THOUGHT	14	14	12	12	11	15	16	13	8	115
LOT	8	8	8	3	4	4	27	18	10	90
START	1	1	2	3	3	6	4	4	6	30
GOOSE	4	2	2	2	2	3	3	2	5	25
GOAT	17	15	9	16	13	9	8	10	13	110
PRICE	50	48	0	10	10	0	43	39	0	200
Total	116	106	55	66	62	64	128	114	54	765

The full dataset of measurements, after having excluded the items discussed in 3.6,[1] includes a total of 765 tokens. There are a total of 264 pairs of sung/recited words in the dataset, made up of 95 pairs for Dylan, 58 for Andrew, and 111 for John. There are also 173 interview tokens, and a further 64 tokens that were sung or recited but were not in a pair. Several sung or recited tokens ended up without a pair partner because one item in the pair was excluded from the analysis. Most commonly, this was because the recited word was unstressed and pronounced with a reduced vowel, while the sung word used an unreduced vowel realisation. Table 4.1 shows how the tokens are distributed across the three singers, for each of the eight vowels, in the three conditions: singing, reciting, and interview.[2] Note that the number of pairs for each singer, for each vowel, is not shown in this table but included below with the results of the paired tests (in Table 4.2). The variables occurred at very different frequencies in the songs, leading to a large

[1] Excluded tokens included: GOOSE, GOAT, and PRICE tokens that preceded non-front vowels; GOOSE tokens that were preceded by /j/; and unstressed words realised with a reduced vowel.

[2] A note on terminology: I use the word 'condition' in the sense that there are three experimental conditions in this study. By extension, the term will also be used occasionally to distinguish between singing and speech more generally.

number of tokens for TRAP, THOUGHT, GOAT, and particularly PRICE, and very few tokens for START and GOOSE.

4.1. Statistical Analysis

The software package "R" (R Development Core Team, 2009) was used for all graphing and statistical purposes. After exploratory analysis of the data, it was found that it would not be appropriate to perform a multiple regression analysis on the data due to multi-collinearity issues. For example, the independent variables pitch and condition (singing versus reciting) are correlated both with F1 and with each other. A model trying to account for variation in F1 would not be able to separate the effects of these variables.

A simpler option would be to do a t-test comparing the mean formant values in singing versus speech. However, the assumptions of the t-test are not met by this data. The data do not come from a normal distribution, being strongly right-skewed, and there is consistently greater variance in singing than speaking, meaning that the assumption of equality of variance is not met. These issues are exacerbated by the fact that the token count for most tests is less than 30 observations. This means that t-tests could lead to misleading results. Wilcoxon tests provide a non-parametric alternative, for which the centre of the data is treated as a median rather than a mean.

For the paired data, Wilcoxon matched pair signed-rank tests[3] were used to test whether the median of the differences between paired items is significantly different to zero. For the unpaired data, a Wilcoxon rank sum test (also known as the Mann-Whitney test) was used. This test ranks all the values in the combined sample, and then determines whether the rankings are significantly different across the two samples, singing and speaking. For the matched pairs test, it is only possible to achieve a significant result if there are at least six pairs, so tests were not conducted on START and GOOSE for any of the singers, or on LOT for Andrew.

A test of F1 and F2 was done for all point measurements, including the *a* and *b* measurement points for the diphthongs. For diphthongs, an additional measurement was made to determine the degree and direction of the vowel's movement. This was

[3] Two-tailed tests are used throughout the analysis because there is no a priori reason to assume that one condition should have a higher or lower value than the other. If one-tailed tests had have been used, all cases where $p<.1$ would have become significant at the .05 level.

measured by subtracting the point *a* measurement from the point *b* measurement, for both F1 and F2. This set of six measurements for a given diphthong allows the statistical testing of both the vowel's position and the vowel's movement.

4.2. General Comparison of Singing/Reciting Pairs

The results of the paired Wilcoxon tests are summarised in Table 4.2, which shows the median of differences, in Hertz, between pairs of sung and recited vowels (with the number of pairs in brackets) for each vowel, for each singer. There are significant differences for all three singers, on all vowels where the token counts were high enough to perform a test. The vowels vary as to whether the differences occur on measurements of F1, F2, or diphthong movement.

In reading the table, it should be noted that the difference being measured is singing minus speaking, so a positive value in F1 difference means that singing was opener than reciting, while a negative value means singing had a closer vowel than reciting. A positive value in F2 difference means that singing had a greater F2, and was thus fronter than reciting; a negative F2 difference, conversely, means that the recited vowels were fronter than the sung ones. For ease of reading, p-value has been represented as *** for $p<.001$, ** for $p<.01$ and * for $p<.05$. A near significant result ($p<.1$) is represented with a full stop. Non-significant results are represented with NS and cases where there are not enough pairs to warrant the test are marked with NA. Note that Table 4.2 does not provide the actual formant values for the vowels, but rather the median of differences between paired items. Averaged formant values for each vowel will be reported in section 4.4. For the rows showing differences in F1 and F2 movement, the phonetic interpretation of the differences is not straight-forward; the reader is advised to refer to the relevant graphs to get a feel for the directionality of the differences. The table is provided largely for reference rather than for discussion.

Table 4.2: *Median of Differences (Hz) between Singing and Reciting Pairs (sing-recite) for F1 and F2 of each Vowel for each Singer, with Results of Paired Wilcoxon Signed Rank Tests. Number of Pairs Shown in Brackets next to F1 Difference*

	F1						F2					
	Dylan		Andrew		John		Dylan		Andrew		John	
	diff. (n)	p	diff. (n)	p	diff. (n)	p	diff.	p	diff.	p	diff.	p
DRESS	313 (8)	**	382 (9)	**	112 (13)	**	51	NS	33	NS	-287	**
TRAP	444 (10)	**	281 (9)	**	75 (14)	**	28	NS	29	NS	-104	**
THOUGHT	356 (12)	***	466 (11)	***	172 (13)	***	288	***	482	***	262	***
LOT	308 (8)	**	323 (3)	NA	76 (18)	*	91	NS	200	NA	307	***
START	159 (1)	NA	186 (3)	NA	-84 (4)	NA	-121	NA	-180	NA	-80	NA
GOOSE (a)	-25 (2)	NA	-60 (2)	NA	-22 (2)	NA	77	NA	824	NA	-19	NA
GOOSE (b)	-23	NA	-5	NA	76	NA	-468	NA	-462	NA	-723	NA
GOOSE (b-a)	2	NA	55	NA	98	NA	-545	NA	-1285	NA	-704	NA
GOAT (a)	-8 (11)	NS	56 (12)	NS	-95 (8)	.	13	NS	20	NS	-258	**
GOAT (b)	10	NS	16	NS	-22	NS	-262	**	-376	***	-642	**
GOAT (b-a)	20	NS	-37	NS	83	.	-255	***	-408	***	-454	*
PRICE (a)	351 (43)	***	245 (9)	**	96 (39)	***	318	***	421	**	267	***
PRICE (b)	222	***	37	NS	57	*	262	***	339	.	106	**
PRICE (b-a)	-127	***	-193	*	-28	NS	-63	**	-53	NS	-178	***

(*** for p<.001; ** for p<.01; * for p<.05; . for p<.1; NS for not significant; NA for cases where there are too few pairs to warrant a test for significance)

4.3. Detailed Comparison of Singing/Reciting Pairs

To illustrate the results for the paired items in detail, 24 graphs are presented in this section, one for each of the three speakers for each of the eight vowels. The F1/F2 space is graphed with both axes inverted, so that the space has the orientation of the vowel quadrilateral. It should be noted that each graph is 'zoomed in' to the data, so the graphs are all on different scales. This means that the reader should take care to look at the axes of each graph before comparing between them. The reason for this presentation choice was that it allowed the labelling of individual data points on each graph.

The monophthongs are marked with an F1/F2 point and the diphthongs are represented with arrows. The labels for singing tokens are in bold, with a closed circle for monophthongs, and a thick line for diphthongs, while the labels for reciting data are in plain text, with open circles and thin lines. For the diphthongs, the start of the arrow represents the first measurement in the vowel (point *a*), and the tip of the arrow represents the second measurement (point *b*). The label for each of the diphthongs is just next to the tip of the arrow.

For words where the vowel under analysis occurs at the beginning or end of the word, the neighbouring word is shown on the label to provide information about phonological environment of the vowel. The symbol $ is used in cases where a vowel is preceded or followed by silence. Where there are multiple pairs of the same word in the same context, each pair is given a number so that they can be referred to in the text where required, and so that they can be matched with their partner.

4.3.1. DRESS

Dylan's eight DRESS pairs (Figure 4.1) provide a good example of the kind of consistent differences found between singing and reciting for many of the vowels in the dataset. The sung vowels have consistently more open realisations (represented by higher F1). There is some variability in terms of F2 in both conditions, and in some cases these differences are matched across pairs, as in *heavy*. Note that the spread in terms of F1 is greater for the sung words. This may be caused by their varying pitches, as will be discussed in 4.5.2. The results for Andrew (Figure 4.2) also show this greater F1 variability in singing than reciting. Note that there are three falsetto tokens of DRESS on *when*, which have a higher than average pitch of around 330Hz, and have a relatively low F1; however, this difference could also be caused by the fact that *when* is a function word (justification for the inclusion of function words in the main analysis was provided in section 3.6, above).

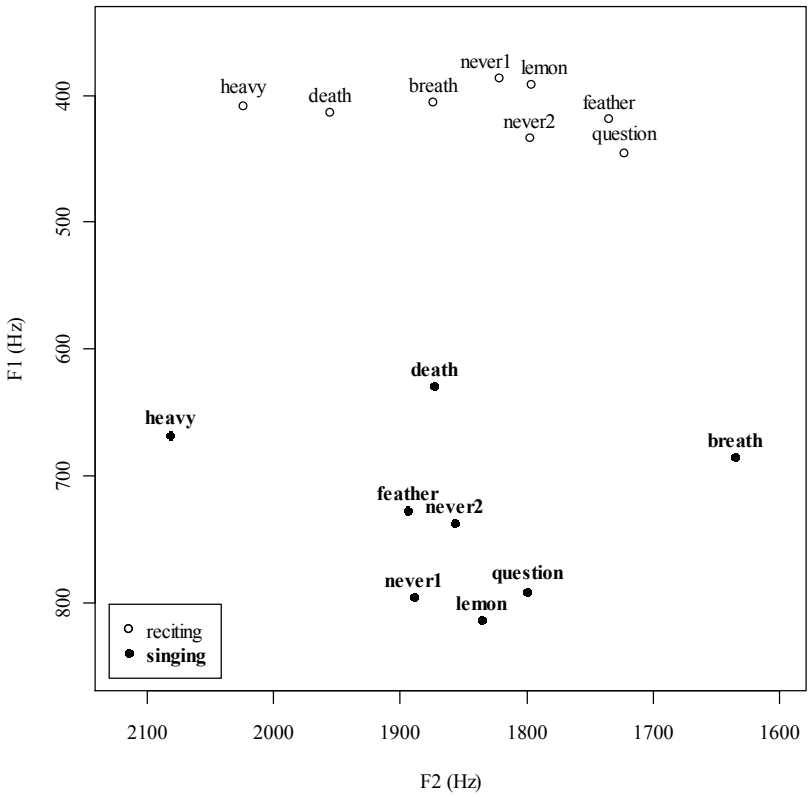

Figure 4.1: Dylan DRESS pairs

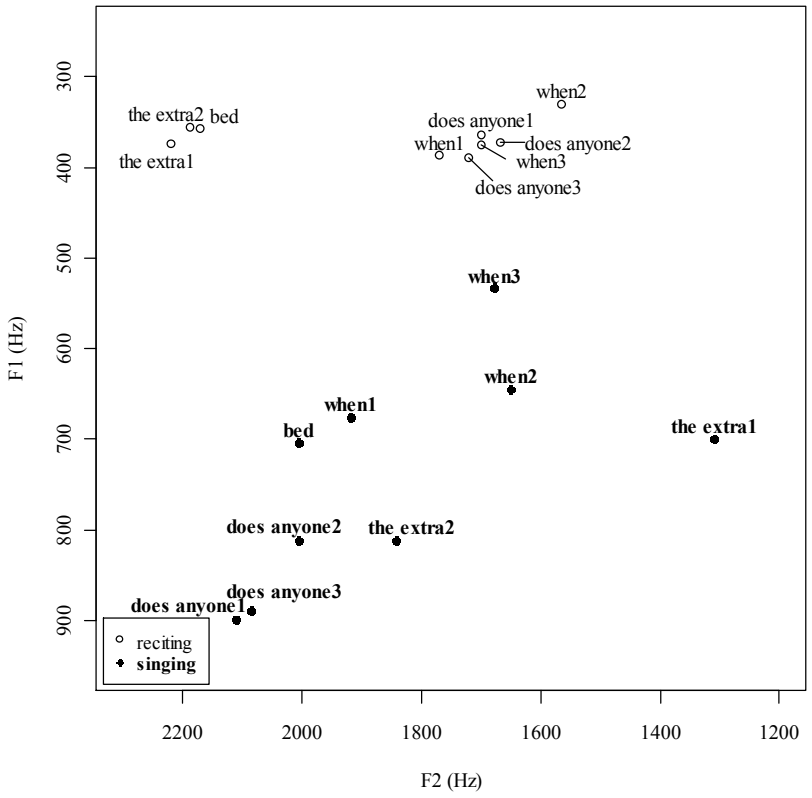

Figure 4.2: Andrew DRESS pairs

Considering the results for DRESS, then, it can be seen that the vowel is significantly opener in singing than reciting for all three singers ($p<.01$ for all singers). The magnitude of this difference, however, is far less for John than it is for the others, the F1 is on average 112Hz higher in singing than reciting for John, and around 350Hz higher for Dylan and Andrew. Additionally, John's recited DRESS vowels are significantly fronter than their sung counterparts ($p<.01$), a difference which does not occur for the other singers.

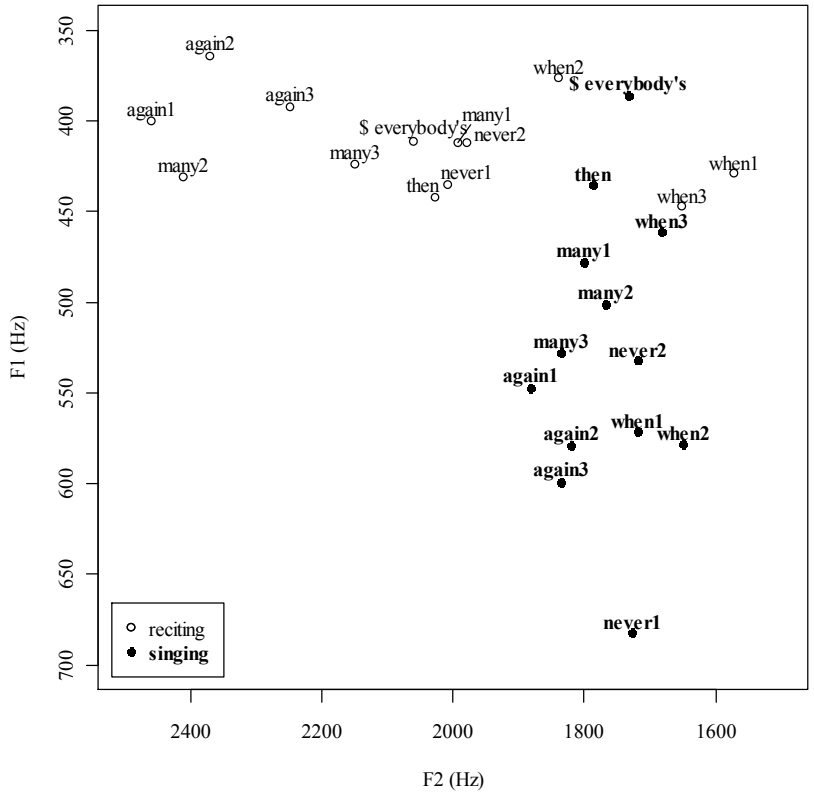

Figure 4.3: John DRESS pairs

4.3.2. TRAP

The results for TRAP (Figures 4.4-4.6) are very similar to those for DRESS, with significantly opener vowels in singing than reciting for all three singers ($p<.01$ for all singers), and with the magnitude of this difference being less for John than for the others. John's TRAP, like his DRESS, is fronter in reciting than in singing ($p<.01$).

At this point, it is worth noting the way stress is realised differently in singing and speech. In speech, unstressed syllables are often reduced to a schwa-like vowel quality. In the data for these singers, vowel reduction was more common in reciting than singing. This may be because on average the sung vowels are longer. The sense of prosody is maintained in singing by placing stressed syllables on prominent beats, and

unstressed syllables on off-beats or less prominent beats, while maintaining the full quality of the vowel. Examples of this are particularly found in function words such as *and*. For example, several of Dylan's TRAP vowels (e.g. *plastic* <u>*and*</u>, *$* <u>*and*</u>, <u>*than*</u>, *heavy* <u>*as*</u>) were realised as [æ] in singing and [ə] in speech. Since the reduced tokens were not included in the main analysis, these pairs do not appear in Figure 4.4, though two cases where the vowel in *and* was deemed as receiving a fully realised vowel in both singing and reciting can be seen.

In terms of the role of pitch on F1, note Dylan's *wrap* pair. The sung and recited words have a pitch of 199Hz and 155Hz, respectively, which is a smaller pitch difference between the conditions than occurs in most paired items. It is of note that the F1 distinction is still very clear in this case. The relationship between pitch and F1 will be discussed in detail in section 4.5.2.

In John's TRAP pairs (Figure 4.6), there are a group of tokens of the word *that* (*that3* through to *that8*) which are overlapping across singing and reciting. These are from the song 'Tell that Boy'. *that1* and *that2* are from different songs, 'Wintersun' and 'Ballad of Archibald Baxter', respectively. Note that while *that2-8* are all demonstratives, *that1* is a complementiser.

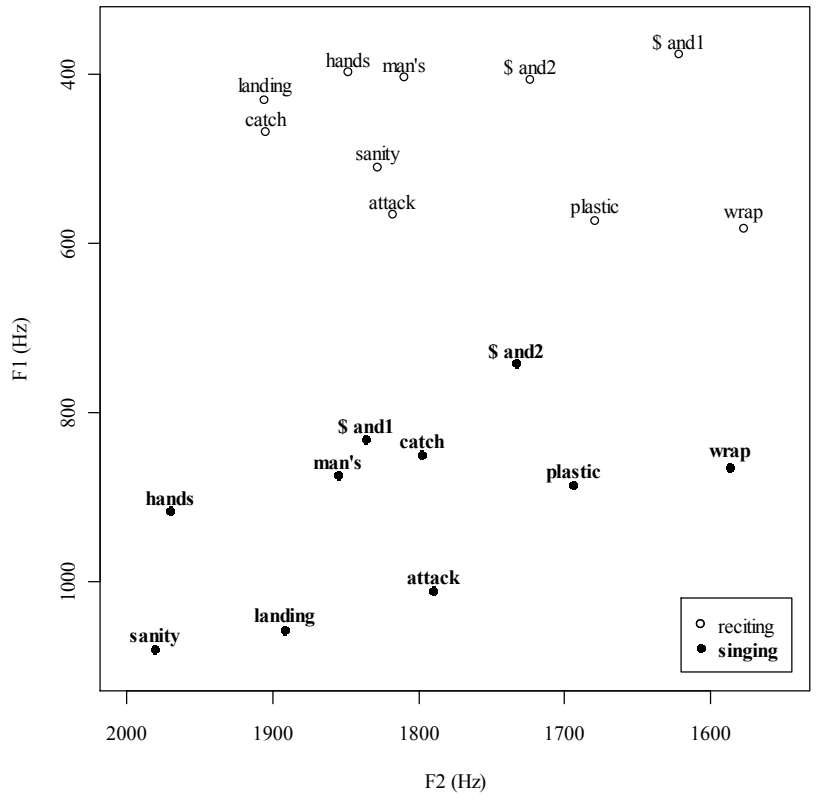

Figure 4.4: Dylan TRAP pairs

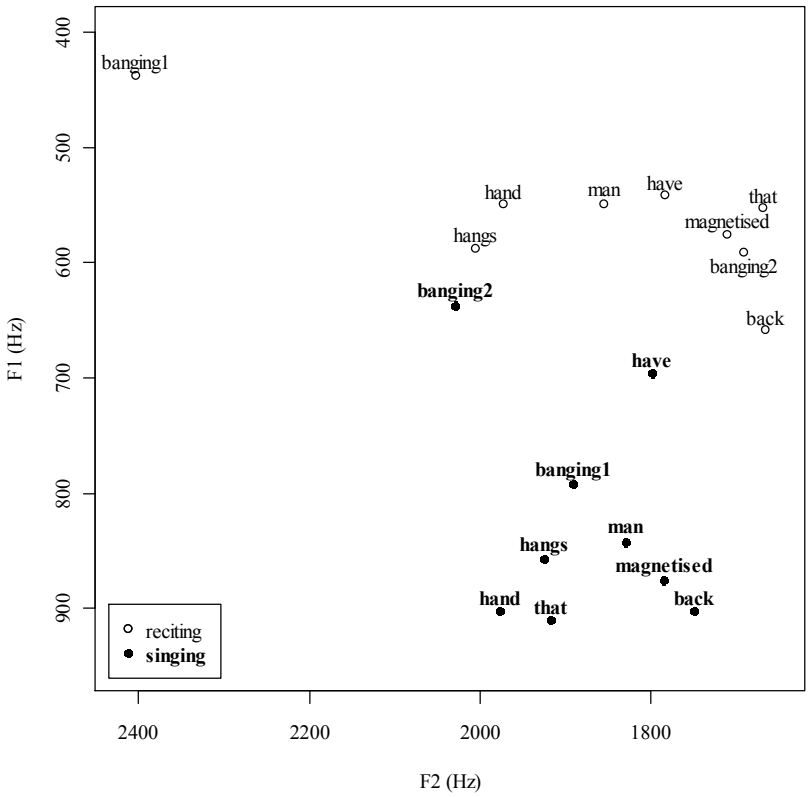

Figure 4.5: Andrew TRAP pairs

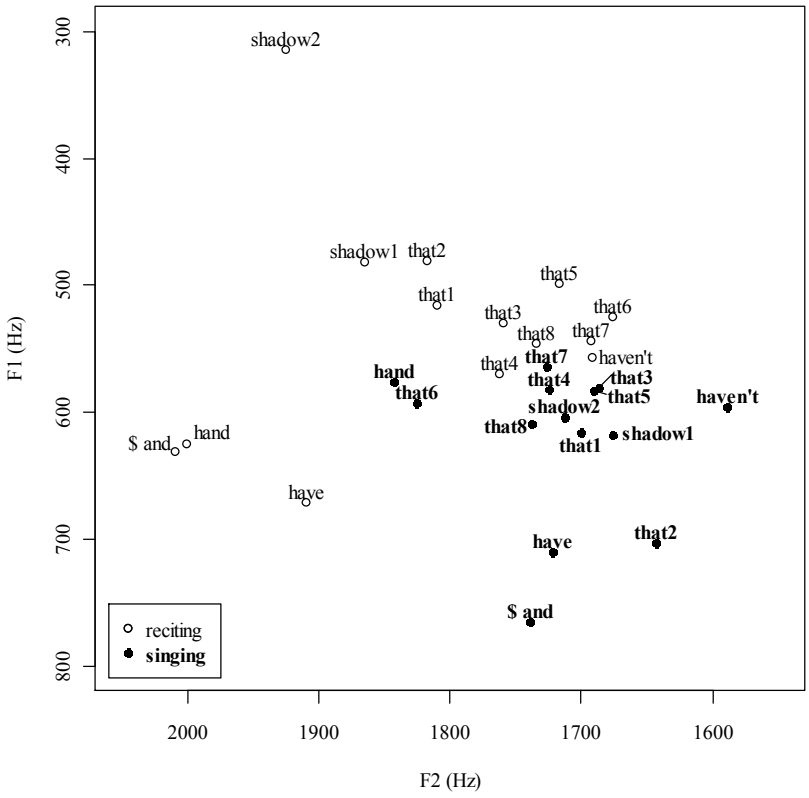

Figure 4.6: John TRAP pairs

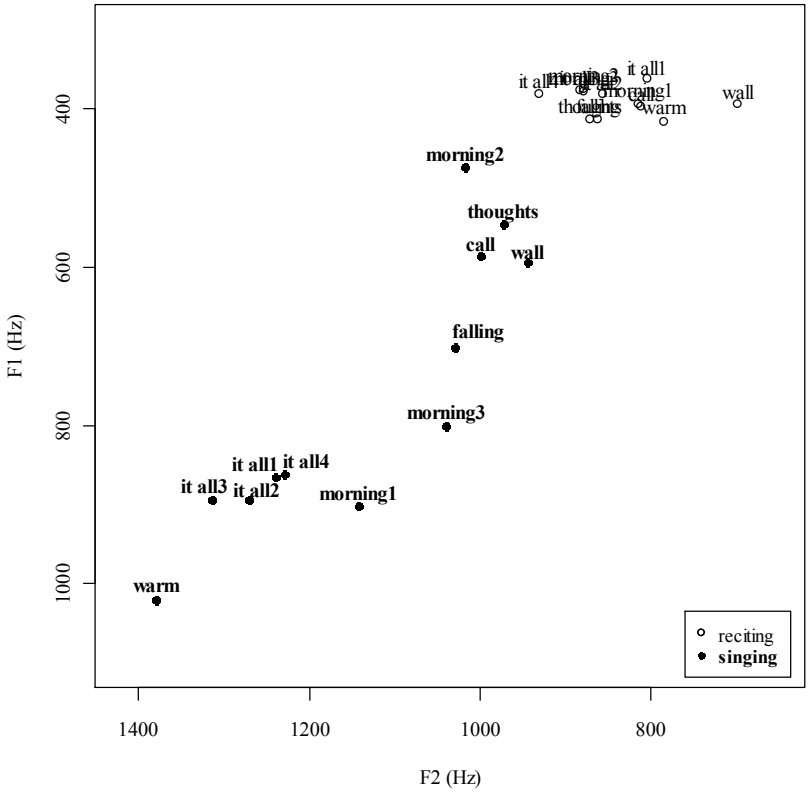

Figure 4.7: Dylan THOUGHT pairs

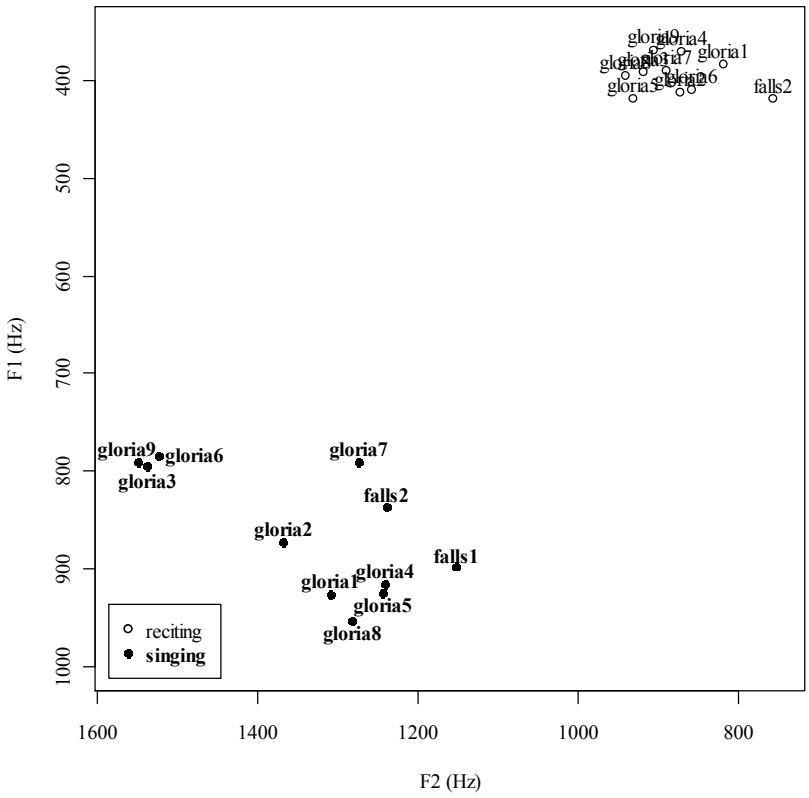

Figure 4.8: Andrew THOUGHT pairs

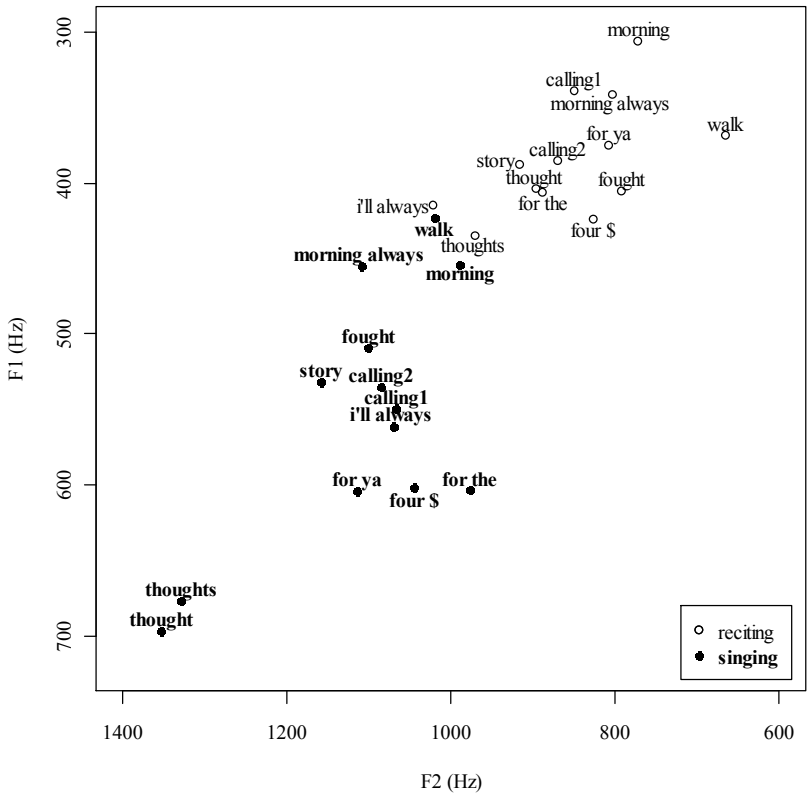

Figure 4.9: John THOUGHT pairs

4.3.3. THOUGHT

THOUGHT is opener and fronter in singing than speaking (Figures 4.7-4.9). This difference is highly significant ($p<.001$) for F1 and F2 for all singers. Again, the size of the F1 difference is much smaller for John (172Hz) than it is for Dylan and Andrew (averaging around 400Hz). The loss of lip rounding is impressionistically apparent in singing as opposed to speech, however this cannot be reliably separated from the fronting element of F2 shown in this acoustic data. There seems to be an overall tendency that higher F2 (fronting or rounding) and higher F1 (opening) go hand in hand. This is especially noticeable for Dylan and John.

In Dylan's THOUGHT vowels (Figure 4.7), we see examples of tokens repeated in the same musical context behaving in different ways. The four repetitions of THOUGHT in the phrase *it all* (which occurs twice per chorus), cluster very closely together. This sort of stability is what previous research has assumed occurs across repeated sections. This contrasts with repetitions of the word *morning.* The two tokens *morning2* and *morning3* occur in the repeated bridge section of the song, and receive very different vowel qualities. Previous studies have excluded multiple repetitions of sections like choruses, but this example suggests that the pronunciation of a word can vary even when it is

repeated in the same context. Notice that overall, the amount of variability in singing is much greater than is seen in reciting.

In Figure 4.8, the large distance between the spoken and recited pairs of Andrew's THOUGHT vowels is noticeable. Also, the tokens cluster relatively closely together within each condition. Note that the three tokens of *Gloria* which are offset from the rest of the group were from the same line as the other falsetto tokens mentioned above, having a much higher than average pitch.

4.3.4. LOT

LOT, like THOUGHT, tends to be opener and fronter in singing than reciting (Figures 4.10-4.12). There were insufficient tokens to test for significance for Andrew. For Dylan, the sung LOT vowels are significantly opener ($p<.01$), but not fronter. For John, the sung tokens are significantly opener ($p<.05$) and fronter ($p<.001$). As with all the other vowels discussed so far, John's F1 difference is small (76Hz higher in singing than reciting) when compared to Dylan and Andrew (differences of around 315Hz).

Note that Dylan's two paired instances of the phrase *rely on* go against the trend, with the vowel in *on* being fronter in reciting than singing. If these two pairs are removed, there is a significant F2 difference between conditions, with sung tokens being fronter (and/or less rounded) than recited tokens. The different pattern seen for these pairs seems to be caused by the insertion of a /j/ glide in the recited tokens, causing assimilation, which is not present in the sung versions. Another interesting LOT pair for Dylan is in the word *follow*, which has a smaller difference between the sung and spoken vowels than is seen in other pairs. The sung token is somewhat lip-rounded, causing it to sound noticeably NZ-like.

For John's LOT vowels, there is a lot of inconsistency in the acoustics of the word *on* in the phrase *go on*. This could be caused by a varying degree of assimilation to the offglide of the preceding GOAT vowel in the word *go*.

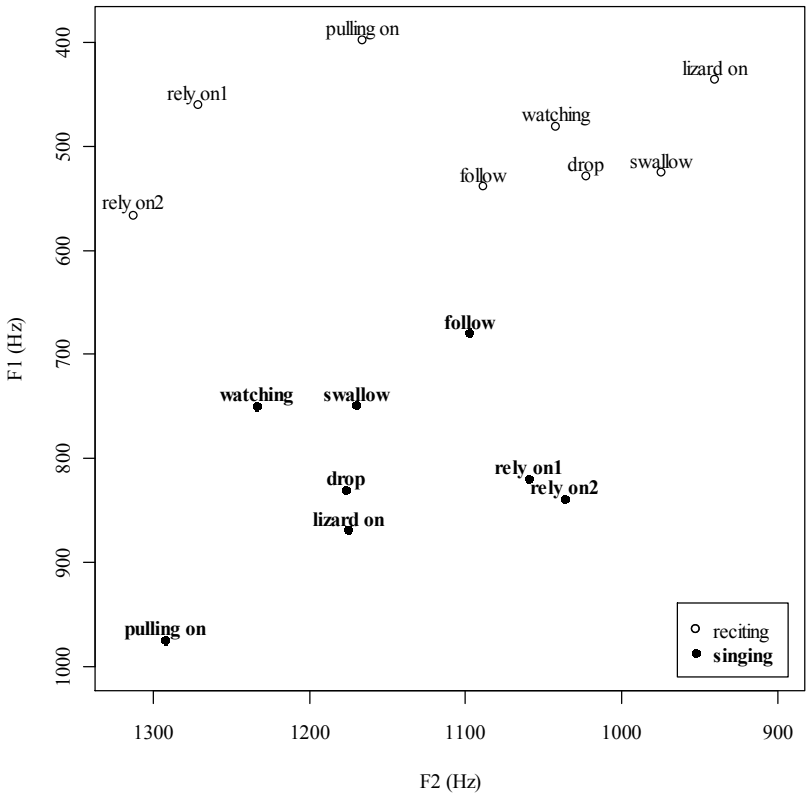

Figure 4.10: Dylan LOT pairs

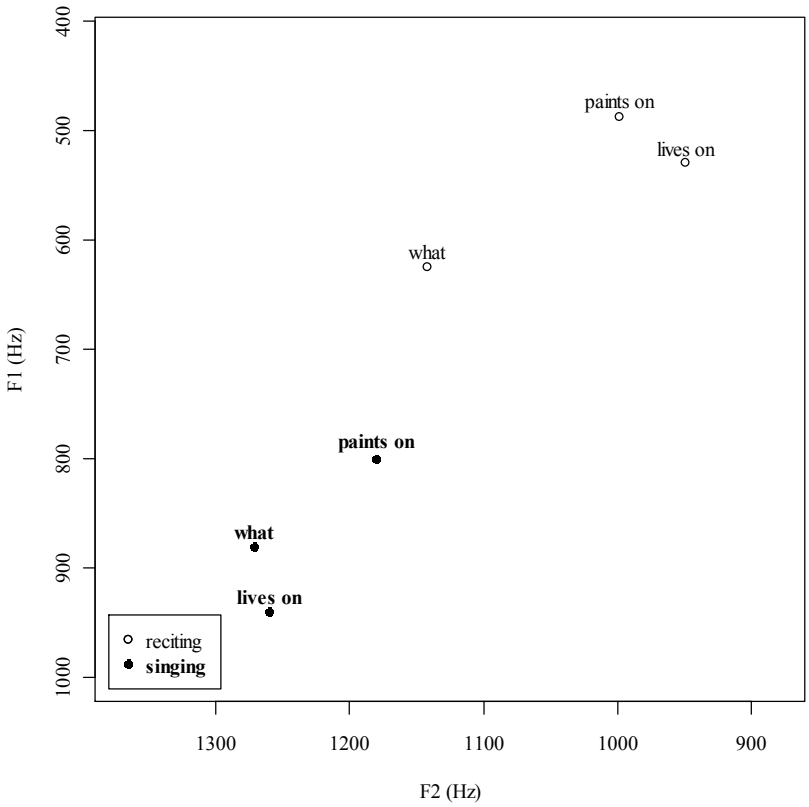

Figure 4.11: Andrew LOT pairs

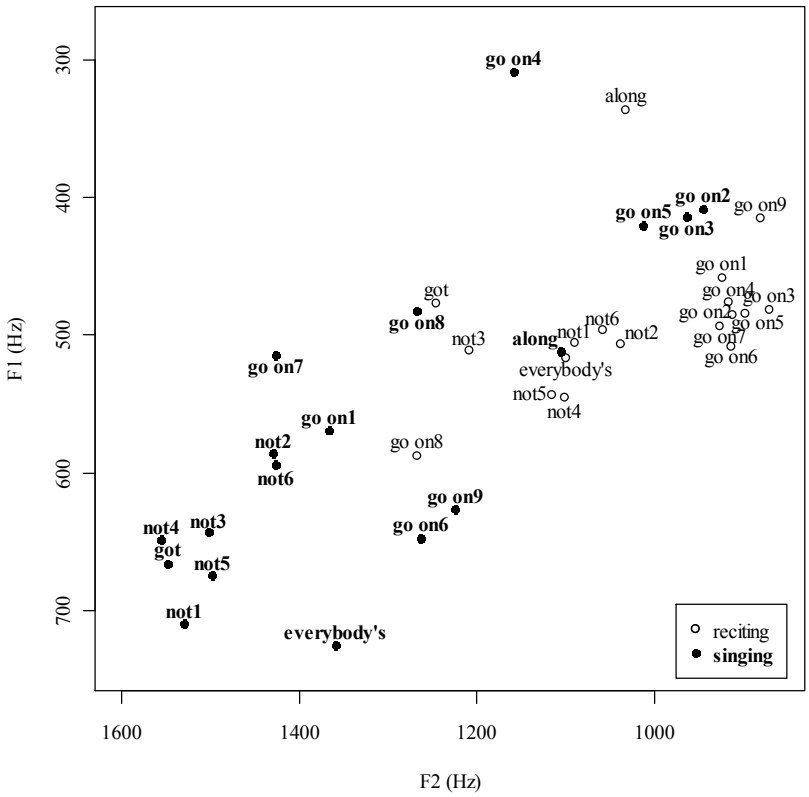

Figure 4.12: John's LOT pairs

"gonna"

Figure 4.13 presents all of the instances of the word *gonna* which occurred in John's song 'Ballad of Archibald Baxter'. These vowels are not included elsewhere in the analysis due to their different behaviour. The relative position of the sung and spoken pronunciations of the word *gonna* is reversed compared to the LOT vowels, with the sung tokens being closer and backer than their spoken counterparts. This different behaviour is not surprising since *gonna* has a somewhat unusual status as a contraction of *going to*. My a priori assumption that the vowel in *gonna* is actually LOT was, perhaps, not well justified, perhaps it is more like STRUT.

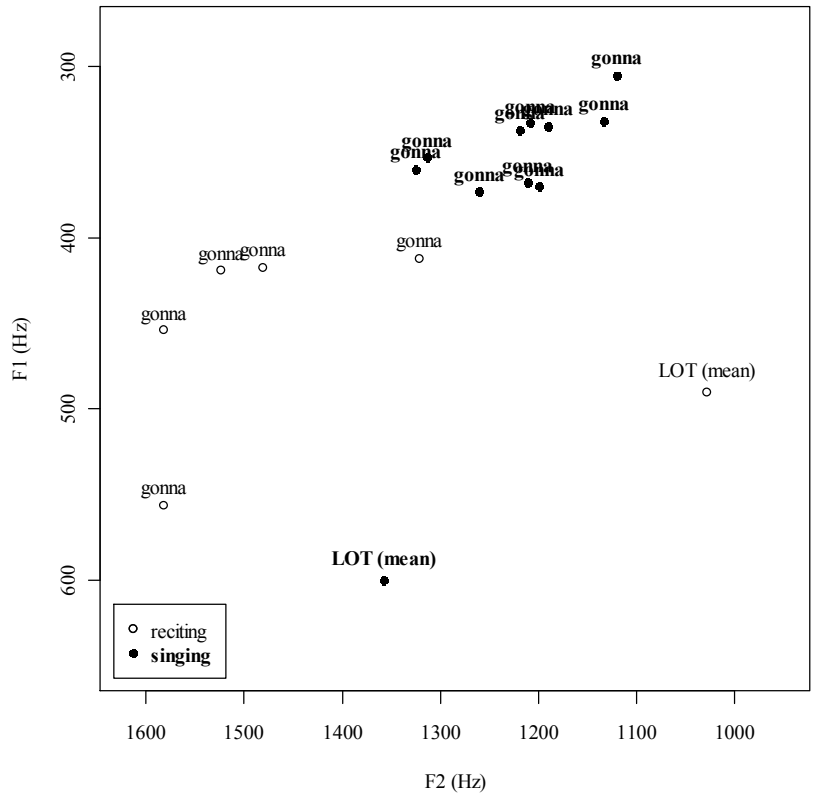

Figure 4.13: John's *gonna* tokens, with mean values of all LOT pairs.

4.3.5. START

START differs from THOUGHT and LOT in that the spoken vowel is fronter than the sung vowel (Figures 4.14-4.16). Though the raw data is in this direction for all three singers, there are insufficient tokens to warrant a Wilcoxon test for any of them. For Andrew and Dylan, START is opener in singing than speech, while for John the opposite pattern is found, with vowels closer in singing than speech. Note that John's pronunciation of the word *ask* is START-like, rather than TRAP-like.

Andrew's two tokens of the word *car* vary greatly in length; *car $1* is 313ms, much shorter than *car $2*, which is an extended 1694ms. The realisations of these two vowels are fairly similar, with the longer token being slightly opener and fronter. Section 4.5 will discuss the possible role of duration in the differences between singing and speaking; it is important to note cases where a large difference in duration is not associated with a large difference in vowel realisation, as well as noting cases where it is.

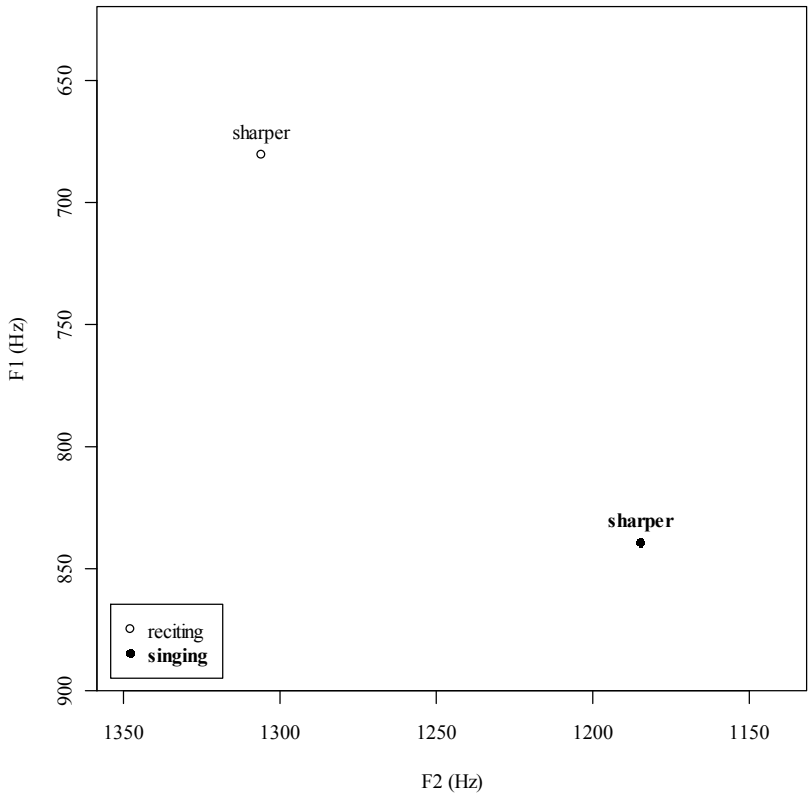

Figure 4.14: Dylan START pair

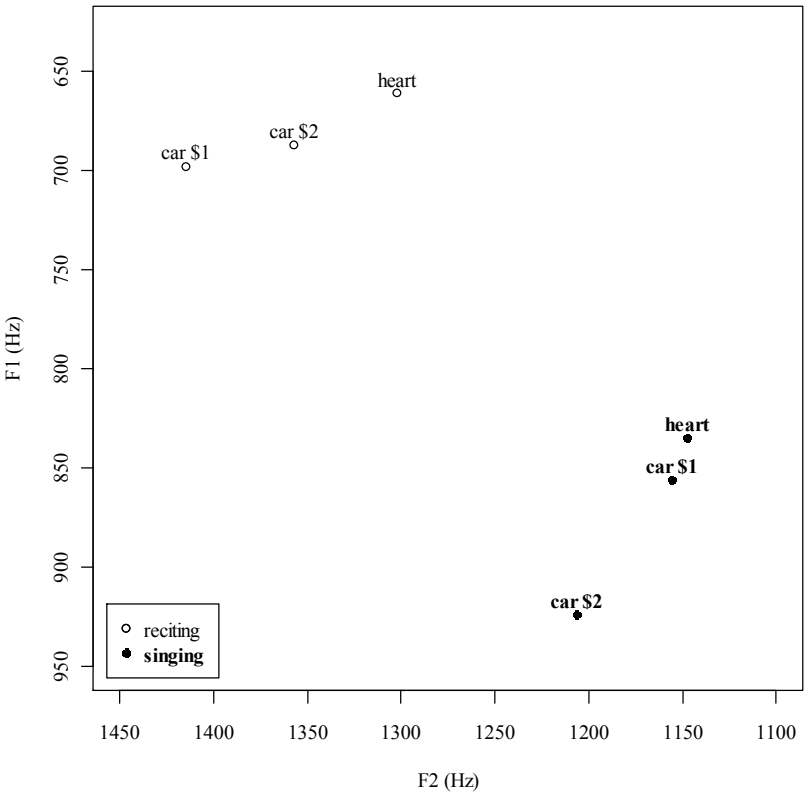

Figure 4.15: Andrew START pairs

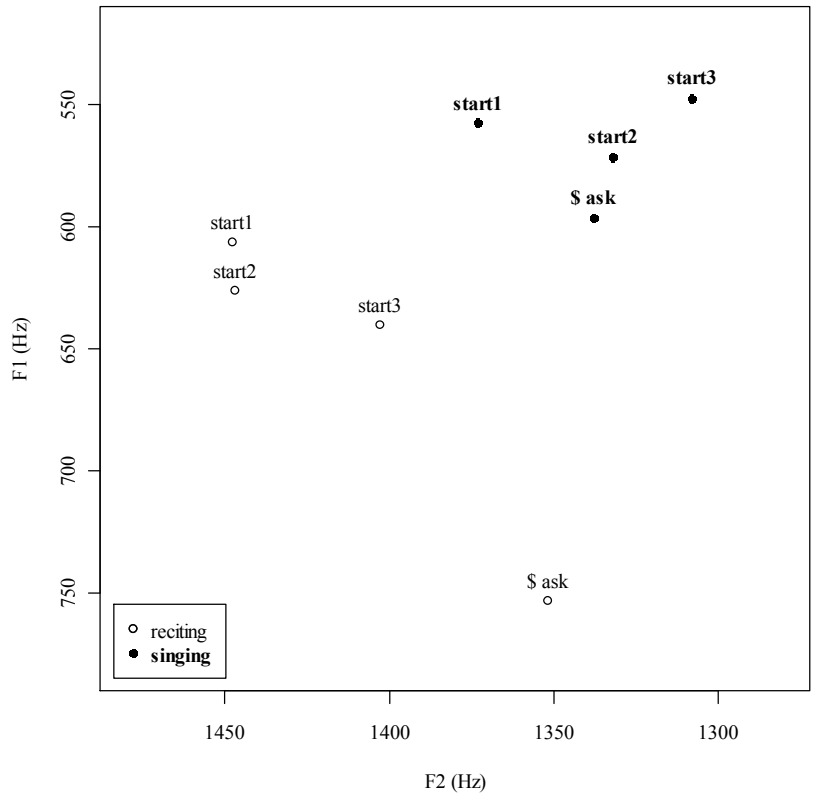

Figure 4.16: John START pairs

4.3.6. GOOSE

As mentioned earlier, **GOOSE** is realised as a diphthong in both speaking and singing. This diphthong moves in opposite directions in the two styles (Figures 4.17-4.19). Recited GOOSE raises and fronts while the sung equivalent retracts, and also raises in most cases. All three singers show this pattern in their two pairs each, though there are insufficient tokens to test for significance.[4] The F2 change between the beginning and end of the vowel in singing also seems to represent a rounding gesture. As well as the dramatic difference between conditions for F2, the spoken vowels start in a more open position and then raise while the sung tokens show less movement on the F1 dimension, with some tokens being raised throughout. It could be reasoned that the vowels in Andrew's two sung tokens of the word *blue* move a long way because of their long duration, however with durations of 231ms for *blue car1* and 636ms for *blue car2*, this does not seem to be the case since *blue car1* actually moves further.

[4] Tests for significance will be carried out on the larger, unpaired, dataset in section 4.4.2 below.

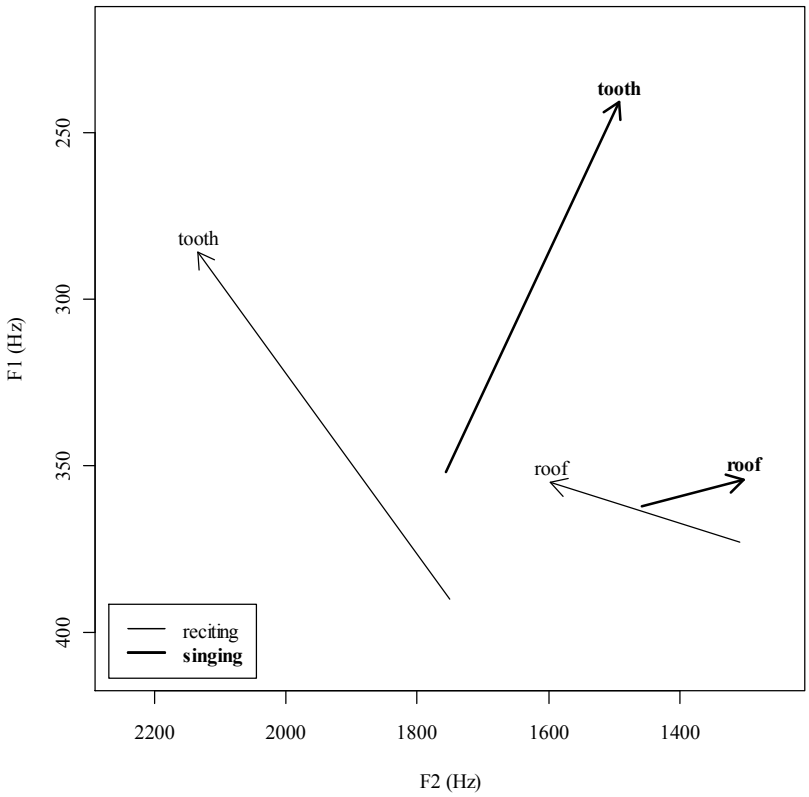

Figure 4.17: Dylan GOOSE pairs

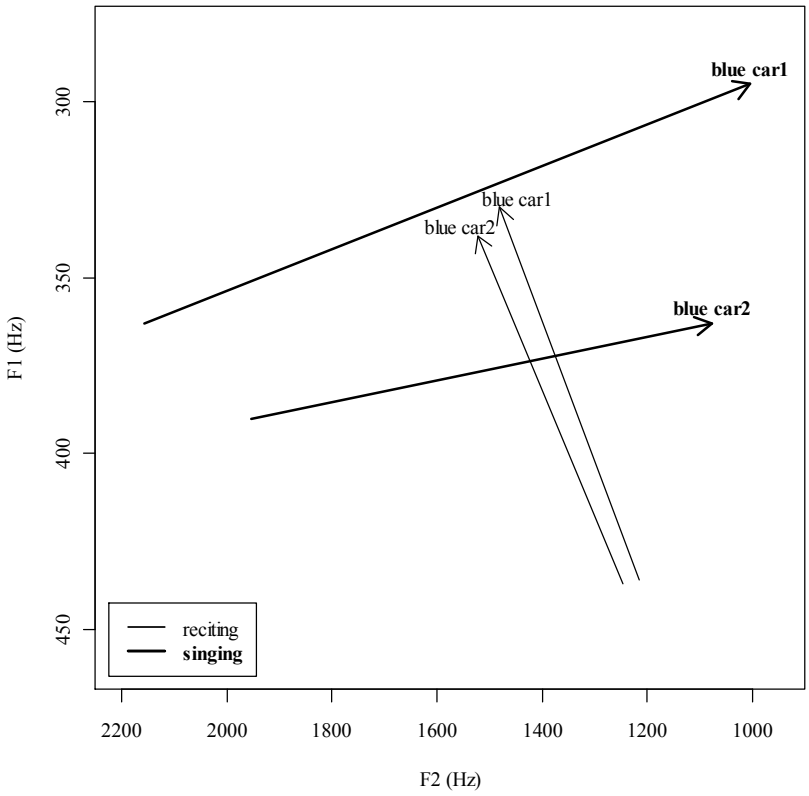

Figure 4.18: Andrew GOOSE pairs

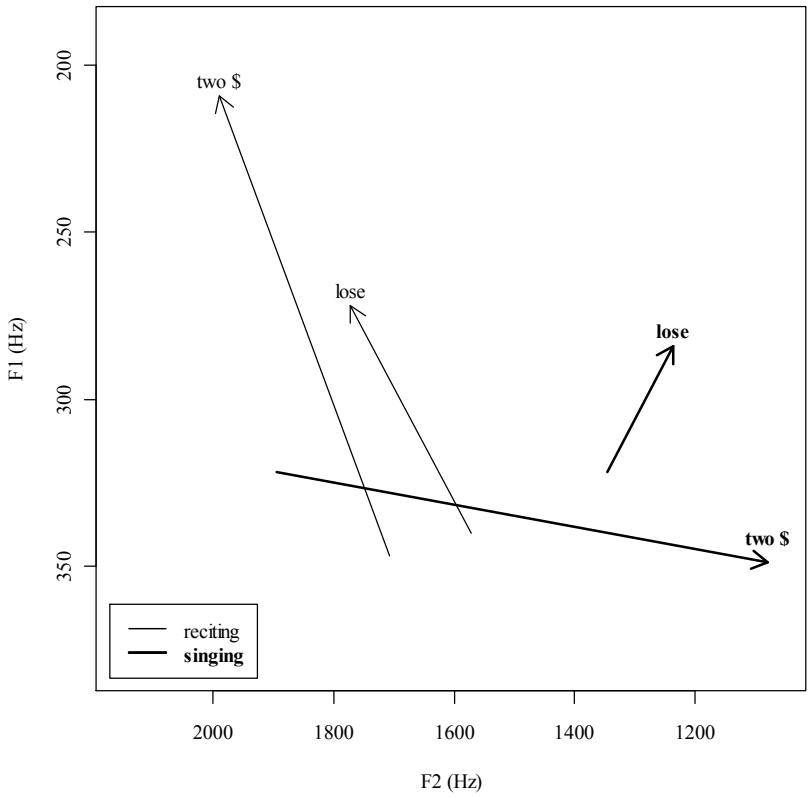

Figure 4.19: John GOOSE pairs

"you"

It should be remembered that this analysis of GOOSE excludes a large portion of the data, following /j/ (notably the word *you*), which shows its own interesting variation across conditions. There were 17 occurrences of the word *you* after excluding those which were followed by /w/. The GOOSE vowel in *you* is monophthongal and fronted in speech, but moves backwards from the /j/ in singing. The sung vowels only move backwards in singing if the vowel is longer than about 100ms. All of the *you* tokens, including those from the interview data, are shown in Figure 4.20. To differentiate between the three singers, each arrow is labelled accordingly. A clear distinction can be seen between singing and speech, despite the effect of the preceding /j/. This result supports the findings for the other GOOSE pairs; sung tokens retract while recited tokens either move to a fronted position, or remain in a fronted position when preceded by /j/. Note Andrew's outlying sung token. This token sounds rather different from the others, something more like *ya* than *you*.

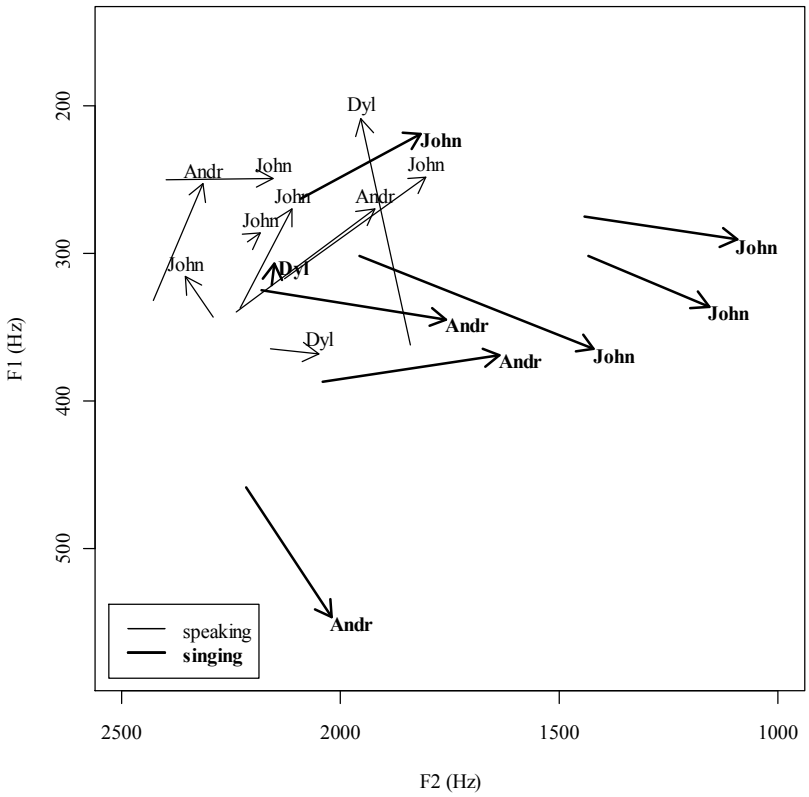

Figure 4.20: All instances (from interviews as well as recited and sung pairs) of the word *you* showing all three singers, with singers' names next to tip of arrow (Dyl=Dylan; Andr=Andrew)

4.3.7. GOAT

GOAT is similar to GOOSE in that the vowels move up and forward in speech and up and backward in singing (Figures 4.21-4.23). This difference in F2 movement is significant for all singers ($p<.001$ for Dylan and Andrew, $p<.05$ for John). For Dylan and Andrew, the GOAT vowel starts in the same position for both singing and speaking (no significant differences), though the opposing direction of the F2 movement is reflected by the fact that the end of the vowel is significantly fronter in speech than singing ($p<.001$ for Andrew and $p<.01$ for Dylan). For John, the whole GOAT vowel is retracted in singing, with significant differences across conditions for both measurements of F2 ($p<.01$ for both F2a and F2b). Interestingly, the nucleus of John's GOAT vowel is raised in singing as compared to speech, though the difference does not quite reach significance ($p=0.0547$). The lowering of F2 in singing could also be caused by an increase in lip rounding for the offglide segment of the vowel.

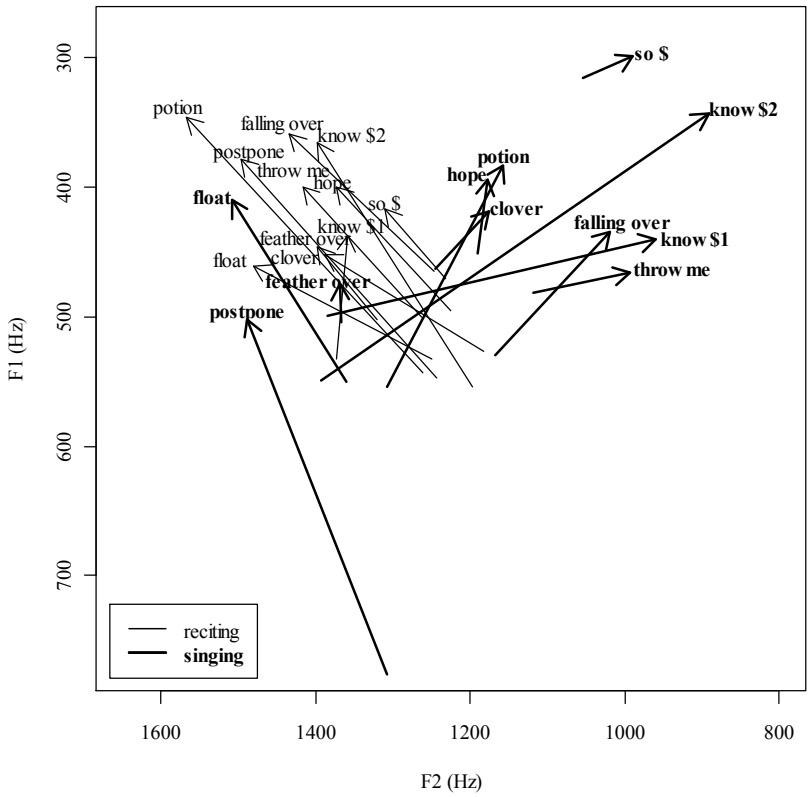

Figure 4.21: Dylan GOAT pairs

Note Dylan's three cases of sung GOAT which go against the trend by fronting (*postpone*, *float*, and *over*), and three more which are different in that they raise slightly but move back only a little (*hope*, *potion*, and *clover*). These vowels are salient and distinctive in the fact that they align relatively closely with spoken realisations of GOAT. These tokens are very different to the results for GOAT in the chorus of Dylan's song, for example in the word *so*,[5] which is very long (1200ms) and is monophthongal, staying in a stable raised back position.

[5] There are many other tokens of *so* which are similar to the one shown, but they are not paired due to the fact that the sung tokens were followed by silence and the recited tokens were run into the following word in all cases but the one recited token shown.

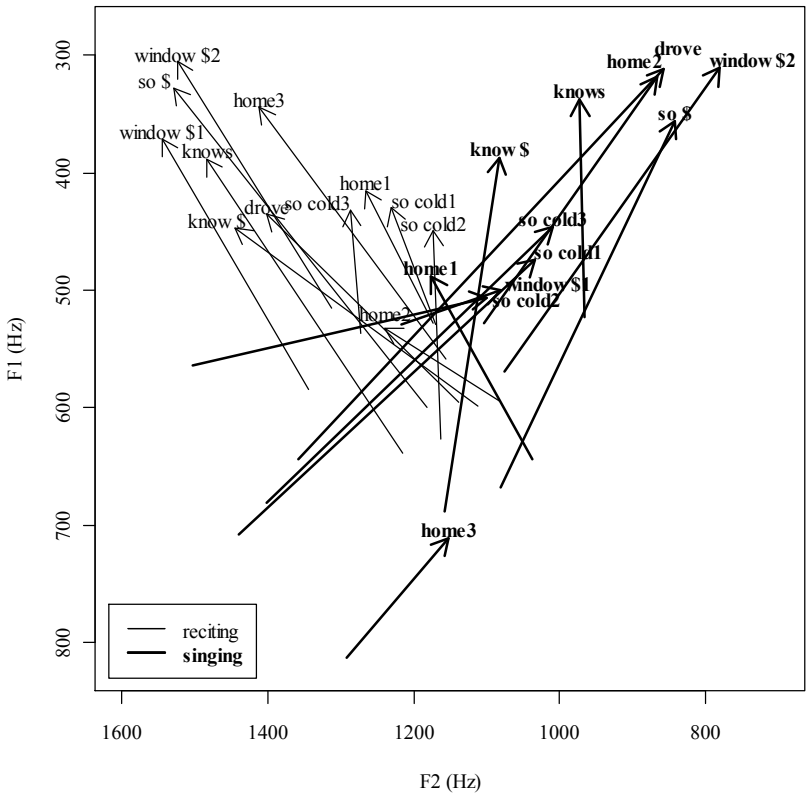

Figure 4.22: Andrew GOAT pairs

It is interesting to compare Dylan's *know $2* with Andrew's *know $*. Both tokens are long (901ms and 1749ms respectively), but Dylan's long token is sung over two notes, dropping by a perfect 5th interval. For Andrew, the diphthong, which has steady pitch, gradually glides from point *a* to point *b*. For Dylan, on the other hand, there are two steady states, with the nucleus and offglide being clearly separated at the moment of the pitch change. The recited pair for this token is a gliding token for Dylan, suggesting that the reason for the sudden change in formants is the sudden change in pitch. This is a pattern seen throughout Dylan's song for cases of vowels sustained over a pitch change, such as the word *I* in the chorus of the song. There is not enough relevant data to analyse this more fully, but the effect of sudden pitch changes on the production of sung vowels should be kept in mind in future studies.

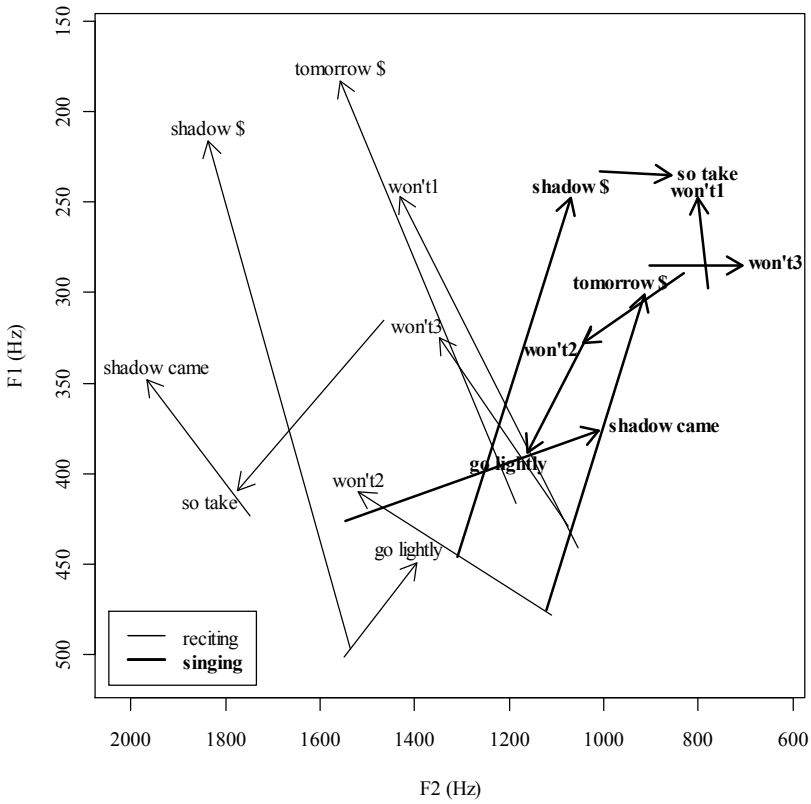

Figure 4.23: John GOAT pairs

As mentioned in the profile for Dylan in 3.4.1, he is particularly conscious of his pronunciation of GOAT, and actually provided a series of performative examples in the interview. Figure 4.24 shows the realisations of all of these demonstrations, with a descriptive label above the head of the arrow for each token. Only those cases where Dylan was overtly displaying the GOAT vowel as part of his discussion of that vowel are included in the graph. First, he produced a hyper-NZ sung example of *so*, followed by a normal sounding sung American *so*, then a spoken NZ *falling over*, and finally an attempt at a spoken AmE *falling over*. The examples of spoken NZE and sung AmE match well with the data seen above, with the especially raised, retracted and monophthongal GOAT for the AmE sung renditions of the word *so* from his song. Dylan's performance of sung NZE GOAT is a perfectly accurate match with the spoken data, demonstrating the level of awareness and control he has over this variable. Dylan's attempt to perform a spoken AmE GOAT is, however, quite different from the sung examples.

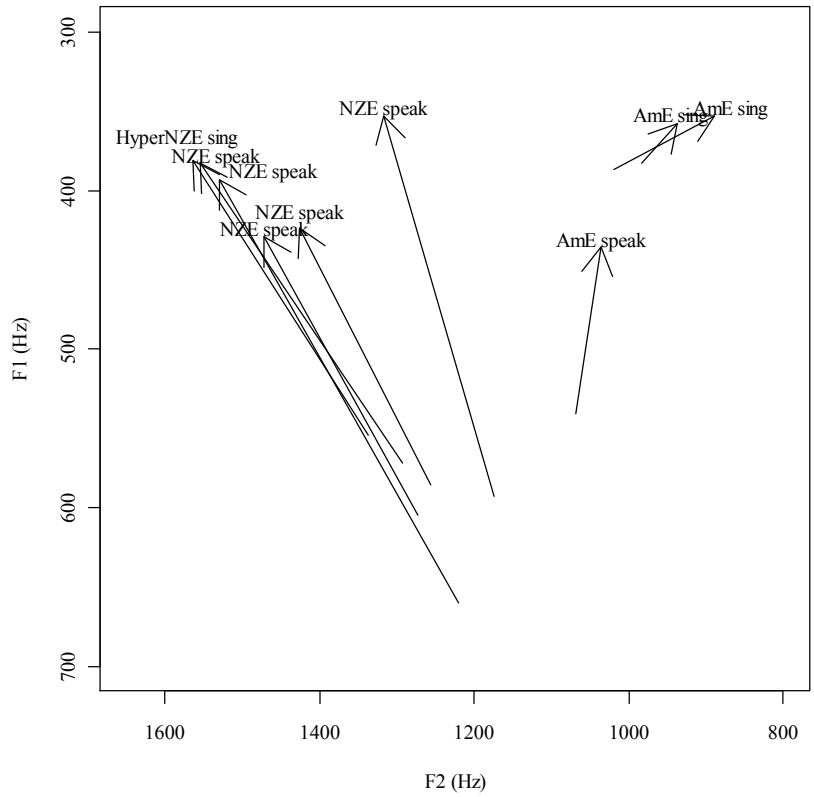

Figure 4.24: All tokens where Dylan demonstrated a GOAT vowel in the interview, with the style of each demonstration marked next to the tip of each arrow

4.3.8. PRICE

PRICE, which is usually described as being monophthongal in AmE style pop singing accents, appears for these singers to be a diphthong which fronts and raises in both singing and speech. There are, however, several significant differences between the conditions. Firstly, the overall position of the vowel is opener and fronter in singing than speech for all singers, at least for the vowel nucleus. All four position measurements were significantly different for Dylan ($p<.001$ for all) and John ($p<.001$ for both F1a and F2a, $p<.05$ for F1b, and $p<.01$ for F2b), while only the nucleus of the vowel was significantly different for Andrew ($p<.01$ for F1a and F2a).

The movement of the vowel is also different across conditions. While the vowel tends to front and raise in both conditions, sung PRICE starts from an opener and fronter position and therefore raises more but fronts less than its spoken counterpart. For Dylan (Figure 4.25) both movement differences are significant, with the vowels raising significantly more in singing ($p<.001$), and fronting significantly more in speech ($p<.01$). For Andrew (Figure 4.26), PRICE raises significantly more in singing than speech ($p<.05$), while there is no significant difference in F2 movement. For John (Figures 4.28-4.29),

the spoken vowel starts further back and therefore fronts significantly more than the sung one ($p<.001$), but the difference in F1 movement across conditions is not significant. In absolute terms, the amount of movement in both conditions is much greater for Dylan and Andrew than it is for John.

The PRICE vowel had a much higher token count than any of the other vowels, particularly for Dylan. This means that it is difficult to display all of the pairs in a transparent way on the graphs, with most of the word labels in Figure 4.25, for example, being difficult to read. This problem was mitigated slightly in Figure 4.28 by positioning the labels for recited tokens at the start, rather than the tip of the arrow.

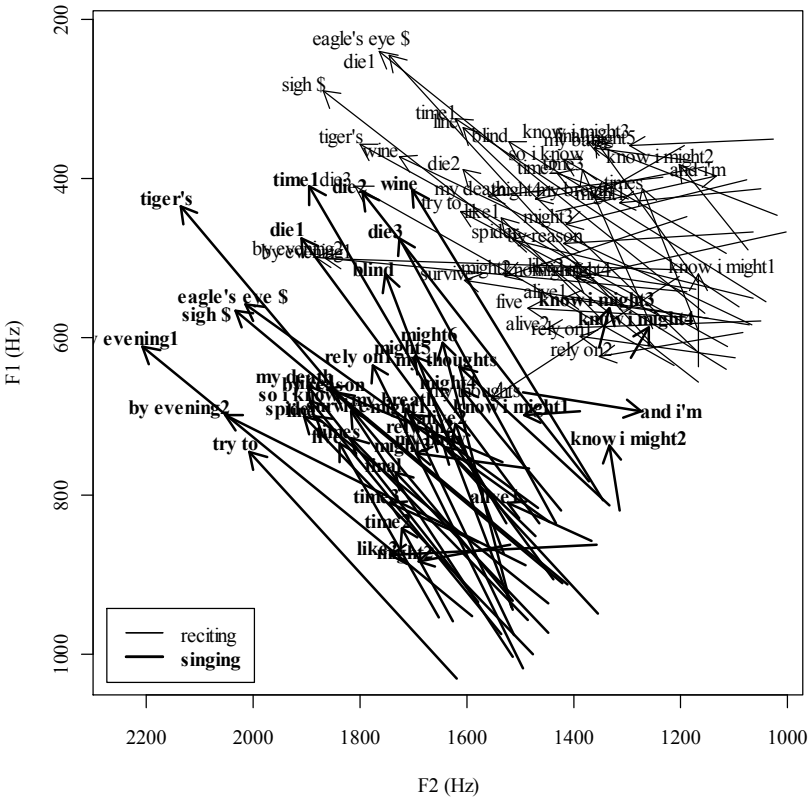

Figure 4.25: Dylan PRICE pairs

For both Dylan and Andrew, there is a tendency for short tokens to be less diphthongal. There are five tokens of Dylan's PRICE which do not move far from the position of the nucleus. All of these tokens have relatively short vowels, ranging from 68-106ms. The majority of the sung tokens, which are more diphthongal, tend to have durations above 250ms. It is also noticeable that the short tokens are all in the case of the word *I* or *I'm*. Figure 4.27 repeats the graph of Andrew's paired tokens, but showing the duration of the vowel (in ms) instead of word labels. This figure shows the tendency for shorter PRICE vowels to lack the fronted and raised offglide typical of the longer tokens.

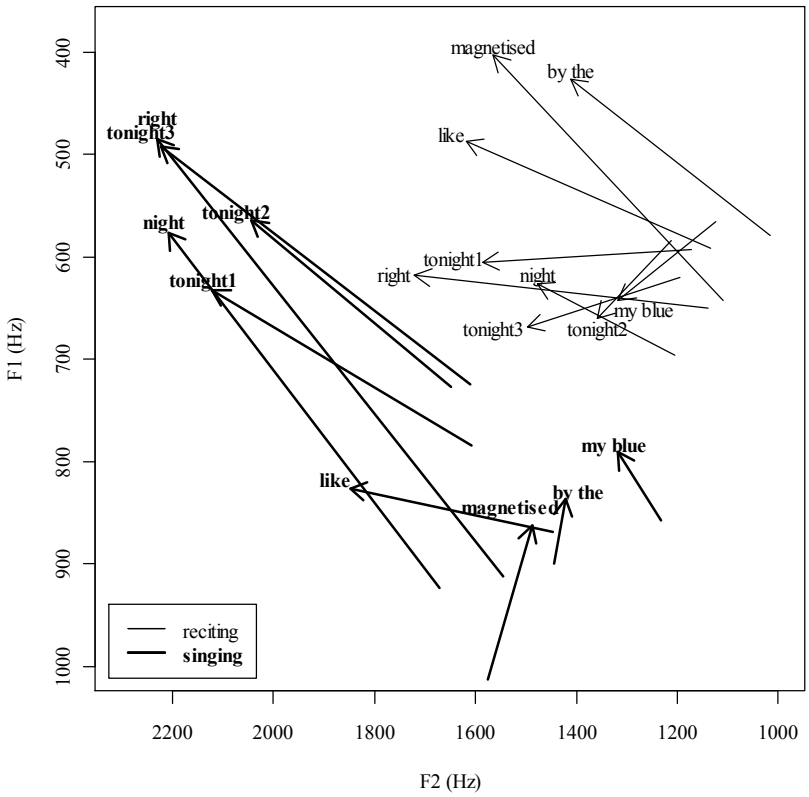

Figure 4.26: Andrew PRICE pairs

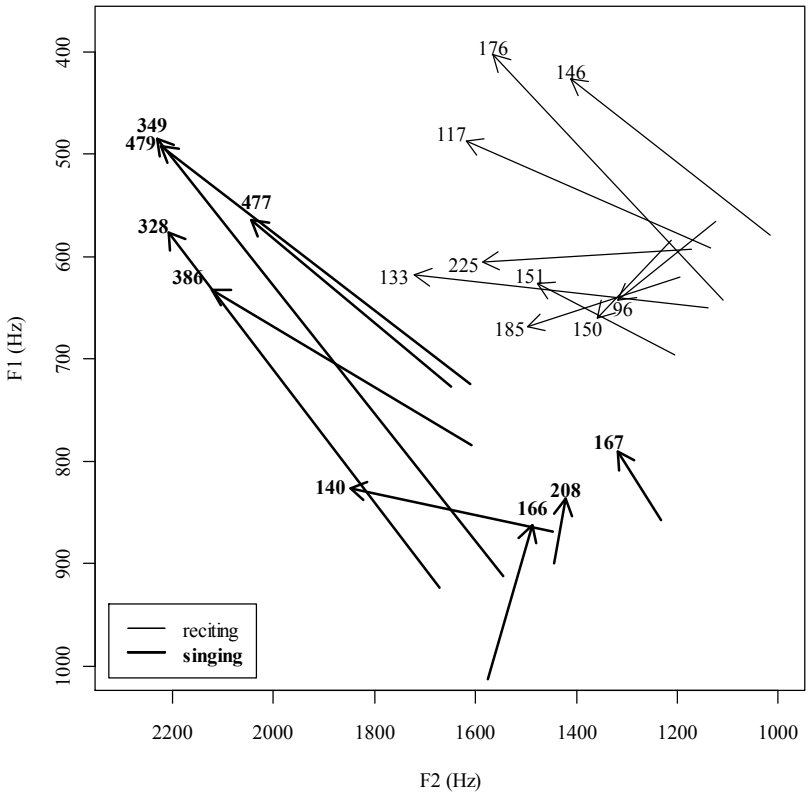

Figure 4.27: Andrew PRICE pairs showing duration (ms) of each vowel

John's PRICE vowels have been represented in two separate figures due to an effect of following environment. The realisation appears to differ according to whether or not the vowel precedes a bilabial or labiodental consonant.[6] This is the only significant environmental effect found in the entire dataset. Figure 4.28 shows all of John's pairs excluding those preceding /m,b,v,f/.

Figure 4.29 shows the pairs which occurred before /m,b,v,f/ (henceforth, the prelabial pairs), which tend to be less diphthongal (shorter distance of movement), and also tend to move backward rather than forwards in singing. The pronunciation of recited PRICE in this environment is different to the non-prelabial environments, too, though it is not different in any consistent way.[7] The mean duration of the pre-labial tokens is only 50ms shorter than for the tokens with other following environments (average durations 250ms for pre-labial and 302ms for others), suggesting that the difference is not just to do with duration as it appeared to be for Dylan and Andrew. This environmental effect is somewhat puzzling, since it does not occur for the other singers. It is possible that this effect is due to the words *I* and *I'm* rather than to the following environment.

In Figure 4.28, one of John's most notably NZ-like sounding tokens occurs, on *fight1*. The vowel moves forward without much raising, and has a retracted nucleus in almost exactly the same position as its spoken pair partner.

---

[6] A paired Wilcoxon signed-rank test on just the non-prelabial pairs found that the absolute position of the vowel was significantly different between conditions only for the nucleus of the vowel ($p<.001$ for both F1a and F2a). The difference between singing and reciting in terms of movement was significant for both F1 and F2, with the sung tokens raising more and fronting less than their recited counterparts ($p<.05$ for F1 movement, and $p<.001$ for F2 movement). The prelabial tokens are left in for the main analysis since it would be inconsistent to remove them from John's data and not the other singers. I did check how all the results looked if they were excluded and it does not change the directionality of any of the differences between singing and speaking, only the degree.

[7] There may be some issues with the measurement strategies for PRICE. For instance, the reason some of John's recited PRICE vowels show opening (increasing of F1) between measurements *a* and *b* is because the F1 of these vowels actually started low-ish, then raised in the middle portion of the vowel, and then lowered again towards the end. Impressionistically, the vowel still sounds like an opening diphthong because of the transition from the middle to the end of the vowel, but these measurements have missed that middle section.

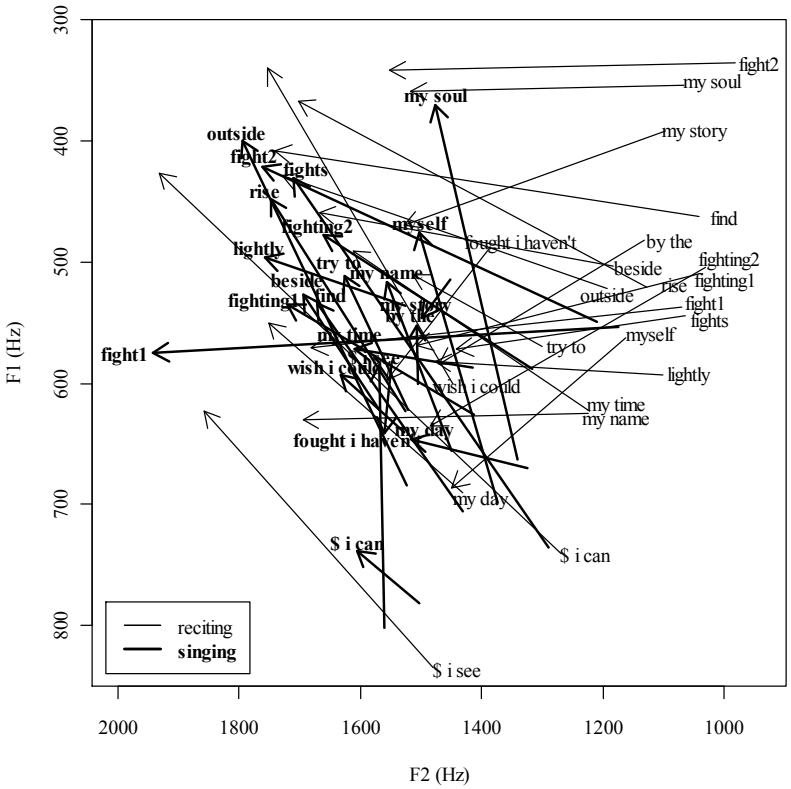

Figure 4.28: John PRICE pairs, excluding those followed by /m,b,f,v/; word labels on recited tokens are plotted at the start, not the tip, of the arrow in this figure.

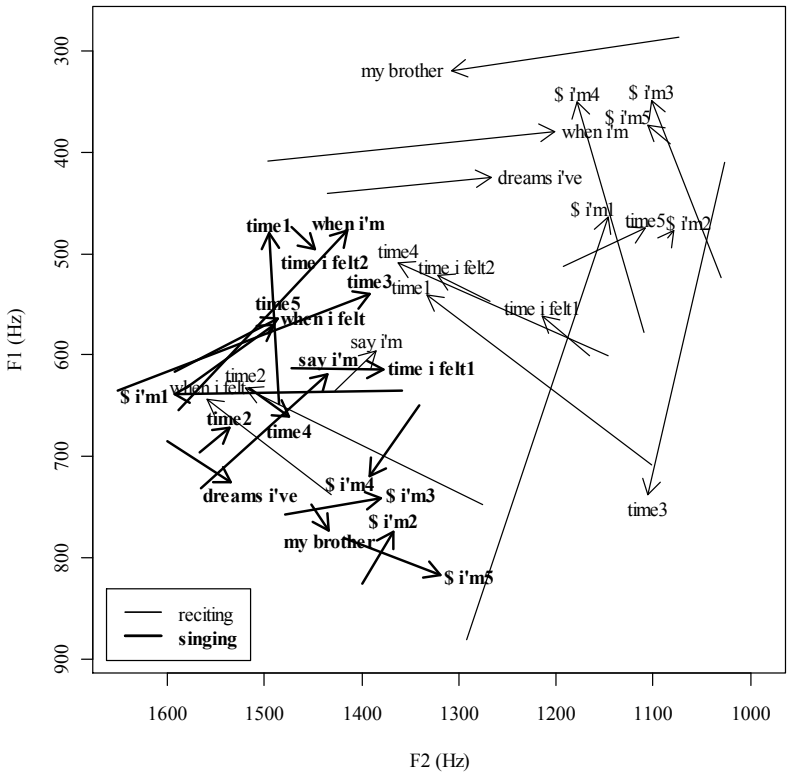

Figure 4.29: John PRICE pairs followed by /m,b,v,f/

The prelabial tokens for John, and the short tokens for Dylan and Andrew are all cases of the impressionistically monophthongal PRICE which is probably close to the variant of the vowel which has been described as typical of AmE style pop singing. It is difficult to tease apart the effects of duration and following consonant, complicated all the more by the fact that these monophthongal PRICE vowels include most cases of the word *I*, *I'm* and *I've*, which are no doubt some of the most frequent lexical items in all of popular music.

It should also be noted that there is a difference in the realisation of the longer, diphthongal tokens and their spoken pairs, aside from those already mentioned. A comparison of the spectrograms of several tokens reveals that the sung vowels tend to have a long steady state followed by a short offglide while the spoken vowels tend to show a more continuous gliding movement. The existence of the long steady state nucleus in the sung tokens may contribute to the impression that the sung tokens are monophthongal.

Summary

All of the eight vowels have now been described with respect to the paired data. To summarise, there are significant differences between singing and speaking which occur on all dimensions; vowel height, vowel frontness, and direction of diphthong movement. Overall, it seems that there is a tendency for F1 to be higher in singing than in speech, though this tendency is much weaker for John, and is actually reversed for his START and GOAT vowels. In the next section, we will look at the averages of vowel realisations across the larger dataset.

4.4. Analysis of Full Dataset

In this section I will show the overall vowel space for the averaged sung and spoken vowels for each singer. This analysis is based on a two way division between singing and speaking, pooling the data from the reciting and interview conditions together. Before presenting the vowel space graphs, I will justify the conflation of the two spoken conditions by showing the relative lack of difference between them.

4.4.1. Comparison of reciting and interviews

The vowel formant values in the reciting of song lyrics and in the interview data were relatively similar, suggesting that the reciting style was fairly representative of the singers' more casual speech style. Since there are very few significant differences

between these conditions, the detailed results are not included in the body of this chapter, but are provided in the first three tables of Appendix C. These tables show the mean formant values for vowels in interviews and reciting, with results of unpaired Wilcoxon rank sum tests for each vowel, for each singer. The differences which did reach significance are plotted below, in Figure 4.30.

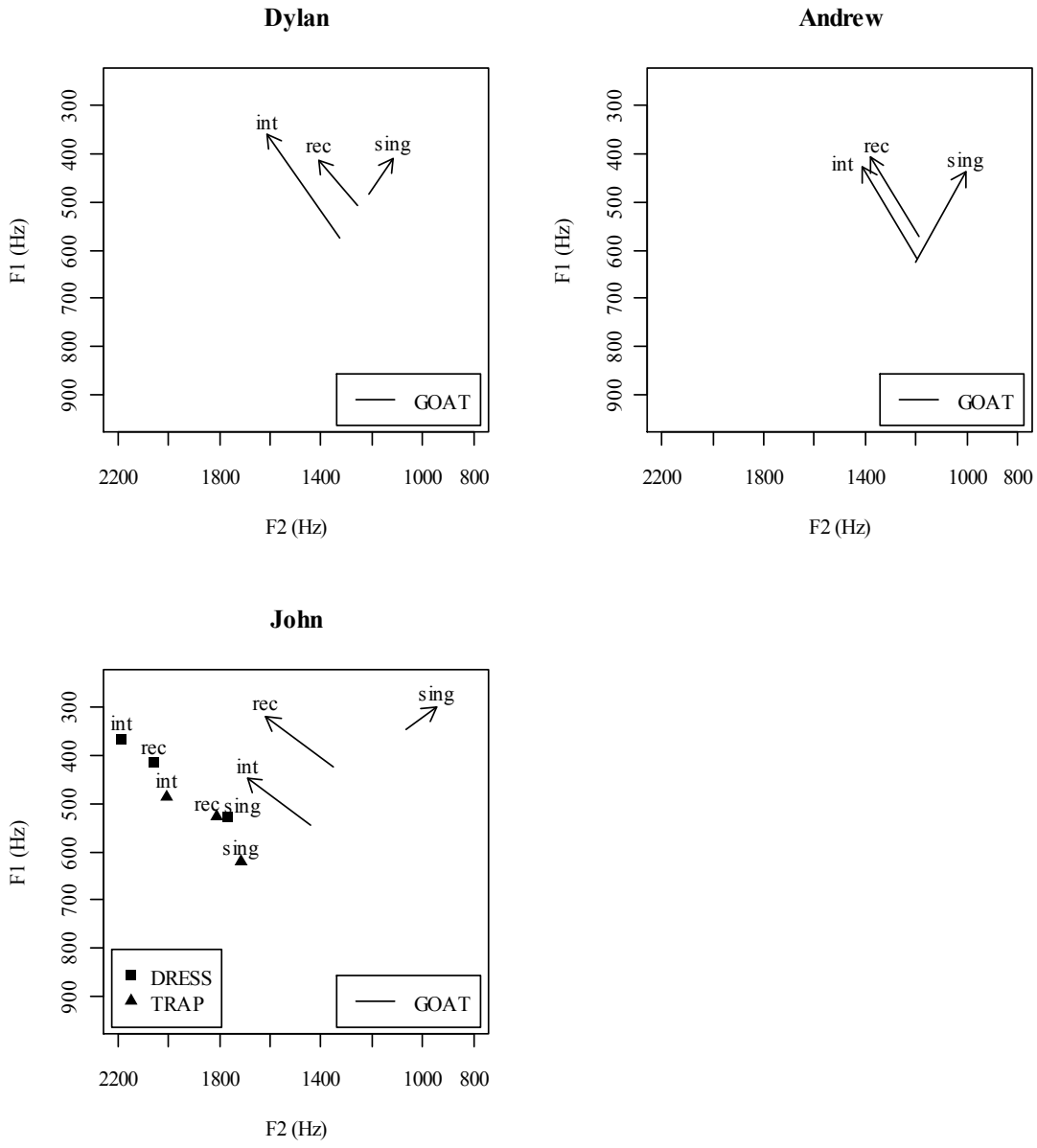

Figure 4.30: Comparison of the interview (int), reciting (rec), and singing conditions for vowels with significant differences between interview and reciting (GOAT for all three singers, and also DRESS and TRAP for John).

For all three singers there was a significant difference between interview and reciting data for GOAT. For Andrew and John, this difference was only in terms of the position of the vowel, not its movement. Dylan's interview tokens, however, have a greater closing movement than his recited tokens, and they front significantly more. The only other

significant differences were for the height of John's DRESS and the frontness of his TRAP. In all these cases, the recited and interview vowels pattern together when compared to the sung vowels, with the recited tokens falling in between the interview data and the singing data. This is what we would expect since the reciting and singing data are inherently more similar, being based on the same words. Since there were only a few significant differences between reciting and interview data, and since even these differences pattern together when compared with singing, it is fair to say that the reciting data is in general quite similar to the interview data. These two conditions can therefore justifiably be pooled together as 'spoken' data. In the next section, I will use the expanded dataset created by grouping interview and reciting together. This enlarged dataset will also include all the sung tokens for which there was no pair in reciting, and vice versa.

4.4.2. Average values for all spoken and sung vowels

The vowel space graphs in this section show the mean F1/F2 measurements for all the data in the expanded dataset, giving a clear picture of the differences between singing and speaking. The three graphs are on the same scale to allow the reader to compare the relative positions of vowels between singers.

The vowel spaces for Dylan (Figure 4.31), Andrew (Figure 4.32) and John (Figure 4.33) show the same patterns that were seen in the paired data, with large differences apparent between singing and speaking overall. The (unpaired) Wilcoxon rank sum tests also follow the patterns seen for the paired data, though more of the differences reached significance since the token count was higher. The results of these tests are shown in Table 4.3, Table 4.4, and Table 4.5 with the difference between conditions, and also the mean formant (and formant movement) values for each vowel. The results are summarised below, making reference to both the tables and the figures.

DRESS and **TRAP** are opener in singing than speech for all three singers ($p<.001$ for both vowels for all singers), with the magnitude of this difference being less for John. John's spoken DRESS and TRAP are also fronter in singing than speech ($p<.001$ for DRESS and $p<.01$ for TRAP).

THOUGHT and **LOT** are opener and fronter in singing. For THOUGHT, this difference is highly significant ($p<.001$) for F1 and F2 for all singers. LOT is significantly opener for all singers ($p<.001$ for Dylan and John, $p<.05$ for Andrew), and significantly fronter for

Andrew (p<.05) and John (p<.001), but not fronter for Dylan. The F1 difference is smaller for John than for Dylan and Andrew.

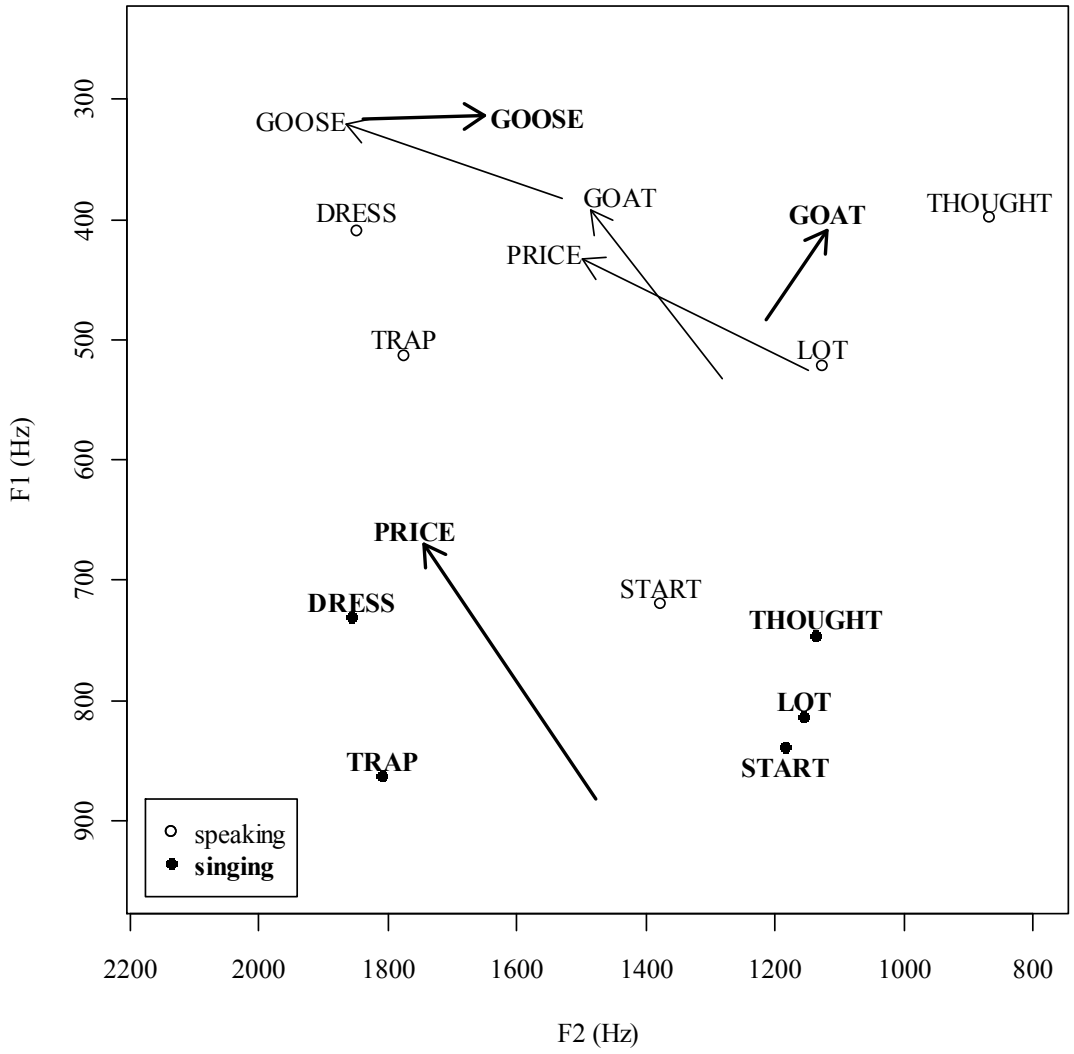

Figure 4.31: Dylan's vowel space, showing mean vowel positions for all spoken and sung vowels. Monophthongs are represented by points, and diphthongs are represented by arrows, with labels next to the tip of the arrow.

START is fronter in speech than singing for Andrew (p<.01), and also Dylan to a similar degree though there were insufficient tokens to test the difference for Dylan. There is no significant F2 difference for John. For Andrew and Dylan, START is opener in singing than speech (p<.01 for Andrew, insufficient tokens for Dylan), while for John the opposite pattern is found, with F1 being significantly lower in singing than speech (p<.01). For both Dylan and Andrew, THOUGHT, LOT and START occupy a similar acoustic space in singing, even though they are distinct in speech.

Table 4.3: *Mean Formant Values for Dylan's Sung and Spoken Vowels, with Mean Differences between Points* a *and* b *of Diphthongs. Significance Levels for all Wilcoxon Tests also Shown*

	F1				F2			
	sing (n)	speak (n)	diff.	p	sing	speak	diff.	p
DRESS	731 (18)	409 (8)	322	***	1858	1849	9	NS
TRAP	863 (22)	512 (14)	351	***	1809	1776	33	NS
THOUGHT	745 (26)	398 (16)	347	***	1138	869	269	***
LOT	814 (16)	521 (8)	293	***	1155	1129	26	NS
START	839 (3)	719 (1)	120	NA	1185	1378	-193	NA
GOOSE (a)	316 (4)	407 (4)	-91	*	1840	1519	321	NS
GOOSE (b)	314	315	-1	NS	1650	1764	-114	NS
GOOSE (b-a)	-2	-92	90	NS	-190	+245	-435	*
GOAT (a)	483 (24)	533 (17)	-50	.	1215	1283	-68	NS
GOAT (b)	409	392	17	NS	1119	1487	-368	***
GOAT (b-a)	-74	-141	67	**	-96	+204	-300	***
PRICE (a)	882 (48)	525 (50)	357	***	1478	1148	330	***
PRICE (b)	670	432	238	***	1745	1499	246	***
PRICE (b-a)	-213	-93	120	***	+267	+350	-83	*

(*** for p<.001; ** for p<.01; * for p<.05; . for p<.1; NS for not significant; NA for cases where there are too few pairs to warrant a test for significance)

GOOSE raises and fronts in speech and retracts in singing. All three singers show this pattern with the difference in F2 movement being significant for both Dylan and John ($p<.05$), and approaching significance for Andrew ($p=.095$). As well as the dramatic difference between conditions for F2, the spoken vowels start in a more open position and then raise while the sung tokens show no movement on the F1 dimension, being raised throughout. The beginning of Dylan and John's vowels is significantly opener ($p<.05$) in speech than singing, and the difference in F1 movement is also significantly different across conditions for John ($p<.05$).

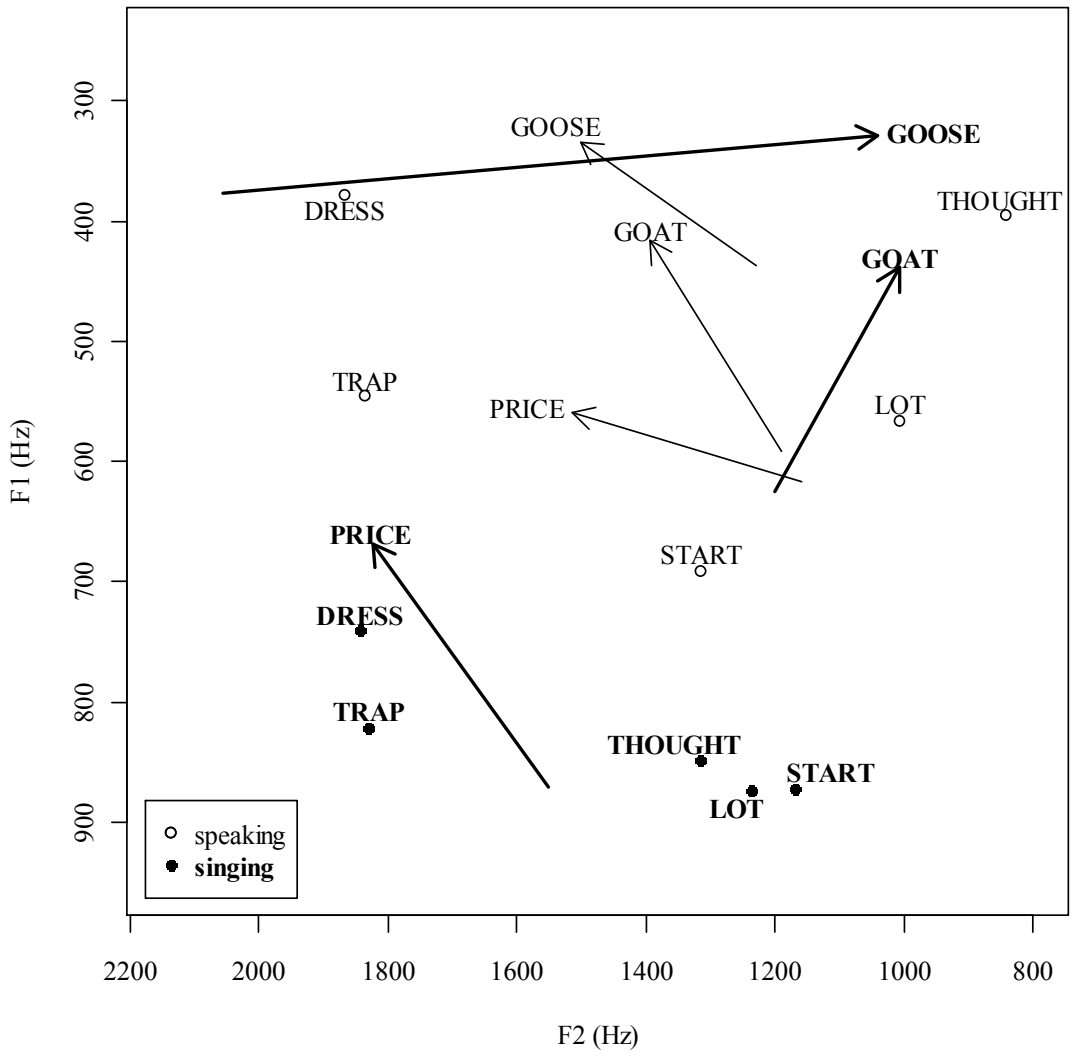

Figure 4.32: Andrew's vowel space, showing mean vowel positions for all spoken and sung vowels. Monophthongs are represented by points, and diphthongs are represented by arrows, with labels next to the tip of the arrow.

Like GOOSE, **GOAT** raises and fronts in speech and raises and retracts in singing (the difference in F2 movement is highly significant for all singers, $p<.001$) and the movement is of a similar magnitude for all singers (spoken GOAT fronts by ~200-250Hz while sung GOAT retracts by ~100-200Hz). For Dylan and John, the vowel raises more in speaking than singing ($p<.01$ for Dylan, $p<.05$ for John) while for Andrew, the vowel raises to the same extent in both conditions. GOAT starts in the same position across conditions for Dylan and Andrew and then finishes up fronter in speech than singing ($p<.001$ for both), while the whole vowel is higher and backer in singing than speech for John (F1a and F2a $p<.01$, F1b $p<.05$, F2b $p<.001$).

Table 4.4: *Mean Formant Values for Andrew's Sung and Spoken Vowels, with Mean Differences between Points* a *and* b *of Diphthongs. Significance Levels for all Wilcoxon Tests also Shown.*

	F1				F2			
	sing (n)	speak (n)	diff.	p	sing	speak	diff.	p
DRESS	741(22)	378 (9)	363	***	1845	1868	-23	NS
TRAP	822(24)	546(11)	276	***	1831	1836	-5	NS
THOUGHT	848(26)	395(12)	453	***	1317	844	473	***
LOT	874 (8)	566 (3)	308	*	1237	1008	229	*
START	872 (9)	691 (3)	181	**	1170	1315	-145	**
GOOSE (a)	377 (5)	447 (2)	-70	NS	2055	1325	730	.
GOOSE (b)	329	354	-25	NS	1041	1520	-479	.
GOOSE (b-a)	-48	-93	45	NS	-1014	+195	-1209	.
GOAT (a)	625(22)	591(16)	34	NS	1202	1191	11	NS
GOAT (b)	439	416	23	NS	1008	1394	-386	***
GOAT (b-a)	-187	-175	-12	NS	-194	+203	-397	***
PRICE (a)	871(10)	617(10)	254	***	1552	1160	392	***
PRICE (b)	668	559	109	NS	1824	1516	308	.
PRICE (b-a)	-203	-58	-145	*	+272	+355	-83	NS

(*** for p<.001; ** for p<.01; * for p<.05; . for p<.1; NS for not significant; NA for cases where there are too few pairs to warrant a test for significance)

PRICE fronts and raises in both singing and speech, with more fronting in speech than singing, and more raising in singing than speech. For Dylan both movement differences are significant, with the vowels raising significantly more in singing (p<.001), and fronting significantly more in speech (p<.05). For Andrew, PRICE raises significantly more in singing than speech (p<.05), while there is no significant difference in F2 movement. For John, the spoken vowel fronts significantly more than the sung one (p<.001), but the difference in F1 movement across conditions is not significant.

The overall position of the vowel is opener and/or fronter in singing than speech for all singers. All four position measurements were highly significant for Dylan (p<.001), while only the start of the vowel was significantly different for Andrew (p<.001 for F1a and F2a). For John, both the start and end of the vowel were significantly opener in

singing, and the start of the vowel was also significantly fronter in singing ($p<.001$ for both F1a and F2a, $p<.01$ for F1b). As seen in the paired section, the averaging of values for this vowel conceals some variation which occurs on the basis of vowel duration and following environment.

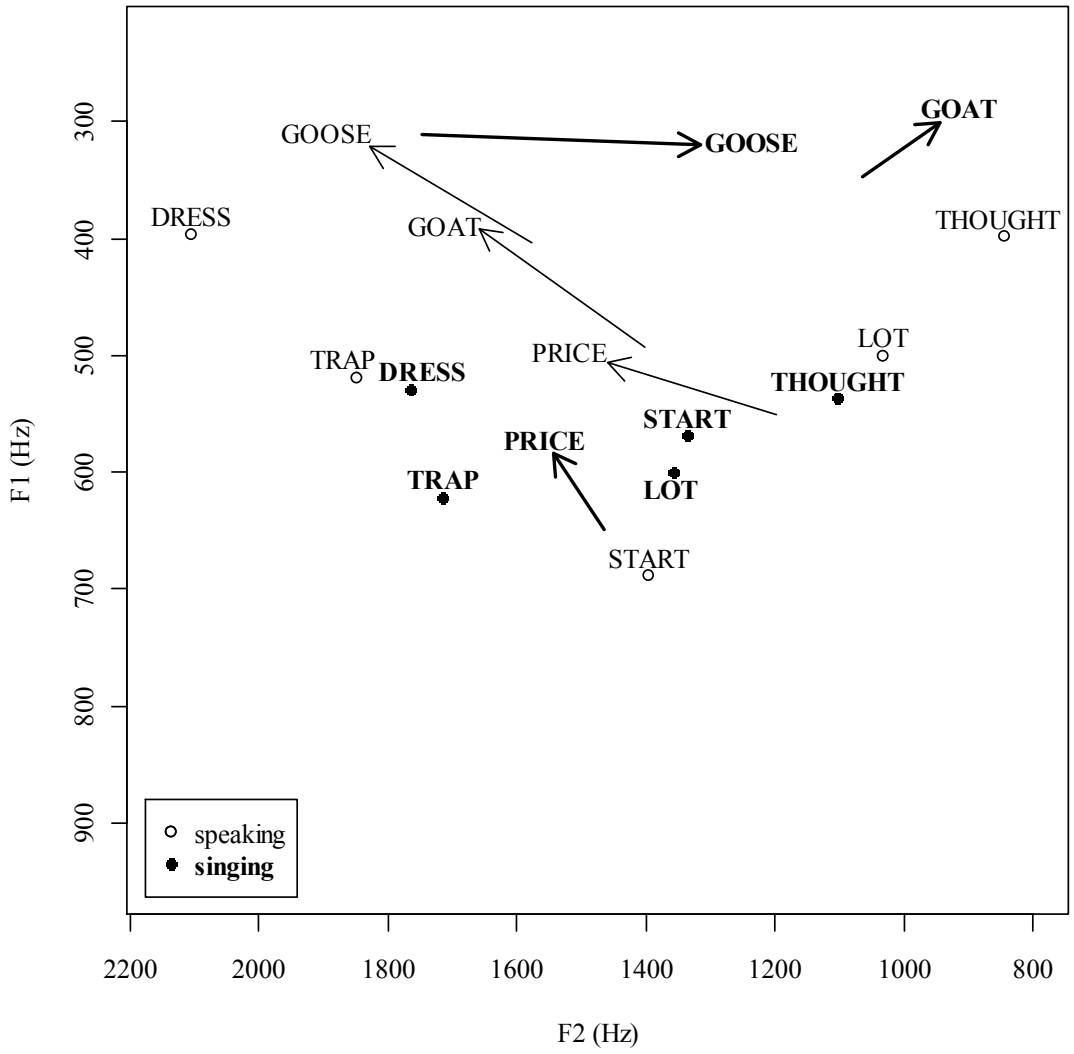

Figure 4.33: John's vowel space, showing mean vowel positions for all spoken and sung vowels. Monophthongs are represented by points, and diphthongs are represented by arrows, with labels next to the tip of the arrow.

Table 4.5: *Mean Formant Values for John's Sung and Spoken Vowels, with Mean Differences between Points* a *and* b *of Diphthongs. Significance Levels for all Wilcoxon Tests also Shown.*

	F1				F2			
	sing (n)	speak (n)	diff.	p	sing	speak	diff.	p
DRESS	529 (21)	396 (13)	133	***	1766	2107	-341	***
TRAP	621 (19)	519 (14)	102	***	1716	1850	-134	**
THOUGHT	536 (21)	398 (16)	138	***	1105	845	260	***
LOT	600 (28)	500 (27)	100	***	1358	1033	325	***
START	568 (10)	687 (4)	-119	**	1338	1396	-58	NS
GOOSE (a)	311 (7)	403 (3)	92	*	1747	1577	170	NS
GOOSE (b)	320	321	-1	NS	1317	1827	-510	.
GOOSE (b-a)	+9	-81	90	*	-430	+250	-680	*
GOAT (a)	347 (23)	492 (8)	-145	**	1065	1401	-336	**
GOAT (b)	301	392	-91	*	946	1660	-714	***
GOAT (b-a)	-45	-100	55	*	-119	+258	-377	***
PRICE (a)	649 (39)	551 (43)	98	***	1466	1197	269	***
PRICE (b)	583	506	77	**	1544	1460	84	NS
PRICE (b-a)	-65	-45	-20	NS	77	263	-186	***

(*** for p<.001; ** for p<.01; * for p<.05; . for p<.1; NS for not significant; NA for cases where there are too few pairs to warrant a test for significance)

4.5. Pitch and Duration

Having presented detailed data on F1 and F2, I now turn to an examination of the other variables which were recorded for analysis, namely pitch and duration. Duration was mentioned above as affecting the realisation of PRICE. Pitch is also likely to play an important role in the realisation of vowels, particularly in terms of F1.

There is a significant difference between singing and speech in terms of both pitch and duration, with sung vowels averaging almost an octave higher and over three times longer than spoken vowels. The musical interval of an octave is equivalent to the doubling of a given pitch when measured in Hertz. A t-test of the tokens in the full

dataset showed that this difference is highly significant for all singers (p<.001 for both pitch and duration).

Table 4.6 shows the mean pitch (in Hertz) and the mean duration (in milliseconds) with standard deviations, for each singer, for all 765 vowel tokens in the dataset. John has a lower pitch than Dylan and Andrew in both singing and speaking, and also a shorter mean duration than the other two singers. The standard deviation around the mean is greater for singing than speaking, especially for duration. This finding can be compared to the results for F1 and F2, where sung vowels were also much more variable than spoken vowels. The variability in pitch and duration in singing may be one of the reasons for the variability in vowel pronunciation.

Table 4.6: *Mean Pitch and Duration according to Condition, with Results of t-test*

		sing		speak		
		mean	s.d.	mean	s.d.	p (t-test)
pitch (Hz)	Dylan	205	48	133	22	***
	Andrew	254	63	121	19	***
	John	162	43	99	18	***
duration (ms)	Dylan	422	401	142	115	***
	Andrew	483	506	129	78	***
	John	275	122	113	55	***

(*** for p<.001)

4.5.1. Differences between choruses and verses

Some of the variability in the pitch and duration of sung vowels can be accounted for in terms of the section of the song in which the vowel occurs. One of the hallmarks of a popular song is the use of a repeated chorus, which usually contains the song's 'hook' (a catchy melody, for example). Stereotypically, choruses are set apart from verses by being louder and having greater emotional intensity. This increased intensity can be achieved in many ways, though it is common for a chorus to use a higher part of the singer's pitch range, and to include words with extended vowel lengths.

Dylan and Andrew's songs were analysed to determine whether the chorus sections are higher in pitch and have longer vowels than the verses. This distinction was not analysed for John because all of his songs have repeated catch-lines rather than full choruses. This is an interesting finding in itself, and its implications will be discussed at the end of section 5.1.3. The median duration for the vowels collected in Dylan's song is 374ms (n=75) in the chorus and 212ms in the verses (n=48). Median pitch is 208Hz

in choruses and 200Hz in verses. The duration difference is highly significant (Wilcoxon rank sum, $p<.001$). For Andrew's song, the median vowel length is 412ms in the chorus (n=30) and 314ms in the verses (n=41). Median pitch is 296Hz in choruses and 225Hz in verses. The difference is highly significant for pitch ($p<.001$) but not significant for duration.

Since choruses seem to be higher and have more drawn out vowels than verses, they could be described as being more 'singy' than verses (to use the adjective coined by Coddington, 2004), viewing singing and speech as poles on a continuum, rather than categorically distinct modes. For Dylan, there is some evidence that his pronunciation is also more singy in chorus sections, which seem to accentuate the differences between singing and speaking, while in verses these differences diminish slightly, see Figure C.1 in Appendix C for a comparison of Dylan's vowel realisations in choruses and verses.

This is an interesting finding in light of Simpson's (1999, p. 360) claim that "the less a singer 'sings', so to speak, the weaker the influence of the external code and the stronger the approximation to the singer's own vernacular usage". While this discussion is speculative, it highlights one way in which pitch and duration may be connected to pronunciation in singing. They are two of the factors which may reliably differentiate singing and speech in a vocal sense, and as such, they may also provide a way to objectively assess where a given stretch of language falls on the continuum between 'singy' and 'speaky'.

4.5.2. The effect of pitch on F1

The fact that singing has consistently higher-pitched and longer vowels than speech introduces a major question. Are the differences between singing and speech based on a socially motivated change in style, are they an artefact of the music-related differences in pitch and duration, or do both of these factors have an influence? This is the multi-collinearity issue mentioned in 4.1, which makes it difficult to ascertain the effect that condition has on vowel pronunciation independent of possible pitch and duration effects. The rest of this section is devoted to resolving this important issue, focusing particularly on the effect pitch may have on F1.

For F2 differences, there are no reasons to suspect that pitch or duration cause the differences found between conditions. There is no evidence in the literature to suggest that higher pitch or longer duration would cause the opposing directions of F2 movement seen between singing and speech for GOOSE and GOAT, for example – these

differences appear to be due to other factors, arguably those associated with identity construction.

There are, however, two reasons we might suspect pitch to be playing a role in the F1 differences. Firstly, there is evidence in the literature that singers tend to open their jaw as pitch increases, which would lead to a more open vowel and a higher F1 (Austin, 2007). Secondly, there is evidence that singers, particularly those who are not classically trained, raise their larynx as they sing higher notes (Howard, 2009). This larynx-raising would lead to a shortening of the pharyngeal cavity, and thus a higher F1. It should be noted that this change would result in an F1 difference without being directly related to vowel height. Since this study is based on acoustic data, not articulatory measurements, it is difficult to tease apart the possible roles that jaw opening and larynx raising might play.

Given the potential arguments for an effect of f0 on F1, it is worthwhile investigating the data more closely in order to assess whether the F1 pronunciation differences are an artefact of singing's higher pitch. If it is true that the increase in F1 is caused by singing being higher than speaking, then we should find a positive correlation between pitch and F1 within each condition.[8] The argumentation presented in the next few pages is of a more technical style than the description of results thus far. The main outcome of the analysis will be summarised in general terms at the end of the section.

For each singer, a Spearman correlation test was conducted for each of the eight vowels, for both the sung and spoken data.[9] The results of these tests are shown in Appendix C, in Table C.4. Only a handful of these tests reached significance, though it was seen that in most cases, the direction of the correlation coefficient was positive. For this reason, it was decided that it would be informative to build regression models that would test for an effect of f0 and duration on F1, within each condition.

Six linear models (for singing and speaking, for each singer) were built with F1 as the dependent variable and f0, duration and vowel as explanatory variables. Unsurprisingly, there were significant effects for most of the different vowels, since they have

---

[8] Because pitch and F1 are both higher in singing than speech, there is already a strong correlation between pitch and F1 in the dataset as a whole. We are interested in the effect of pitch on F1 *independent* of condition. To search for this we need to look for an F1/f0 correlation *within* each condition.

[9] The spoken data included tokens from both reciting and interviews.

inherently different F1 values. The linear model accounts for these differences and holds them constant, to determine the independent effects of pitch and duration.

For Dylan's singing, there is a significant correlation effect of f0 upon F1, with F1 increasing as pitch increases. Vowel duration was not related to F1. For Andrew's sung data, no relationship between pitch and F1 was found, however duration did have a significant effect, though very slight. For John, the results were not straight-forward, with F1 decreasing with increasing pitch for low pitches (up until about 200Hz), and then increasing with pitch for higher tokens. There was no effect of duration on F1 for John. None of the speakers had any main effects of pitch or duration on the F1 values for vowels in speech.[10]

From these results we can conclude that for Andrew, because pitch does not affect F1 in either condition, the F1 differences between singing and speaking are not artefacts of the higher pitch in singing. There is, however, an effect of duration, such that longer sung vowels tend to be opener for Andrew. This duration effect, though, is far too subtle to account for the large F1 differences seen between conditions.

For Dylan, there is a clear effect of pitch on F1 in the sung data. For John, there is an effect of F1 increasing with pitch, but only for sung vowels above about 200Hz. For John's lower notes, F1 decreases with increasing pitch. For the sung data, then, Dylan and John seem to have increased jaw opening and/or larynx-raising at higher pitches. Does this account for the F1 differences between singing and speech? To investigate this question, the figures below plot F1 against f0, for the five vowels which consistently show an F1 difference between conditions: DRESS, TRAP, LOT, THOUGHT and PRICE. Lowess smoother trend lines are fitted through the spoken and sung data (with the smoother setting f=0.85), and the vertical axis, representing F1, is inverted as it was in previous F1/F2 graphs. If the F1 difference between conditions is an artefact of singing's higher pitch, the sung data should meet up with the spoken data in the portion of the graph showing tokens at a lower pitch.

---

[10] For Dylan's singing, the regression model estimated that for each 1Hz increase in pitch, F1 increases by 1.23Hz, holding all other variables constant. For Andrew's singing, F1 was estimated to increase by 0.53Hz for every 10ms increase in duration, holding all else constant. For John's singing, there was curvature in the residuals of the basic linear model, so the model was refit with a quadratic term for pitch ($f0^2$). Both f0 and $f0^2$ were then significant. There were no statistically significant effects of pitch on F2.

In Figure 4.34, for Dylan, the correlation between pitch and vowel height is apparent, and importantly, so is the effect of condition, with the two trend lines being clearly separated on the left-hand side of the graph, which shows lower pitched tokens. For John (Figure 4.35), we can see the way the F1 of sung vowels decreases at first, and then increases, with increasing pitch. There is also a distinguishable gap between conditions, though to a much lesser degree than for Dylan. The important point is that the two trend lines are separate on the left-hand side of the graph, where tokens have similar pitches.[11]

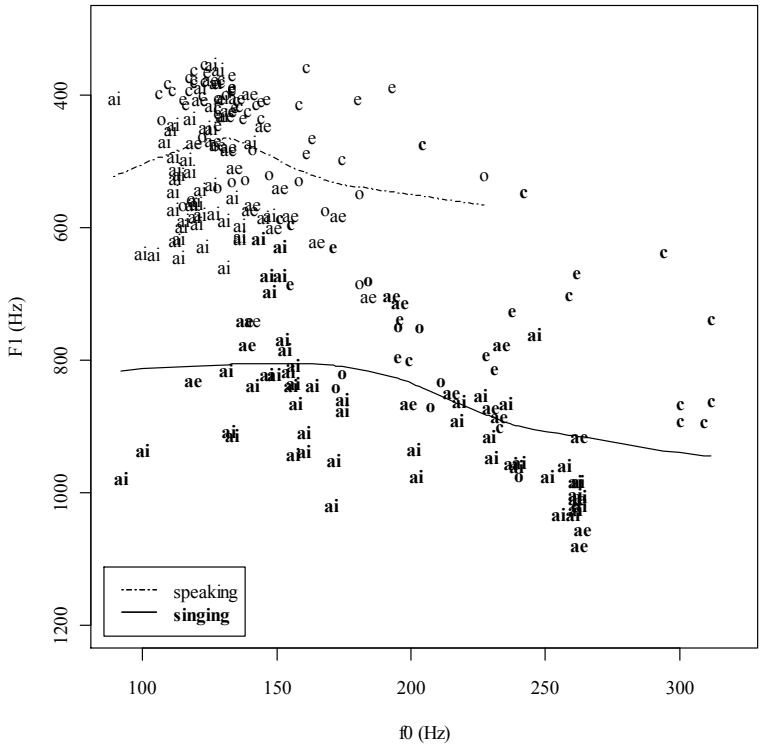

Figure 4.34: Relationship between f0 and F1 for Dylan. e=DRESS, ae=TRAP, o=LOT, c=THOUGHT, and ai=PRICE (nucleus).

[11] Note that to ensure that the smoother lines on these graphs were legitimate, it was necessary to check that each vowel type was randomly distributed in terms of pitch. For Dylan's singing, THOUGHT had a significantly higher pitch than the other vowels, however, it can be seen from the graph that the high pitched THOUGHT tokens do not alter the interpretation that there is a large condition difference that is not driven by pitch.

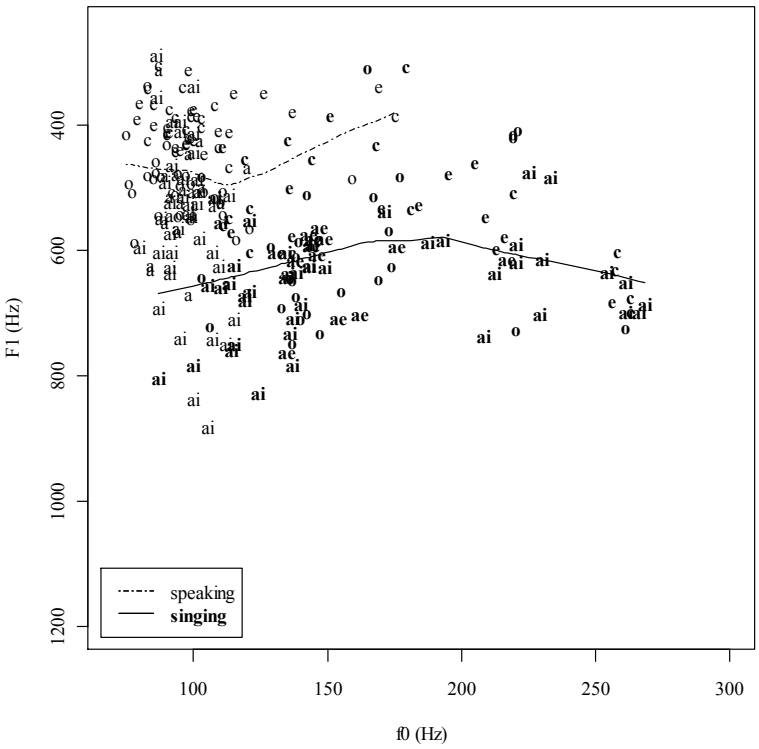

Figure 4.35: Relationship between f0 and F1 for John. e=DRESS, ae=TRAP, o=LOT, c=THOUGHT, and ai=PRICE (nucleus).

From these graphs we can conclude that for Dylan, the higher F1 seen in singing than speaking involves both an effect of the higher pitch used in singing, but also other influences, arguably a socially motivated style difference. It remains to be seen whether the pitch-induced portion of this variation is caused by larynx-raising or jaw-opening, or both. For John, while the F1 difference between sung and spoken vowels is not as clear as it is for Dylan, there is very little evidence that the difference that does exist is an artefact of the higher pitch found in singing. The potential reasons for John's comparatively low F1 in singing will be discussed in the next chapter.

This somewhat technical discussion has been necessary in order to establish that even though sung vowels are higher in pitch and longer in duration than spoken vowels, the differences between sung and spoken vowel realisations that have been presented in this study are not merely artefacts of these music-related differences. What this section has shown is that for Dylan, opener vowels are used on higher pitched notes, but vowels are also more open in singing than speech, irrespective of pitch. While the analysis shows that the findings are not solely a result of the greater pitch and duration in singing, it has *not* ruled out the potential influence of a preference for sonorous vowels in singing. That is, we cannot automatically assume that the F1 differences are related to the

construction of social meaning. This issue will be addressed in the next chapter as the role of sonority is discussed.

4.6. Consonants

To conclude this presentation of the differences between singing and speech for these three NZ singers, and to move towards my interpretation of these data, I will briefly present the results of an analysis of two consonant variables. Though vowel production has been the focus of this study, it would seem neglectful not to briefly touch on some consonantal data. Non-prevocalic /r/, in particular, has been the most extensively studied variable in relation to pop singing. The main motivation for conducting an analysis of these features, however, is to provide another angle on the evidence that differences between singing and speaking are not artefacts of singing inherent effects, which will be discussed further in the next chapter.

I chose to analyse two widely studied consonant variables: non-prevocalic /r/ and (ING). The analysis of both variables is based on a binary impressionistic judgment; presence versus absence of /r/, and /n/ versus /ŋ/. In some cases I looked at the spectrogram to check my decisions.

Non-prevocalic-/r/

There were no cases of non-prevocalic /r/ in the reciting of lyrics for any of the singers. This is not surprising since NZE is mainly non-rhotic. Interestingly, neither Dylan nor John produced any instances of non-prevocalic /r/ in their singing either. Dylan had 16 sung tokens which could have been pronounced with /r/, and John had 34 such words, including several NURSE environments, which may favour realisation of /r/. As discussed in 2.3.3, non-prevocalic /r/ has a very high level of salience for NZ singers as a marker of American-ness in singing. So the fact that Dylan and John both avoid it is consistent with their negative stance towards the 'blatant' use of AmE features.

Andrew did produce non-prevocalic /r/ in singing, with 6 out of 20 potential /r/s being realised. 5 of the 6 /r/s occurred after the NURSE vowel (*her* and *burn*), and one after SQUARE (*hair*).[12] This result is consistent with Andrew's interview, where he expressed no negative opinions about AmE in singing.

[12] There were a total of seven NURSE environments and four SQUARE environments for Andrew.

(ING)

The realisation of (ING) as [n] has been described as a vernacular feature in the UK, America, Australia and New Zealand. Bell & Holmes (1992) found NZ speakers to realise (ING) as [n] at rates ranging from 15 to 46 percent, varying according to sex, age and socio-economic class. Comparing their results with studies from other regions, they conclude that for NZE, "ING reduction is at much lower frequencies than in the American and British data" (Bell & Holmes, 1992, p. 240). It is possible, then, that in singing pop music the [n] variant could index American-ness and [ŋ] could index NZ-ness. Note that (ING) reduction, at least in American speech, may also be used to construct various stances such as relaxed, easygoing and unpretentious, while the velar form can index stances like educated and articulate (Eckert, 2008; Kiesling, 1998).

None of the three singers produced any apical variants of (ING) in the reciting of their lyrics (total n=42), which lends support to the idea that the use of [ŋ] for (ING) is typical in NZE, at least in this reciting style. They varied in their sung pronunciation, with both Andrew and John using the apical form categorically (n=12 for each singer). For Dylan, 5 out of 18 sung instances of (ING) were realised as /n/. It is interesting that John aligns with Andrew on this feature, a finding which will be discussed further in the next chapter, where I will draw all of the results together for discussion.

The results for these consonant variables reinforce the overall finding of this vowel production study: that there are many differences in pronunciation between singing and speaking for the three singers under analysis. The consonant variables exemplify this trend in an uncomplicated way, avoiding the many collinearity issues that make some of the vowel formant differences more difficult to interpret. An important aspect of the results is the finding that singing is different to speech not only in terms of segmental pronunciation, but also in terms of its overall prosodic style, with higher pitch and longer vowels. This finding provides a good reminder that the study of singing accents constitutes a different endeavour to the majority of existing sociolinguistic research; we are dealing with two very different modalities, or channels, of verbal behaviour. The differences between the two modes are not just phonetic, nor even linguistic; they run through all aspects of the two forms of expression.

Chapter 5 Vowel Production Study: Discussion

In this chapter, I will consider the results of the vowel production study in relation to the hypothesis that American-influenced singing accents are to some extent automatic for many pop singers, and are 'easy' to perform, while the use of New Zealand pronunciation features in this context requires more effort. I will discuss this issue with reference to Bell's (1984, 2001) distinction between the responsive and initiative dimensions of style, and with reference to Eckert's (2008) notion of the indexical field, particularly emphasising the way indexical associations shift according to socio-contextual frames. I will also consider the way singing inherent effects might relate to the norms of pop singing pronunciation.

The main finding of the comparison of sung and spoken vowel production was that there are acoustic differences between singing and speech for all three singers. One difference is that singing tends to have more open vowels than speaking, and another is an opposing direction of movement between singing and speaking for the GOAT and GOOSE vowels. The details will be discussed for each singer in turn over the next few pages, considering the results of the quantitative analysis alongside examples of other variables, in some cases paying attention to a wider sample of the singers' songs. I will also examine how the results relate to each singer's interview responses.

5.1. Discussion of Results for the Three Singers

The purpose of this section is to provide an overall summary and interpretation of the results for each of the singers, complementing the discussion with some impressionistic descriptions of variables other than the eight vowels studied acoustically, and also providing some examples from other songs by the singers. Overall, there are surprisingly few cases of explicitly NZE-like vowels. Dylan and John both have moments of NZ-ness which are marked in their contrast to the otherwise normative American-influenced singing accent. Such moments of NZ-ness are entirely absent in Andrew's singing, in whose case, however, there are some quite marked American features which deserve comment. It must be emphasised that what follows includes some subjective interpretations of the social meanings of various tokens. It is not certain that others would agree with all of the descriptions I present. The reader is referred to the lyrics of the songs which are printed in Appendix D, and the recordings of the songs which are included on the attached CD.

5.1.1. Dylan

Dylan Storey was positively oriented to the NZ accent in singing but felt it was difficult to sing with NZ vowels. He tries to use 'something in between' the AmE and the NZ models. There are some signs of this pro-NZE stance, most notably in the several instances of fronting GOAT vowels in his song. However, these few tokens are the exception. Dylan's singing style is very different to his speech with, for example, opener DRESS and TRAP and opener, fronter, and less rounded THOUGHT and LOT. His sung PRICE vowel has an opener, fronter nucleus than its spoken counterpart. Sung PRICE also tends to have a long steady-state followed by a very short off-glide, which makes it sound impressionistically monophthongal. My overall impression of Dylan's accent is that it is a typical American country-blues singing accent with a subtle NZ twist.

If we were to repeat Trudgill's description of the AmE accents used by British artists in the early 1960s, we would say that Dylan has identified "Americans" as an esteemed group in this context, and he is attempting to imitate what he considers to be an American accent. If this is the case, shifts to AmE should only occur on features that are seen as stereotypically American; we might also expect to see hyper-American forms and inconsistent application of the 'rules' of AmE. This is not what was found in the analysis of Dylan's singing. Significant differences between speech and singing were found for all variables *except for* those features which are stereotyped as Americanisms in singing, notably non-prevocalic /r/ and, as will be discussed below, BATH. By including non-salient variables in this study we can come to a different conclusion than Trudgill did in his study. Rather than seeing Dylan as 'trying less hard to sound American' through his /r/-lessness in singing, for example, this is actually one of the places where he is *trying harder* to *not* sound American. That is, the backdrop to the lack of /r/s is one of AmE-influenced vowels.

When asked for examples of differences between singing and speaking, Dylan said that the "first one that springs to mind" is the GOAT vowel. This also happens to be the only vowel in which Dylan uses NZ-like variants on a number of occasions. On these tokens, he seems to be making an effort to project a local identity. The large majority of his sung GOAT vowels are very different to their spoken counterparts, retracting and rounding, rather than fronting, and the overall difference between conditions is significant. GOAT seems to represent a phono-opportunity (Coupland, 1985) for Dylan to occasionally project his New Zealand-ness, it is something he has caught hold of and

learnt to deploy for stylistic purposes. Upon listening to other songs on Dylan's album, it became apparent that he uses fronting GOAT quite frequently. Interestingly, he tends to use the fronting variant more in verses than in choruses, suggesting that verses are perhaps closer to his speech style than are choruses, as was suggested in section 4.5.1. The chorus of the target song 'Sold it all away' is the most overtly AmE sounding section of the song, with the long PRICE nuclei, and the retracted GOAT vowels.

There are a couple of other cases of distinctively NZE-influenced singing aside from the fronting GOAT vowels. One such example occurs in a line in the first verse of 'Sold it all away'. The phrase "and I'll follow my baby" contains a series of NZ-like features. Firstly, the *and* is Dylan's most raised sung TRAP vowel (the token *$ and2* in Figure 4.4). Next, we have Dylan's most raised sung LOT vowel in the word *follow* – unlike most of Dylan's LOT vowels, it also is somewhat lip-rounded, which may be the reason for its relatively low F2 (see Figure 4.10). The GOAT vowel in the second syllable of *follow* is one of Dylan's fronting sung GOAT vowels. And finally, Dylan sings /bəˈbi:/, not /bəˈbə/ (which would be the typical American sung pronunciation), which comes across as distinctively NZE-sounding.

In one of Dylan's other songs (not analysed acoustically), 'The Skies above Crete', Dylan makes a definite move *away from* an AmE style when he sings consistently in each chorus /kɑnt/, with a very open, retracted DANCE vowel. Perhaps British associations caused by the World War II subject matter may play a role in this BritE-like vowel. It could also be taken as a case of qualitative overshoot (Gibson & Bell, fc), where a hyper-performed non-American DANCE is used, being much further back than the [a] used for DANCE in spoken NZE. It is of interest that each occurrence of this word is directly after the line "search for your taniwha[1]", a bold claim of New Zealand or, more specifically, pro-Māori identity. This kind of co-occurrence of identity claiming features in a short passage, which occurs in both of the above examples, is something which was discussed in Bell (2001) as a common aspect of initiative referee design. It should be emphasised that it is the NZE-like, not the AmE-like segments which seem to require active stylisation.

[1] Māori word for a mythical monster.

5.1.2. Andrew

Andrew Keoghan expressed no desire to use NZE pronunciation in his singing. There are no traces of NZE singing from Andrew, his pronunciation style seems to be fully in line with the American-influenced norm. The comparison of his speech and singing revealed large and consistent differences across all of the vowels studied, not just those which have some status as stereotypes of AmE. Andrew is the least aware of specific segmental features of his singing accent and he produces the most consistently American style of pronunciation. This result would not be expected if singers are assumed to use an American accent as an act of identity, constrained by lack of ability to identify their target group, and lack of exposure to that group. The differences between singing and speech are similar for Andrew and Dylan, except that all of Andrew's GOAT vowels retract. Use of non-prevocalic /r/ in Andrew's singing style sets him apart from Dylan and John and reflects his lack of any anti-AmE or pro-NZE sentiment.

The moments that stand out as salient to me are those which are particularly American, that is, where a more neutral variant might have been used. The use of non-prevocalic /r/ is one example (though use of /r/ is still at a fairly low rate overall), and there are two instances of non-prevocalic /r/ which are particularly noticeable, clustering together in the phrase "her hair hangs". The reason the non-prevocalic /r/s are noticeable is because they are very long (the /r/ segments are 275ms and 173ms). Another interesting feature in this line is Andrew's 'unmerging' of NEAR and SQUARE (this is similar to the findings mentioned in 2.3.3, about the singer from the Brunettes). The word *hair* (pronounced [hɛəɹ]) has a very open nucleus (F1=830), contrasting with its recited counterpart (pronounced [hiə]) which has a raised nucleus (F1=344), characteristic of the NZE NEAR/SQUARE merger. The other interesting feature of this phrase is the alliteration on /h/, which is emphasised in the delivery with 193ms, 175ms, and 151ms of glottal frication before the onset of voicing in the three words. The elongation of /h/ and /r/ in this line may be examples of the sorts of features which made Andrew refer to his performance of this song as being a bit too "edgy" and not quite himself.

The similarity of F1 for all of the open vowels may be a sign of Andrew's classical singing training, with a fairly stable jaw position for these vowels. The overlapping realisations of THOUGHT, LOT, and START may also relate to the merger of all of these vowels in American Englishes. If this is the case, this is a very sophisticated application of the 'rules' of the target variety.

Andrew's experience with virtuosic singing in both classical and jazz styles is noticeable in his performance of 'Gloria', evidenced by his large pitch range with frequent use of falsetto, and also through a nuanced and almost theatrical use of vocal dynamics and emotional expression. Having sung in a range of styles, it makes sense that Andrew takes on the singing accent most appropriate to the pop music style, in the same way that he would have adopted operatic pronunciation norms when singing passages from Puccini's Turandot.

5.1.3. John

John Guy Howell, of the Broken Heartbreakers, expressed a very positive attitude towards NZ accented singing, and stated that he had made a conscious decision to sing more in his speaking voice. However, these intentions are not clearly reflected in the results of the vowel production study. His DRESS and TRAP vowels are opener and more centralised in singing, his LOT and THOUGHT vowels are opener and fronter in singing, and are also unrounded. The movement of F2 in his GOAT vowels, like for Dylan and Andrew, is significantly different in the two conditions, with a retracted and rounded off-glide in singing and a fronted, less rounded off-glide in speech. Unlike Dylan, John has no clear exceptions to this pattern. GOOSE shows a similar pattern in terms of F2 movement, fronting in speech and retracting in singing. Finally, the nucleus of spoken PRICE is significantly further back than its sung equivalent.

There is a noticeable difference between John and the other singers in that his F1 is consistently around 200Hz lower than the average of Dylan and Andrew's F1s. There are several possible reasons for these F1 results, which require some detailed discussion to tease apart. I will consider this issue before discussing John's singing style more generally.

At face value, John's low F1 values appear to represent close vowels. In general, F1 is treated as being directly related to vowel height in terms of tongue position; however, this conceals some complexity which should be explored in this case. Firstly, the lower formant values could be caused by a vocal tract size difference between singers. This seems unlikely, however, since neither F1 or F2 in speech, nor F2 in singing are lower than the values for Dylan and Andrew.[2] It would be unusual for the effect of a larger

[2] This was ascertained through multiple regression models including singer, vowel and the singer by vowel interaction as independent variables.

vocal tract to be restricted to such a specific context. A second potential cause of the low F1 values is that John may have a tendency to nasalise vowels in singing. The effect of nasalisation on F1 is complicated, but for most of the vowels studied F1 would likely be lowered by nasalisation.[3] If this is the case, the F1 results could not be seen as representing vowel height in a straightforward way.

The F1 difference between singers would also be found if John had an overall tendency towards a more closed jaw position in his singing than Dylan and Andrew. Impressionistically, this seems like the most plausible explanation. Evidence for this view comes from the uniformity of the F1 difference between the singers, with John's whole vowel space 'shifted up' as compared to Dylan and Andrew, not just on vowels which showed F1 differences between singing and speech for the other singers, but for all non-high vowels.

If John does use a more closed jaw setting in his singing style than Dylan and Andrew, we must still consider *why* this is the case. Since jaw release is an important part of singing pedagogy, it could be related to the fact that John is the only one of the three singers not to have received any vocal training. While this may be the case, it is hard to ignore the fact that the F1 *distance* between singing and speaking is so much less for John than it is for Andrew and Dylan, especially in light of John's comments about having decided to sing more in his speaking voice. This jaw position brings the F1 values in singing closer to those in speech for some vowels (DRESS, TRAP, THOUGHT, LOT), though it actually *causes* an F1 difference between conditions for others (GOAT, START).

This suggests that the F1 differences between John and the other singers may not be related to differences between New Zealand-ness and American-ness, but rather to John having a less 'singy' style. There is other evidence for this view. John's singing exhibited the lowest average pitch and duration, and both of these had lower standard deviations. These factors all make his singing more similar to his speech. Also, as mentioned earlier, John's songwriting style tends not to have climactic chorus sections, instead using repeated catch lines, which are comparatively 'speaky'. It seems that John may enact his decision to sing more in his speaking voice in a very literal way, by

[3] This is due to an acoustic effect of the nasal cavity's first resonant frequency (see Kingston, 2007, p. 417, for details on the acoustic effects of vowel nasalisation).

making his singing style more like his speech in terms of jaw setting, and through aspects of his songwriting.

On the basis of the above discussion, we cannot conclude definitively that John is trying to project a NZ place identity by producing raised vowels, particularly since he follows the AmE-influenced pattern in other aspects of his pronunciation, notably the F2 movement of GOAT and GOOSE. There are only two other clear signs of non-American-ness in the four target songs for John. As with Dylan, these are for non-prevocalic /r/ and BATH. His singing style is /r/-less, and the single token of the word *ask* is produced with a START, rather than a TRAP vowel. The realisation of BATH words with TRAP is one of the most salient stereotypes of American-accented singing. As with Dylan, it seems that it is on the variables that are stereotyped as Americanisms where John is more able to portray his NZ place identity. This suggests that the AmE variants used on all the other vowel variables represent a kind of automatic default style, while NZE variants may require more effort.

As mentioned at the end of the results chapter, John consistently uses [ɪn] for (ING). When I asked John about this feature (after analysis), he said that this is a conscious choice because he dislikes the way [ŋ] sounds in singing. The use of [ɪn] may also act as an index of country music. The Broken Heartbreakers' other lead singer also consistently uses [ɪn]. It would be interesting to see whether this feature has genre-specific indexical ties.

It seems unlikely to me that the negative social meaning John has for sung [ŋ] is about the index between [ŋ] and NZ-ness. Rather, it is likely to be related to the negative relations between the velar form and formality, effortfulness and pretentiousness (Eckert, 2008). Each variant can have positive and negative forms, depending on the orientation of the speaker. The velar form can be either esteemed as articulate, or derided as pretentious. The apical form, inversely, is positively tied to unpretentiousness, and negatively tied to inarticulateness. In singing pop music, it is generally more important to project easygoing, relaxed and unpretentious than it is to project formal, educated or articulate. John's consistent use of [n] cannot be taken at face-value as a marker of American-ness. It does provide an interesting example of how the indexical meaning of a variant can shift according to the contextual frame. In this case, the use of [ŋ] in speech does not carry negative connotations for John, but in singing it does. This could be caused by associations with esteemed musicians;

pronunciation of (ING) as [n] is common in country music, for example. It could also be that while it is appropriate to project 'articulate' in speech, it might be more appropriate to project 'relaxed' in singing.

The song 'Angela', which was released on the Broken Heartbreakers' second album, stands in stark contrast to the four target songs in its pronunciation style. The opening line is "England is far away, why are you there? You've been away so long I forget to care". The PRICE vowels in *why* and *I* sound notably NZ-like; they are both realised as gliding diphthongs that have raised and retracted nuclei. The words *so long* cluster together as distinctive markers of NZ-ness, with a fronting *so* and a rounded, backish *long*. The line "London's a shitty town" features an aspirated, not flapped, intervocalic /t/, a saliently non-American feature. The use of NZE features continues throughout the song, particularly on the diphthongs, especially FACE.

There is a contrast between this song and the four songs which were analysed acoustically, and it comes across as performative and stylised. 'Angela' is an earnest song with personal lyrics about family and specifically NZ based references. The difference between these songs shows how important song topic might be. But it also demonstrates the way effort is required to use NZE features in pop singing. When I mentioned the difference between 'Angela' and his other songs, John said that the use of NZE in this song was intentional, and that the vocal had involved re-recording to get the NZ accent to "come through strongly". It is therefore interesting that he was quite successful in performing a NZE accent in this song, but only having practiced and having applied conscious attention to his pronunciation. This appears to be a case of what Bell (1984) called an inverted audience design, which is distinctive in requiring initiative style.

5.2. American-Influenced Singing as a Responsive Style

To summarise the discussion of each of the singers presented above, there are significant differences between singing and speaking for all of the vowels studied, demonstrating that the modification of pronunciation in singing is not restricted to variables that are salient as 'Americanisms'. This finding suggests that American accented singing is the default, automatic way to sing popular music. It is the singing of New Zealand variants that requires an initiative act of identity. This contrasts with Trudgill's (1983) description of singers trying to imitate American accents, 'putting on'

the accent, and is much more in line with O'Hanlon's (2006) observation that the American variants may actually be hard to 'take off'.

Below, I present four points which consider the results of the vowel production study with reference to the claim that singing with AmE-influenced vowels is normative, and to some extent automatic for at least the three NZ singers studied in this thesis.

- The singers use AmE-influenced variables overall, not just on stereotyped variables.
- For variables which do have a status as salient Americanisms, the singers with pro-NZE stances, Dylan and John, do not use the American variant. This ratifies their stated positions of wanting to project a NZ identity where possible.
- The interview responses in themselves provide evidence for this position, with Dylan stating explicitly that although he does not want to use AmE in his singing on ideological grounds, it is difficult not to. John states very clearly that he likes how NZE sounds in singing, and even that it might give the band a distinctive edge for overseas listeners, and yet states that he can't imagine fully singing in a NZE style.
- Perhaps most convincingly, when NZE is stylised in singing, it requires conscious effort of one kind or another. For Dylan, he is very aware of the differences between his sung and spoken GOAT vowels, and it is on this vowel that he has a few very distinctive tokens which match his spoken, NZE, style. For John, he re-recorded his vocal to a place-themed song, in order to get his accent to 'come through strongly'. This shows that for John, it is *possible* to use much more NZE style vowels in singing, but it requires focus and effort.

These questions about whether it is 'easy' or 'difficult' (to quote Dylan) to sing in AmE and NZE accents relate to Bell's distinction between the responsive and initiative dimensions of style. Table 5.1 is an adaptation of the table published in Bell (2001), which lists the terms used in different sociolinguistic frameworks to make a distinction between normative and innovative language styles.

Table 5.1: *Terms for Responsive and Initiative Style in Speaking and Singing Frames (Adapted from Bell, 2001)*

Frame	Responsive	Initiative	
Speech	Style	Stylisation	(Bakhtin, 1981 [1935])
	Situational	Metaphorical	(Blom & Gumperz, 1972)
	Audience design	Referee design	(Bell, 1984, 2001)
	Unmarked	Marked	(Myers-Scotton, 1993)
	–	Crossing	(Rampton, 1995, 1999)
	Relational	Identity	(Coupland, 2007)
Pop singing	Institutionalised referee design:	Inverted audience design:	(Bell, 1984)
	AmE	NZE	

At the bottom of the table, I have added my interpretation of the situation for pop singing in New Zealand, based on the evidence discussed above. What I am claiming is that singing pop music in an American-influenced style involves *responsive referee design*, while using NZE in pop singing requires *initiative audience design*.[4] This 'inversion' of the normal situation was discussed by Bell (1984) with reference to the use of local pronunciation features in British punk music. He described the use of American features in singing as a case of 'institutionalised referee design'. This term refers to situations where referee design, which is usually an initiative act in Bell's model, becomes so associated with a certain situation that it becomes responsive. A clear example of this occurs in diglossic speech communities, where different linguistic codes are used in distinct domains of language use.

Once institutionalised, referee design no longer needs to be *stylised*, it is not a case of *crossing*, since the speaker has full 'ownership' of the style in this restricted context, and it will not be *marked* in the same way that initiative referee design is likely to be. The use of AmE features by a New Zealander in the singing of pop music is therefore an unmarked style that is determined by the situation, rather than by a desire to change the situation. It is relational, in that it helps the singer to 'fit in', and finally, even though this style may not be based on the speech patterns of its audience, it is in a sense

[4] The table risks presenting the distinction between responsive and initiative style as if it were a dichotomy, which it is not – both sides of the distinction are available at all times to at least some extent in situated language use.

designed for that audience because it presents the style which they expect to hear. The important issue of listeners' expectations will be the topic of chapter 6.

5.3. Singing Inherent Effects

This section returns to the idea that singing and speech constitute very different vocal activities, and that differences between them may affect the pronunciation style used in popular music. It was stressed in chapter 2 that in the interpretation of these results, I should seek to distinguish the extent to which the differences found are to do with style as opposed to being caused by singing inherent effects. This section considers the possibility that such effects have caused the differences between singing and speaking that were found in the vowel production study. Two possible singing inherent effects have been discussed; the first being the higher pitch of singing and the second being the preference in singing for sonorous vowels.

It was hypothesised that all of the singers were likely to show a positive correlation between pitch and F1, on the basis of the findings of studies like Austin (2007), discussed in 2.2.2. The results partially supported this hypothesis, with Dylan showing clear evidence of having higher F1 when singing at higher pitches. It was established in section 4.5 that this effect was too minor to account for the large F1 differences seen between singing and speaking overall. But there is another singing inherent factor which relates directly to jaw position and vowel openness which may prove to be important.

Morrissey (2008) argued that there is a preference in singing for more sonorous variants. The American variants of most of the vowels considered in this study have an advantage over their NZE counterparts on sonority grounds. There are no clear cases of vowels which are more open in NZE than AmE, which makes it difficult to distinguish between social influences and the effect of sonority. The results of the comparison of singing and speaking presented in this thesis provide evidence that both forces are at work.

Some evidence that sonority plays a role in the overall tendency for vowels to be opener in singing than in speech comes from a comparison of Dylan and Andrew's sung vowels with reported values for spoken AmE vowels.[5] The F1 of the sung vowels is much higher than that found for spoken AmE. If the singers were emulating a spoken

[5] In Peterson & Barney (1952), Hillenbrand et al. (1995), Hagiwara (1997), and Clopper, Pisoni & de Jong (2005).

American accent it is unlikely that the vowels would be as open as they are. It is more likely that this vowel openness is related to singing technique than dialect or style.

Though the effect of sonority is likely to play some role, there is also overwhelming evidence that the differences between singing and speaking found in the study are motivated by identity concerns. Sonority is mainly related to F1 differences, so the differences found that relate to consonants, and to F2 differences are not likely to be affected by the singing inherent effects outlined above. Strong indications that sociolinguistic forces are driving the differences in F1 come from other non-sonority-related differences, in diphthong movement for GOAT and GOOSE, and the difference in F2 for the START vowel (which is significant for Andrew, and of a similar magnitude for Dylan though not significant). Additionally, the GOOSE vowel is not opener for singing than for speech, so it seems that if there is a sonority effect, it may only apply to non-high vowels. It is also of interest that for Andrew and Dylan, THOUGHT, LOT and START occupy a similar acoustic space in singing, even though they are distinct in speech. This may be an instantiation of the CAUGHT-COT merger which has occurred in most dialects of AmE and is likely to be reflected in normative popular singing accents.

One obvious American marker is found in the realisation of consonants, for which the variants are less distinct in terms of sonority. For Andrew, the use of non-prevocalic /r/ can be considered an Americanism, and for both John and Andrew the use of [ɪn] for (ING) is also likely to be based on the pop singing model, even if it is not an overtly American feature.

The fact that AmE vowels are generally more sonorous may in fact be one of the reasons why the AmE singing accent has had such a strong, ongoing influence internationally. It would also help to explain the opinion of several singers (e.g. in Coddington, 2004), and expressed in this study by Dylan, that “it just seems easier to sing in an American accent”. Because of this entanglement, we cannot easily resolve the degree to which the F1 differences between singing and speaking for New Zealanders are socially motivated stylisations (the emulating of American-ness) or the result of singing-inherent effects (the preference for sonority).

A preference for sonorous vowels in singing provides a motivation for sung vowels to be even more open than AmE speech, and it also may play a role in the enduring association of normative singing accents with America. AmE might be inherently more ‘singable’ than NZE. However, this is not a reason for either the F2 based differences, or the consonant differences. However, these may be mutually reinforcing factors.

The fact that singing may be more sonorous than AmE means that it is, by definition, not AmE, at least not spoken AmE. Rather, it is an American-influenced pop music singing variety of English, with its own features. I will use the term AmE-influenced, rather than AmE from now on, to acknowledge the role which singing inherent effects play on pronunciation.

5.4. The Indexicalities of Singing Styles

Thirty years after Trudgill first studied 'American' pop singing accents, and fifty years after some of the songs he analysed were recorded, the American model is still strong, and the question quoted at the beginning of this thesis is still open to debate:

> why do singers modify their pronunciation? (Trudgill, 1983, p. 143)

Developments in sociolinguistic theory have provided new tools for answering this question. As Beal (2009, p. 229) notes, Trudgill's arguments were "framed within the sociolinguistic theories available at the time of publication". At that time, dialects and registers of speech were seen as being associated with social categories and different situations of language use. Taking into account the insights of earlier sociolinguistics, we can now think of a speaker's linguistic repertoire as containing a vast set of linguistic resources which are available to index a wide range of social meanings, in the situated practice of active contextualisation. The web of associations between phonetic forms and social meanings constitutes an indexical field (Eckert, 2008) which is always in flux, changing as situations change, and according to differences in the "perceptions of 'the same' situation across cultural or social groups, across individuals and for the same individual over time" (Coupland, 1985, p. 155).

In this view, linguistic variability is a resource available to be actively deployed to achieve identity goals, whether those goals are to conform to expectations or to negotiate new identities. From this perspective, our questions change. Having established that the Beatles often flap their /t/s, use non-prevocalic /r/s and realise PRICE as a monophthong, then, we might not be inclined to ask "why should singers attempt to imitate what they consider to be an American accent" as Trudgill (1983, p. 144) did. Instead, we might ask: what do these features *mean*, both in their own right and in their co-occurrence with one another; and what have they meant for different singers and listeners at different times?

I return now to the notion of framing introduced in chapter 2. Different aspects of identity are salient in different frames: The "identificational value and impact of

linguistic features depends on which discursive frame is in place" (Coupland, 2007, p. 112). I argue in this section that the indexical meanings of AmE-influenced features may be rather different in singing than they would be in speech. For example, an open, unrounded LOT vowel might index 'American' in a conversation, but index only 'pop singing' in a recorded song.

Related to this is the role a speaker plays in a given frame, and the stances they take:

> Stance and role in social identification are often more appropriate concepts than identity: 'in which persona am I to approach this communicative event?' (Coupland, 2003, p. 426).

This is a useful way to look at the differences which occur between singing and speech. Pop singing style is constructed as both an act of identity towards 'pop singers', and as a context-dependent response to an established communicative genre. Stances for pop singers might relate to various music styles, to a distinction between professional and amateur, commercial and indie, serious and comical. The latter may be important since local dialects are often used in singing for comic purposes, and this may be something that a lot of singers wish to avoid in their music.

The previous section established that the results for Dylan and Andrew, with higher F1 in singing than speaking, are likely to be related to both a dialect based style-shift and an effect of sonority. It is of interest to consider the impact of this conclusion on the idea of singing accents being 'American'. Firstly, consider the question of the esteemed group with which the singers wish to identify. This has been described in the literature on singing accents to be the group 'Americans' for many artists. The data suggests that a more accurate description of the model group is 'popular singers', whose country of origin is to some extent irrelevant in the pop singing contextual frame. For example, ABBA may be Swedish, but they are still, no doubt, influential pop singing role models for many other singers. This foregrounds the specific nature of the role models, who are not Americans, but rather pop singers; thus backgrounding place identities.

The reason there are similarities in the pronunciation style of so many popular singers of various nationalities is that "the first genuinely international popular music" was American (Potter, 1998, p. 151). The continuing use of AmE-influenced features is therefore an epiphenomenon, the influence of influences (Coddington, 2004; Simpson, 1999). Coddington argued that rather than viewing singing pronunciation as a shift from one's spoken accent towards a foreign target, "it may be more appropriate to view singing pronunciations as 'accents' in their own right" (Coddington, 2004, p. 86),

although they are originally derived from a common source. The idea of defining 'singing accents' in their own right is important, however it could be in danger of treating these 'singing accents' as static entities. A more effective approach might be to try and identify the sorts of indexical meanings conveyed by variables in the pop singing frame.

The view that popular singing is based on prior recordings, rather than on American speech, or on variants that are stereotypically associated with America, sheds new light on Trudgill's (1983) arguments, essentially removing the obstruction created by Le Page's first two riders; ability to identify the model group, and sufficient exposure to the target variety. If the model group is 'pop singers' not 'Americans', then the heterogeneity of AmE dialects is not a problem. A pan-American singing accent may have been established in early recordings of popular music, while changes in sung accents over time are likely to relate as much to music style as to the dialects of singers. The American-style pronunciation of popular music may have undergone a process of focusing over the decades, as well as various genre-related divergences. This would allow singers to emulate the sung accent itself, rather than appealing to stereotypes of American speech. Additionally, positing 'pop singers' as the model solves the problem of singers having insufficient access to the target variety. If the target variety is that which they have heard on records and radio since childhood, then their exposure is much greater than if the required exposure is considered to be spoken AmE.

This point of view opens up the empirical question of what the typical vowel realisations actually are for a normative popular singing accent. Whether or not Dylan and Andrew's vowel realisations represent a normative 'pop singing' variety can only be verified through the study of a much wider range of artists including mainstream American artists.

The significance of the seemingly subtle distinction between indexing AmE features to 'American' as opposed to 'pop singers' becomes more apparent when we consider the metapragmatic ideologies that singers may have about their singing pronunciation. For Dylan, the social meaning of backing GOAT vowels may have switched from 'singing' to 'American'. That is, the meaning of the differences between forms shifted from being one of register to being one of dialect and became ideologically significant. Since Dylan finds it 'painful to blatantly sing American vowels', this shift in social meaning could have quite an effect on his construction of style. Other 'American' variants, such as open DRESS and TRAP, may continue to be indexed to 'pop singers' and not to

'Americans', and consequently have less of an ideological sting to them. The degree to which singers associate AmE-influenced features with 'fakeness' is likely to greatly affect their singing pronunciation. Arctic Monkeys are the extreme case, where they have knowingly (and probably effortfully) constructed a distinctly non-American singing accent in order to set themselves apart from the mainstream, to show they are not 'following a handbook' (Beal, 2009, p. 225). In the final section of this chapter, I will consider how indexicalities come to be, and provide a short case study which considers this question from the perspective of the listener. Before this, however, some limitations of this vowel production study should be mentioned.

5.5. Limitations

One of the most important limitations of this study is that there is a lack of comparative data available for influential American pop singers. While vowel formant data has been published for spoken American English, no such data has been published for commercial American pop singers. The analysis of the New Zealand singers studied here would have benefited greatly from such a comparison, which would make it much easier to distinguish singing inherent effects from dialect based stylistic choices.

Assessing a singer's awareness of their singing accent is an important part of the study of singing pronunciation. Singing pop music is an inherently performative activity, which is likely to involve conscious stylisation. The design and framing of this study did not allow me to deal with these important issues in depth. Asking the singers 'if they've thought about their singing accent' and 'are there any specific examples they can think of' was a rudimentary and problematic way to explore the singers' metapragmatic awareness. It would be better to try and access this information through indirect means. Perhaps asking singers to recite their lyrics in an American accent, or to sing their song in a NZE accent, would provide insights about the degree to which singers have conscious control over specific variables and linguistic styles.

Another limitation of this study has been the lack of engagement with musicological analysis and cultural studies of popular music. The insights of these disciplines will be particularly important as we try to uncover the way linguistic styles contribute to music genre styles. It is clear that musical genre is important to variation in singing pronunciation. It may actually be one of the most salient dimensions of a singer's construction of voice in singing, drawing on and actually superseding the social meanings of regional and social variation. If that is the case, then this study has been

lacking in its consideration of musical genre. Future studies of singing style should consider the visual images (such as clothing and CD cover design) used by the singer in their construction of style, and the actual form of their music in terms of lyrics, harmony, rhythm and production.

5.6. Memory: The Underlying Mechanism?

The ideas presented earlier, about the way indexical fields shift according to contextual frames, raise the question of how the indexicalities come to be in the first place and how indexing might actually work. Presumably, the key concepts here are memory and exposure – people learn patterns and associations because things show up together. If /r/ occurs consistently with commercial pop artists, it begins to 'sound' commercial.

To conclude this chapter, and provide the rationale for the next, I present a small case study of the indexicality of the GOAT vowel, which as we have seen, fronts in spoken NZE, and retracts in the pop singing of at least the three artists studied. If we simplify the phonetics of the situation for the purposes of a quick investigation, and consider fronting and retracting GOAT vowels as being two distinct phonetic variants, then we can ask what the indexical relations of these two variants might be in terms of national identities, music genres and stances towards the mainstream.

Consider the sorts of indexical associations about retracting GOAT vowels which might emerge from hearing the following artists use that variant. All of the underlined examples in the following list have retracting, rounding GOAT vowels:

- Frank Sinatra (USA), 'Chicago';
- Michael Jackson (USA), "we've got so far to go", in 'Rock with you';
- The Beatles (UK), 'The Long and Winding Road';
- Pink Floyd (UK), "we don't need no education", in 'Another Brick in the Wall';
- Crowded House (NZ, commercial), 'You'd better be Home soon';
- Bic Runga (NZ, commercial), "don't come and go", in 'Sway';
- The Muttonbirds (NZ, iconically "Kiwi", sings about local roads and bridges), "she'd seen me come alone" in 'She's been Talking';
- Chris Knox (NZ, indie, one of the founders of the 'Dunedin sound'), "hello my friend", in 'Not Given Lightly'.

Based on the list of retracting GOAT vowels above, and assuming the examples given are relatively characteristic of the singing style of each artist, what can we say about the indexicality of a retracting GOAT vowel in pop singing – what does it mean? Surely not "Americans", not even "*mainstream* pop". The one thing which these examples have in common is that they are sung. Retracting GOAT may therefore most strongly index "pop singing from various countries, in various genres, with varying degrees of commerciality".

Can we say that singers are *putting on* an American accent when they sing a retracting GOAT vowel? Firstly, as has been demonstrated, it is not accurate to call this an American feature. Furthermore, if 'putting on' means something like stylisation, or initiative design, then are they really putting anything on at all if it is what comes most easily and naturally to them? Are they not just being responsive, appropriate, and conventional?

It seems that in singing popular music the normative American-influenced accent is a strong force for these three singers. The singing of pop music involves a highly ritualised and conventionalised language style that is used in a very specific set of contexts. To use a NZE accent in these contexts has a hint of scandal to it, perhaps a remnant of the cultural cringe. If a singer breaks the accepted frame it is very salient, and loaded with meaning.

Bakhtin's assertion that words taste of the history of their use (Bakhtin, 1981 [1935]) is perhaps an important key to the mechanisms which give context-specific language styles their power. The next chapter introduces an account of language which attempts to describe the processes which might allow individuals to acquire this 'taste' for the indexicalities of socially contextualised language. The direction I propose involves turning to the mechanisms of language perception. The next chapter therefore constitutes a major change in focus.

Chapter 6 A Different Perspective: Perception of Speech and Singing

The acoustic comparison of singing and speech presented in the last three chapters has hopefully provided some insight into the causes of American-influenced singing styles. The singers who want to use a NZE accent in singing on identity grounds find it hard to do so in practice, which suggests that AmE in pop singing is a responsive style, not an initiative one. However, this is a difficult argument to verify by studying production data alone, because we cannot be entirely sure that singers' self-reported attitudes and motivations are the same as those which operate in the construction of their singing style. In order to understand why a 'proud New Zealander' would use an AmE-influenced singing style, we need to investigate the issue of that singer's prior exposure to sung forms of language. It seems likely that the reason it is difficult to use NZE in pop singing is because there is a relative lack of previously heard examples of NZE in the pop music context.

From this perspective, it is in their role as music *listener* that singers become predisposed to the use of AmE-influenced features in their own singing. An examination of music *listeners* could therefore be a useful line of enquiry; one which draws on an alternative, complementary, methodology to investigate the causes of pronunciation differences between singing and speech.

In the last decade or so, an increasing number of sociolinguistic studies have focused on speech perception rather than speech production to try and get a different perspective on the mechanisms behind social and stylistic phonetic variation. As will be discussed below, these studies have found that listeners' perception varies according to social factors, such as the identity traits of the speaker. One way to investigate the claim that AmE-influenced singing is automatic rather than effortful is to show that ordinary listeners perceive vowels in an 'American-influenced' way when listening to pop music. If that is the case, it provides evidence that these variants are the most expected linguistic forms in this context, that these variants are the default.

By taking this approach, this chapter marks a turning point in my investigation of phonetic differences between singing and speaking. On one level, it is a move from sociolinguistics to psycholinguistics, and on another, it is the move from looking at the outward manifestations of sung and spoken vowels to an exploration of how these

vowels might be stored and accessed in the mental lexicon. The first half of the chapter introduces a range of literature on speech perception, and in the second half of the chapter, I describe an experiment which explores the perception of vowels in musical and non-musical contexts.

6.1. Theories of Speech Perception

This section returns to the review of literature begun in chapter 2. The material to be reviewed is of a different enough character that it was deemed preferable to deal with it here, in its own right, rather than including it at the beginning of the thesis. The body of research discussed requires the introduction of terms which have not yet been used, and involves something of a change in writing style. This change in style is necessary, however, in order to contextualise the perception experiment presented in the second half of the chapter. Section 6.1.1 briefly describes the theoretical landscape, followed in sections 6.1.2 and 6.1.3 by a more detailed presentation of the literature on exemplar theories of speech perception.

6.1.1. Dealing with variation in speech perception

Most popular models of speech perception have been influenced by generative linguistics, along with its focus on the abstract rules of language and its dismissal of language use in situated contexts. This perspective is associated with the notion that the main objective of speech perception is the conversion of speech in its raw acoustic form to propositional utterances made up of discrete phonemes. Once spectrographic study of speech became predominant in the 1950s, it became clear that there is no strict sound to phoneme correspondence. The invariant phoneme, as represented in the grammar, can be realised in connected speech in a multitude of ways, and its realisation overlaps with that of its neighbouring phonemes (that is, there is *co-articulation* between neighbouring segments). Variability in the speech signal, which is often referred to as a *lack of invariance*, also arises from differing vocal-tract lengths between speakers, and from differing speech rates. And beyond all of these factors, there is dialectal, social and stylistic variation. The gruelling task of a theory of speech perception in the generative setting, then, is to determine how to filter through all that irrelevant noise to get to the invariant phonemes (or sets of distinctive features) that were intended by the speaker. This filtering mechanism is referred to as speaker normalisation, and has been the central concern in many studies of speech perception.

One solution to the problem of the variability in the acoustic signal is to posit that the acoustic signal is not represented in the speech processing system at all. In both the motor theory of speech perception (Liberman & Mattingly, 1985) and the direct realist model (Fowler, 1986), speech perception uses the acoustic input as a superficial messenger signal which allows the listener to determine the speaker's gestural intention. That is, speech perception is achieved by using sound as a cue to the articulatory gestures which created it. Motor theory deals with the lack of invariance in the acoustic signal by rejecting the importance of the acoustic signal itself:

> Association of some particular [acoustic] cue (or set of cues) with a phonetic category will work only for a particular circumstance. When circumstances change, the child's identification of the category will be wrong, sometimes grossly, and it is hard to see how he could readily make the appropriate correction. Perception of the phonetic categories can properly be generalized only if the acoustic patterns are taken for what they really are: information about the underlying gestures. (Liberman & Mattingly, 1985, p. 25)

This is an extreme view, in which the acoustic manifestations of a phoneme are not represented at all in the lexicon. Most other theories of speech perception argue that it is the acoustic signal that is used to access words in the mental lexicon.[1] I will now turn in more detail to these perspectives, which offer an alternative to motor theory, making the assumption that the information stored in speech perception is acoustic, not gestural. As mentioned above, most of these theories treat variability in the speech signal as a problem which must be resolved in order to uncover the abstract, invariant segments required in a generative grammar.

A large portion of this branch of speech perception research focuses on the specific process of spoken-word recognition. The acoustic signal is seen as activating lexical items that are stored in the mental lexicon. Theories differ in how they describe the form of the lexical entries themselves, and in the mechanisms they propose to trigger the perception of one word over another, but most theories describe the process of lexical access using some version of the concepts of *activation* and *competition* between potential candidates.

Marslen-Wilson's (1987) cohort model of speech perception, for example, states that "each individual entry in the mental lexicon is assumed to correspond to a separate computationally active recognition unit" (p. 78). That unit includes acoustic-phonetic,

[1] There is evidence from neuroimaging studies which suggest acoustic, rather than motor, representations of speech. See Coleman (1998) for a review.

syntactic and semantic information, but not social or contextual information. Speech perception involves the matching of incoming information to these recognition units:

> A lexical unit is assumed to become active when the sensory input matches the acoustic-phonetic pattern specified for that unit. The model prohibits top-down activation of these units in normal word-recognition, so that only the sensory input can activate a unit (p. 79).

The cohort model emphasises the 'left-to-right' nature of word recognition with the onsets of words having privileged status, as they activate the initial cohort of possible words. Of the initial cohort, the perceived word will be selected using both bottom-up and top-down (syntactic and semantic, not socio-contextual) processes.

The TRACE model of speech perception (McClelland & Elman, 1986) is an influential connectionist implementation of speech perception, which applies several facets of the cohort theory. Connectionist models treat speech perception as a process of spreading activation through a computer-like parallel distributed processing system, with connections between nodes having variable weights. One of the attractions of connectionist models is that they are slightly more plausible with regards to actual brain processes than other approaches; a quality which has led connectionism to the forefront of cognitive science. The TRACE model uses the variation in the speech signal rather than just stripping it away: "the perceptual system uses information from the context in which an utterance occurs to alter connections, thereby effectively allowing the context to retune the perceptual mechanism on the fly" (McClelland & Elman, 1986, p. 6). This retuning of the perceptual mechanism allows the acoustic signal to be converted to the abstract phonemes which underlie words. This kind of perception has at its heart the goal of ascertaining the propositional meaning of a speaker's utterance. Even though contextual information is used to achieve that goal, the variation is not treated as important in terms of conveying social meaning.

A more recent connectionist model of spoken-word recognition, Norris & McQueen's (2008) Shortlist B, is different from the TRACE model in several ways. It is based on probabilistic Bayesian decision principles, where the selection of a lexical item is a process of hypothesis testing, rather than the whittling down of a list of competing possibilities. Also, while still postulating abstract pre-lexical categories, Shortlist B models these units as probabilities of multiple phoneme combinations rather than a sequence of discrete phonemes. This perspective marks a change in the landscape of speech perception theories and in the understanding of the linguistic system more generally. There has been a shift towards the view that the language system is not

determined genetically by the setting of parameters in a universal grammar, but rather that it emerges using more general brain functions, with particular focus on the brain's ability to calculate ongoing probabilistic functions.

This overview of different speech perception and spoken-word recognition theories is not intended as an exhaustive or detailed review, but rather to provide a backdrop for the approaches to speech perception presented in the next section. These *exemplar* approaches view variability in the speech signal as an important part of speech perception, not a problem to be surmounted.

6.1.2. Exemplar theory: Perception without normalisation

A series of rigorous and influential experiments carried out in the 1990s provided empirical evidence against the long-standing view that spoken-word recognition involves a process of normalisation in order to get at pre-lexical segments, such as phonemes or distinctive features. Goldinger (1996) provides a review of this work as well as presenting several new experiments. The studies found evidence for long-term memory of the surface details of a particular voice. For example, when asked to identify whether a test stimulus had occurred in a previous task or not, participants responded more accurately when the test word was presented in the same voice as in the original task. If speakers filter out voice information when processing speech, this memory for specific voices should not occur. The effect has been found in a range of tasks that explore both explicit and implicit memory, with different levels of processing of stimuli, and for words in different listening conditions, such as white noise.[2]

Explicit memory is accessed in word recognition tasks like the one described above. This technique draws the subject's attention to the remembering of that word by asking them whether or not they heard it in the original task. Implicit memory is examined by seeing whether the previous exposure to a word improves performance on lexical decision tasks. In lexical decision tasks, participants label stimuli as either words or non-words. They can identify a stimulus as a word more quickly if an earlier presentation of that word used the same voice.

The basic finding of these experiments, with several variations according to task and attentional factors, is that speaker-specific surface detail of previously heard lexical

[2] The neurolinguistic studies reviewed in Coleman (1998) also suggest that lexical representations include phonetic detail, rather than abstract phonological units.

items is stored in memory. This memory has been shown to persevere at time scales of up to a week. Longer time scales have not been tested, but there tends to be a forgetting curve as time from initial exposure increases. Memory of surface details is more apparent in implicit tasks than explicit ones. However, explicit memory tasks vary according to the *level of processing* of the first instance of the item. When an item is processed in a more semantic way, the surface details are not retained in memory for as long or as strongly as when the item is processed with an emphasis on the surface form.

These findings are difficult to account for in a theory of speech perception that views the propositional meaning of an utterance as the only goal of speech perception. In such models, all supposedly irrelevant phonetic variation should be removed in the process of normalisation, and should not be stored in long-term memory. Exemplar theories of speech perception provide an alternative stance, stating that instances of language are stored in memory including all the acoustic details which other theories would have normalised away. The repetition of the same voice saying the same word re-activates the memory of the original stimulus.

Before going into any detail about exemplar theory, which hinges on the notion of episodic memory, it is worth providing a very brief description of some of the ways memory is taxonomised in psychology. For a start, there is a basic distinction between *procedural* memory and *declarative* memory. Procedural memory involves the learning of skills and habits. It is the kind of memory that allows you to maintain the ability to ride a bicycle. Declarative memory can be further divided into *semantic* memory and *episodic* memory. Semantic memory could loosely be described as 'knowledge', while episodic memory allows you to recall specific events, or episodes. The unit of episodic memory, according to Tulving (1983), is an act of remembering. An episodic memory is 'relived', whereas a semantic memory does not necessarily include any details about the circumstances in which the memory was stored. For an episodic memory, the recollective experience is the product of not only the memory trace as it was encoded at the time of the event, but also of the conditions of retrieval. Furthermore, the actual encoding of an episodic memory trace is modulated by attention, with the most attended to dimensions of the cognitive scene being most strongly laid down in memory.

Exemplar theories of speech perception (and production) argue that at the core of any person's language system is a large collection of episodic memories of linguistic experiences which are rich in detail (Johnson, 2006; Pierrehumbert, 2001). Both fine-grained acoustic details as well as non-linguistic details may be retained in the

memories, determined by the degree of attention paid to them at the time of the encoding event. As an individual's experience with structured variation increases, exemplars will form distinct clusters which permit indexical categories to emerge (Foulkes & Docherty, 2006). Indexical relations can emerge between language forms and any kind of social or contextual information which is perceived by the subject as a coherent entity. The categories of linguistic structure, then, emerge from linguistic experience rather than being precursors to linguistic competence. The contexts of language use are an essential part of the speech perception:

> A word could be identified to a previous occurrence of a word in a similar context, from a similar source and in a familiar format, rather than by reference to a generalized representation of the word. (Jacoby & Brooks, 1984, p. 3)

Exemplar theories need to show how different episodes are activated by stimuli in an environment, and also how categories can emerge from experience. These issues are addressed by Hintzman (1986) in his presentation of a simulated model of perception, MINERVA 2. In this model, "A retrieval cue contacts all traces simultaneously, activating each according to its similarity to the cue" (Hintzman, 1986, p. 411). The objects in a perceiver's environment are recognised according to how similar they are to previously encountered objects, by activating memory traces of similar objects in similar contexts. MINERVA 2 results in a parallel, distributed response which is dependent on the similarity of the retrieval cue to the relevant dimensions of encoded episodes.

Categories are not specifically stored in memory; instead a category is "a temporary, dynamic structure that springs into being when a retrieval cue occurs". When a cue is similar to many stored exemplars, the commonalities between these exemplars will be apparent in the response to the cue, masking variation and giving the impression of an abstract representation. Thus, while a word can be conceived of as a coherent category through the general similarity across its occurrences, those episodic occurrences persist in memory, and function to both deal with and use contextual variation:

> "The word *eat* could be said to be represented by a very large number of traces that can be activated in parallel, but not all of these traces will be strongly activated in any one encounter with the word. Any particular active subset of these traces is a biased sample from the population, and the active subset will vary in both size and content from one occasion to another, depending on the context in which the word appears". (Hintzman, 1986, p. 423)

This description of language provides a different perspective on the investigation into pronunciation differences between singing and speaking. In singing, the 'active subset'

of memory traces for a given word will be those instances which were sung. Indexical relations form on the basis of associations in memory, and these memories can include details about pronunciation. Resonating with the theories of style presented in chapter 2, the exemplar perspective focuses on the situated context of a language event, and it provides an elegant way to conceptualise the relationships between language forms and social meanings. The compatibility of social constructionist approaches to style and exemplar approaches to language perception and production is a recurring theme in the growing sub-field of sociophonetics (see for example Drager, 2009; Foulkes & Docherty, 2006).

The activation of contextually relevant stored exemplars provides a frame of reference through which to perceive the language event that is occurring at any given moment. The context is defined as whatever dimensions of the environment the perceiver is paying attention to. It includes not only the 'outside world' but also the perceiver's own cognition, thus setting up the potential for reflexivity, performativity and metapragmatic awareness. The next section considers experiments which have explored the way context and social information can affect perception.

6.1.3. Context and social information affect perception

Several studies have found that perception can be affected by listener expectation. In an identification task, Ladefoged and Broadbent (1957) found that by altering the formants of an introductory carrier sentence the perception of a following word could also be altered. For example, participants categorised a stimulus as *bit* when F1 of the carrier sentence was high, and as *bet* when F1 of the carrier sentence was lower. Fry, Abramson, Eimas, & Liberman (1962, pp. 174-175) discuss these results in terms of a listener's shifting frame of reference for vowel perception:

> These results not only demonstrate an effect of context on vowel perception but also support the view which has been generally held for a very long time that in dealing with vowels uttered by a particular speaker, listeners rapidly form an appropriate reference frame against which they judge the quality of and identify the sounds which occur.
>
> …The important point is that the particular phonemic category selected is dependent on context, that is more specifically on the vowel reference frame which is operative for the listener at the time of reception.

The perceptual reference frame can be conceived of in terms of patterns of activation in stored exemplars. A range of experiments have strategically manipulated context to gauge its effect on perception. The most salient aspects of context in remembered

episodes of speech are likely to be those which relate to people, so the studies presented below have all manipulated social information to see how it affects speech perception.

Hay, Drager, & Warren (in press) have examined the perception of words in the NEAR and SQUARE lexical sets, which are merged for younger speakers of NZE. In one experiment, participants performed better at discriminating NEAR and SQUARE words when they had interacted with an American experimenter than when the experimenter had been a New Zealander. A second experiment found that participants were better at assigning NEAR and SQUARE words to their lexical sets if the experiment's instructions were given in a British English accent instead of a NZE accent. The authors explain these results using an exemplar account. Exemplars associated with (unmerged) British and American speakers are activated by the exposure to those dialects. The activation of exemplars of distinct realisations of the vowels aided performance in the perception tasks.

Warren, Hay, & Thomas (2007) report on another set of experiments on NEAR and SQUARE, using a semantic priming methodology.[3] In New Zealand, older speakers tend to keep NEAR and SQUARE separate, while for younger speakers these words are homophones, merging on NEAR. This means that while young people in NZ may sit on a [tʃiə], their parents sit on a [tʃeə]. Neither the young person nor their parents would [tʃeə] at a rugby game. Warren et al. (2007) exploit this asymmetrical situation.

In a semantic priming task with a young voice and young participants, the stimulus [tʃiə] decreased reaction times to the word *shout* (through association with *cheer*). Interestingly, it also primed *sit* to almost the same extent (through association with *chair*). When these participants heard the stimulus [tʃeə], however, *sit* was again primed, but *shout* was actually inhibited, with longer response times than in a control condition without priming. A second experiment repeated the methodology with an older sounding voice (and a new set of young participants), and achieved different results – the asymmetry in priming effects was not found. Each prime facilitated the semantically appropriate word (*cheer – shout*) more than the semantically inappropriate word (*cheer – sit*), and to a similar extent.

[3] Semantic priming experiments measure the effect a given prime word can have on participants' reaction time in a lexical decision task, that is, their speed at assigning the next word they hear as either a word or a non-word. Hearing the word *sun*, for example, will make participants respond more quickly in saying that *moon* is a word.

These results provide a demonstration of the way social information can affect perception. In the experiment with an older sounding voice, "listeners expect the speaker to be a non-merging one" (p. 103). When listening to an older speaker, the phonetic form [tʃiᵊ] does not activate the lexical item *chair* like it does when listening to a younger speaker. A different set of phonetic memories are activated according to the social characteristics of the speaker. This result is relevant to the present study because the same sort of principles that apply to differentiating between a younger and an older speaker should apply to the differentiation between a singer and a speaker, since memories of singers and speakers involve consistently different phonetic realisations.

Other studies have found that even without preceding linguistic material, listeners' expectations about the social characteristics of a speaker can affect perception. In a task where participants matched the vowel they heard in a sentence to another in a resynthesised vowel continuum, Niedzielski (1999) found that participants responded differently according to whether they were told that the speaker was from Detroit or from Canada, according to their stereotypes about differences between Detroit and Canadian accents. Hay, Nolan & Drager (2006) extended this study, looking at listener expectations based on cross-dialectal differences. They found that listeners heard more Australian-like KIT vowels when their response sheet had the word "Australian" rather than "New Zealander" written at the top of it. To ensure that these results were not an artefact of the overt labelling of varieties on the response sheet, Hay & Drager (in press) repeated the experiment replacing the labels with an incidental stuffed toy being placed on the desk where the experiment was conducted. A stuffed kangaroo (indexing Australia) caused participants to perceive more Australian-like KIT vowels than did a stuffed Kiwi (indexing NZ). These results are described in terms of activation of exemplars associated with the relevant speaker groups.

While the above research has shown the effect of expectations about speaker dialect, other studies have found differences in perception according to expectations about the social class and gender of a speaker. For example, Johnson, Strand, & D'Imperio (1999) found that the boundary between two phonemes differed as a function of the gender of a person in a photo presented at the same time as the spoken voice. This effect even occurred when there was no photo, and participants were instructed to imagine the gender of the speaker.

All of these findings suggest that speech perception is not merely about filtering out variation in order to uncover the propositional meaning of a speaker's utterance. The

variation between speakers, rather, is an essential part of the language system, allowing listeners to determine not only the propositional meaning of an utterance but also the social meanings which are carried in phonetic detail. All of the studies outlined above have shown effects relating to the *social* characteristics of a speaker. Social information about speakers is likely to be one of the most attended-to aspects of the environment during an interaction. However, exemplar theories predict that clusters of exemplars can be indexed to any aspect of the cognitive scene that is salient at the time the speech is heard.

When one listens to a pop song, the fact that the lyrics are being sung not spoken is important and salient. The fact that it is language in music makes it distinctive and it is therefore likely that exemplars of sung vowels will form clusters on the grounds of their contextual similarity. The second half of this chapter describes an experiment designed to investigate this claim.[4] If it can be shown that perception differs in the context of music, this will shed some light on the results of the vowel production study. The singers that were investigated do not only sing – they also listen to singing, and it is in that listening that the indexical associations between phonetic forms and social meanings would have initially developed during language acquisition. If there is a strong link, for example, between pop singing and high F1 values in DRESS and TRAP, then this link should be known not only by the singers who perform it, but also by their general listenership. And if this link is entirely normative, then it should not take an initiative act of identity to be performed; it should be the responsive, 'natural' realisation; part of the linguistic routine that is pop singing.

6.2. Perception Experiment

When the introduction to a pop song is playing, listeners develop expectations about what they will hear. Memories of similar music heard in the past become activated – and these memories include acoustic phonetic information. Viewed in this way, it is not only singers who construct different styles in singing and speech, but also their listeners. The aim of this experiment is to illustrate this case of perceptual style shift.

---

[4] As discussed earlier, singing and speech should not be treated as categorically distinct, but rather as two points on a continuum. While I will refer to 'sung exemplars' and 'spoken exemplars'; this should not be taken to imply a separation between two distinct systems. It is the degree to which given instances of singing and speech are different to one another that determines the degree to which the related exemplar clouds will also be distinct.

6.2.1. Hypothesis

This study focuses on the NZE DRESS and TRAP vowels, which are raised and fronted compared to other dialects of English. As shown in the study of vowel production, DRESS and TRAP are also more open in popular singing than in speech for at least the singers analysed in this thesis. The experiment presented here tests whether listeners perceive the boundary between DRESS and TRAP as being at an opener position (at a higher F1) when they hear a word in the context of music. The starting assumption of the study is that in the lexicon of a New Zealander, stored exemplars of these vowels have higher F1 in popular singing than in speech. This means that the distribution for sung exemplars of DRESS and TRAP is, in general, opener in auditory-acoustic space than the centre of the DRESS and TRAP exemplar clouds overall. It is likely that there is considerable overlap between spoken exemplars of NZE TRAP (which is closer in NZE) and sung exemplars of DRESS (which is open in singing). This situation could result in listeners categorising a word with the same formant values differently depending on whether they are listening to spoken or sung language.

It is hypothesised that in the music condition a higher percentage of the stimuli will be categorised as *bed* (rather than *bad*) than in the no-music condition. In the same way that for Ladefoged and Broadbent (1957), listeners categorised a stimulus according to its relation to the carrier phrase, it is hypothesised that respondents will categorise the stimuli according to the relevant frame of reference, be it spoken NZE or sung pop music. It is hypothesised that participants will expect to hear lowered DRESS and TRAP in song and raised DRESS and TRAP in speech and that these expectations will cause them to perceive ambiguous stimuli differently across conditions. If this happens, it will demonstrate that for these vowels the non-NZE variants are the default in the context of pop music. This, in turn, suggests that the differences between singing and speaking found in the vowel production study should not be described as an initiative act of identity, but rather as a response to established norms – the result of habit not effort.

6.2.2. Methodology

Experiment design and procedure

The experiment uses a binary forced-choice word-identification methodology to establish the perceived boundary between two phonemes. Each participant was presented with two conditions: music and no-music. Each condition contained 20 tokens played in a randomised order. The presentation order was always the same in the music

and no-music conditions for a given participant. In the music condition, the stimuli occurred over an instrumental music background, which will be described below.

The experiment was conducted in a quiet university office with stimuli presented on a laptop through headphones. Participants were told they would hear a New Zealander singing (in the music condition) or saying (in the no-music condition) the words *bed* or *bad* and that they were to decide which word they had heard and circle it on a response sheet, an example of which is shown in Appendix F. The instructions, which were read out to each participant, can be seen on this sheet. Additionally, either the word *singing* or *speaking* was included at the top of each response sheet. After the first condition had been run, the participants were handed the second response sheet and they were told, "and now, you will hear a New Zealander *singing/saying* either 'bed' or 'bad'", with emphasis on the word singing or saying which distinguished the two conditions.

In the music condition, the stimuli were played on the first beat of every second bar (that is, in a prominent position on every eighth musical beat), with the participant responding to the task in the gap between each word. In the no-music condition, the stimuli were presented at the same rate but without any background music. In the music condition (but not the no-music condition), every second token was pitch-shifted up one semitone to fit the chord progression and create a sense of melody. The effects of this decision to pitch-shift some of the stimuli will be discussed below. Half the participants heard the music condition first, and half heard no-music first. For half of the participants, the word *bed* was on the left of the sheet and for half it was on the right; the order of the words *bed* and *bad* in the instructions matched their position on the response sheet.

Creation of materials: vowel continuum and musical accompaniment

A vowel continuum from NZE *bed* to AmE *bad* was created by resynthesising the first two vowel formants of a word recorded by myself. In this recording, I aimed for a realisation roughly in the middle of the intended continuum, though this was not absolutely essential since this token was not used in the experiment in its original form. The recording was made through an AKG C4000B condenser microphone and recorded into a Mac Powerbook G4 using Logic Pro 7 recording software and an M-Audio 410 audio interface. The .wav file was recorded at 24bits with a sample rate of 44.1kHz. The aim was to have a vowel similar to a spoken NZE *bed* (maximally close) at one end of the continuum and a vowel in the vicinity of a sung, AmE style *bad* (maximally open) at

the other, with small enough differences between steps for the middle steps to be ambiguous.

Vowel synthesis was done in Praat, using a script created by Paul Warren and modified by Jen Hay. The length of all stimuli is the same (726ms for the whole word, 282ms for the vowel segment), and F3 is held constant across all tokens.[5] The continuum initially contained twelve steps, with the formant values being chosen based on a comparison of the values used for the continua in Hay et al. (2006) and Drager (2005), as well as referring to the values for DRESS and TRAP found in Labov, Ash, & Boberg (2006) for different dialects of AmE. The values given in these sources were used as a guide but could not speak directly to the purposes of the continuum in this experiment, since there was no relevant information about formant values in popular singing.[6] In the end, then, the continuum was based partly on the sources and partly on an impressionistic assessment of the vowel qualities.

An instrumental musical accompaniment was recorded in Logic Pro 7 on a Mac Powerbook G4, using sampled drums from Apple Loops, home recordings of guitar, and recordings of MIDI keyboard and bass instruments. The recording was 2'22" long, in 4/4 time at a tempo of 80 beats per minute. It cycled continuously around a common pop music chord progression (Ab Fm Db Eb[7]). In terms of music genre associations, the backing could be described as being of a generic easy-listening style.

The non-pitch-shifted stimuli were at a pitch of 127Hz which corresponds to the note C3 (one octave below middle C on a piano). This constitutes the major 3rd interval of the Ab major chord. Every second stimulus (the pitch-shifted tokens) fell on the Db major chord, and the pitch-shifted notes were the root note (Db) of that chord. This Db is also the suspended 4th of the Ab major chord, which creates a point of tension which then resolves when the chord progression begins again. This gives the recording a sense of melody despite the use of only two notes, one semitone apart. Examples of the full

[5] The script uses Linear Predictive Coding to manipulate F1 and F2 so that there is a steady state in the middle portion of the vowel which conforms to the formant values specified by the user. Start and end points for the synthesis are also specified, so that a logistic function can be used to ensure a smooth transition into and out of the steady state portion of the vowel. An initial trial of the continuum allowed F3 to move but it was found that having a stable F3 made the continuum more natural sounding.

[6] It should be noted that the perception experiment was carried out before the data for the vowel production study had been collected, so the results of that study could not be used to decide the formant values for the continuum.

music and no-music experiment materials, as well as both the original and the pitch-shifted continuum from *bed* to *bad* are included on the accompanying CD (see Appendix E for track numbers).

Pilot study

The 12-step continuum was placed into the experimental design and piloted on 10 participants. Based on this pilot, two issues with the continuum were noticed. Firstly, there was almost no variation in the responses for the four outermost tokens at each end of the continuum. Ambiguity between *bed* and *bad* only seemed to occur for the four stimuli in the middle. Since the purpose of the experiment was to determine how people would respond to ambiguous stimuli, this 12-step continuum was spread more widely than necessary. Secondly, it was found that the two stimuli at each end of the continuum had a large effect on the participants' response to the next stimulus. An extreme *bed* stimulus, for example, was disproportionately likely to be followed by a *bad* response in the next trial. Based on this pilot, it was decided that the continuum would be restricted to the middle eight stimuli, removing the two most *bed*-like, and the two most *bad*-like tokens from the continuum used for the main experiment. This allowed more trials with ambiguous tokens to be presented.

The other outcome of the pilot study, based on the observation of strong effects of preceding stimulus, was a decision to re-randomise the presentation order of the stimuli after every fifth participant. The presentation order was randomised with the exception of the first trial, which was always one of continuum steps 3, 4 or 5.[7] This was done so that the first stimulus presented to every participant would be in an ambiguous part of the continuum. These first trials would therefore provide a subset of results not affected by preceding stimuli. In all, eight different presentation orders were used. Within each block, 20 stimuli were presented; these were made up of two instances for each of continuum steps 1, 2, 7, and 8, and three instances each for steps 3, 4, 5, and 6.

Main study

The continuum used in the main study included 8 stimuli, after having excluded the outermost steps of the continuum used in the pilot study. The spectrogram of the continuum is shown in Figure 6.1 and the formant values of the stimuli are shown in

[7] Only eight participants had a version of the experiment which began with continuum step 3, which was an oversight since steps 4 and 5 represent the middle of the continuum.

Table 6.1, with the target values that were entered into the Praat script as well as measurements from the resulting soundfiles of synthesised vowels. The outcome of the synthesis process was quite different from the target for some tokens. The resynthesis depends on the availability of appropriately placed harmonics, and this may have been the cause of the difference between the targeted and actual formant values. Unfortunately, I did not think to take independent measurements of the stimuli after resynthesis, so by the time I discovered that the steps were of varying sizes it was too late to correct the issue. However, since the anomalies in the continuum were presented equally in both of the test conditions, this issue does not affect the validity of the results. The range for F1 spans from 450-660Hz, while F2 spans from 2110-1800Hz.

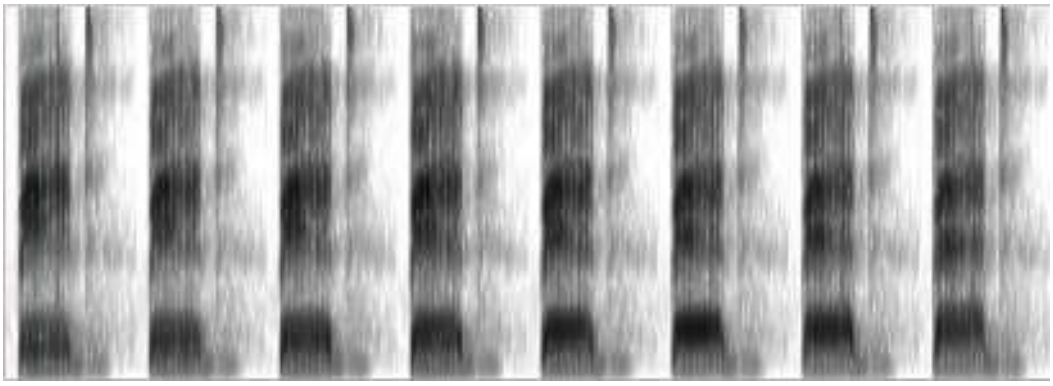

Figure 6.1: Spectrogram of the 8-step continuum from *bed* to *bad* used in the perception experiment

Table 6.1: *Formant Values used for the Resynthesis of Vowels to Create a Continuum from* bed *to* bad*, and Measurements from the Resulting Soundfiles of Resynthesised Vowels*

	Target		Actual	
Continuum Step	F1	F2	F1	F2
1	465	2090	450	2110
2	500	2060	500	2070
3	535	2030	530	2030
4	565	2000	580	2000
5	600	1970	620	1950
6	635	1940	630	1920
7	665	1910	630	1890
8	700	1880	660	1800

Figure 6.2 shows the F1/F2 values for each of the continuum steps in comparison with the averaged DRESS and TRAP vowels for singing and speech recorded in the study of vowel production. As mentioned above, this data had not yet been collected when the

perception experiment was carried out. It can be seen that even though the target values placed into the resynthesis algorithm were at even intervals, the tokens as measured from the spectrogram of the synthesised vowels show some anomalies, particularly in the clustering together of steps 5,6 and 7. When compared to the data for John, the continuum meets the goal of spanning the range from a spoken DRESS to a sung TRAP. It is rather out of place when compared to the values for Andrew and Dylan however, with the *bed* end of the continuum being too fronted, and the *bad* end of the continuum being less open than not only sung TRAP but also sung DRESS. Importantly, however, the *bad* end of the continuum is much more open than the spoken values for TRAP.

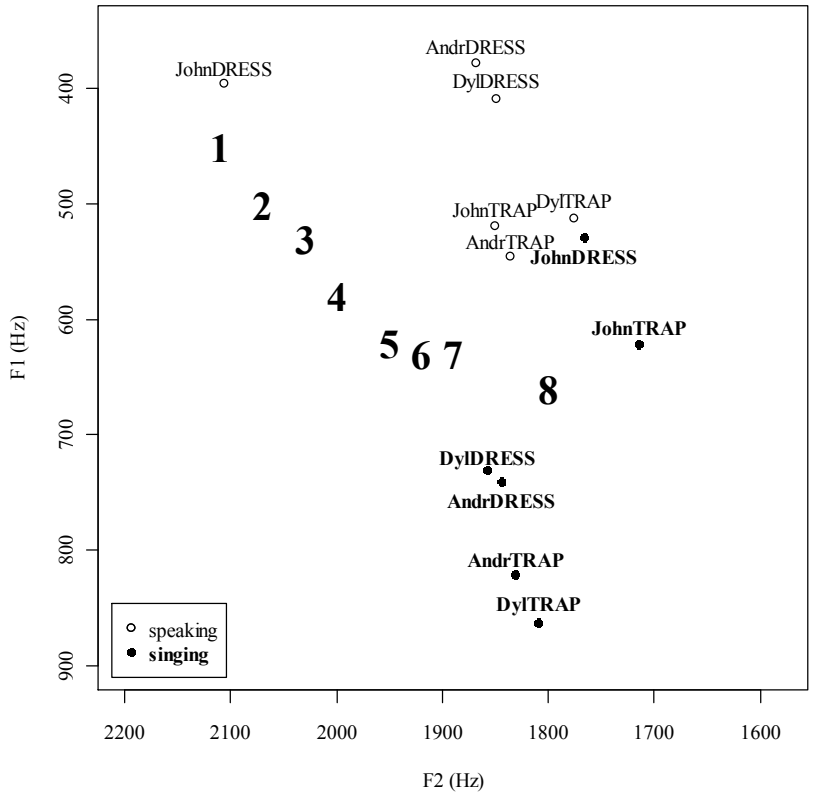

Figure 6.2: Continuum-step numbers, as measured after resynthesis, plotted with average sung and spoken formant values for DRESS and TRAP found in the vowel production study

The pitch-shifting of tokens was done after the vowel resynthesis, leading not only to a higher pitch but also to higher formant values. When designing the experiment, I naively thought that this shift would be so subtle that it would not affect the perception of the vowel. Unfortunately, this was not the case, as will be discussed in the next section. The pitch shifted words, which formed half of the stimuli used in the music condition, were on average 30Hz higher for F1 and 120Hz higher for F2, which should cause the vowels to sound more open and more fronted.

Participants

Posters were placed around the Auckland University of Technology and University of Auckland campuses, and also on the AUT online 'staff noticeboard'. The participants were mainly either university staff or students. There were 37 participants, 24 were female and 13 were male. There was a wide age range, with nine participants each in their 20s and 30s, eight each in their 40s and 50s, and three in their 60s. One of the selection criteria was that participants had lived in New Zealand most of their lives, having spent less than ten years overseas. All but three of the participants had spent less than five years outside of NZ. Information on the participants' occupation, ethnicity, "favourite kinds of music", and fluent languages spoken was also collected, though these aspects of the data have not been analysed since the results do not consider inter-personal differences, as will be described below.

Ethical considerations

Ethical approval was obtained before approaching the participants, and the letter of approval is included in Appendix G. The appendix also includes the information sheet which each participant read before beginning the experiment, and the consent form which was signed. Participants also provided answers to the demographic questions on this form. The recruitment advertisement and the information sheet both kept information about the experiment to a minimum, explaining it in very vague terms. The advertisement described the project as "a linguistics experiment which tests listeners' perception of words", while the information sheet explained to participants: "you will be played a series of words through headphones and circle on a piece of paper which word you heard". As can be seen from the information sheet in Appendix G, the participants did not know that the experiment would involve music. For those who had the no-music condition first, they knew nothing of the musical aspect of the research until they were presented with the instructions for the second condition. After completing the experiment, participants were offered a chocolate bar to thank them for their involvement.

6.2.3. Results

There were a total of 1480 trials, 20 trials in each of 2 conditions, for 37 participants. The results found that the two ends of the continuum unambiguously represented *bed* and *bad*, irrespective of condition. Continuum step 1 was perceived as *bed* in 99.3% of trials, and continuum step 8 was categorically perceived as *bad*. This shows that the

vowel resynthesis created the desired outcome of spanning both vowels in both conditions. Overall, 49.5 percent of stimuli were perceived as *bed*, and 50.5 percent as *bad*. Figure 6.3 shows the percentage of trials categorised as *bed* for each of the eight steps of the continuum, for the music (solid line) and no-music (dashed line) conditions.

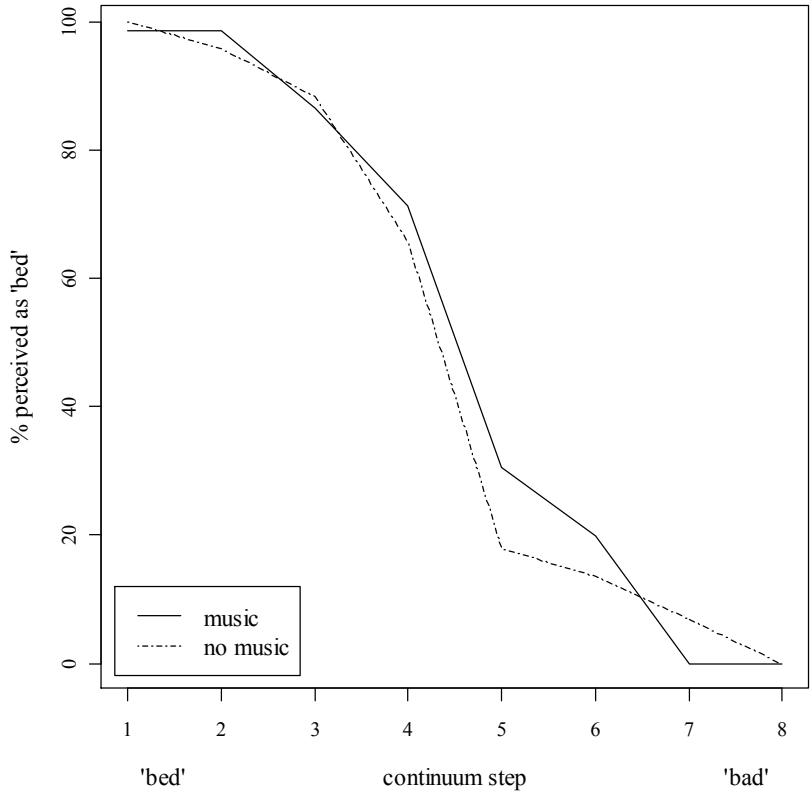

Figure 6.3: Percentage of trials perceived as *bed* for the eight steps of the continuum from *bed*-*bad*, (most *bed*-like=1, most *bad*-like=8), for the conditions with and without music.

It can be seen that the percentage of trials categorised as *bed* was marginally higher in the music condition than in the no-music condition. Overall, 50.9 percent of trials were categorised as *bed* when background music was playing, compared to 48.1 percent when the words occurred with no music. If we consider the most ambiguous tokens in the middle of the continuum, steps 4 and 5, the result is more apparent with categorisation as *bed* being 51.9 percent in the music condition and 44.3 percent in the no-music condition. However, the raw figures are somewhat misleading due to the effect that the pitch-shifted tokens had on responses.

In order to analyse the results statistically, the data were entered into the statistical program "R" and a logistic regression model was built to assess the likelihood of a participant responding *bed* when taking into account all of the variables measured. The model was designed to hold constant any effects of individual variation, and also to control for any variation caused by the altering of condition order (whether the music

condition was presented first or second) and layout of response sheet (whether *bed* was on the left or right column of the response sheet).[8]

The full model was fit with the following independent variables:

- condition (a factor with two values: music and no-music);
- whether or not the stimulus was pitch-shifted (no or yes);
- continuum step (from 1 to 8);
- order (serial position of the trial within the block of 20 stimuli);
- preceding stimulus (the continuum step of the stimulus from the preceding trial, from 1 to 8);
- and participant number.

Table 6.2 shows the output of the final model showing that all of the variables included in the model reached significance.[9] The coefficient column is most important for interpreting the results, and will be discussed for each variable in turn. It should be emphasised that all of these results hold constant the estimated variability in all other variables.

- Condition – the log-odds[10] of responding *bed* are significantly lower in the no-music condition. That is, the hypothesis was supported, with participants significantly more likely to perceive a word as *bed* when it was played in a musical context and when they had been told they would hear singing.

---

[8] In this type of model, we cannot analyse for speaker specific variables such as age and sex, due to the fact that there are 40 responses per individual. This violates the assumption that all observations are independent of one another. In order to overcome this issue and pool together all the data from different participants, the participant number was added as a factor in the model. This means that individual variation is offset by a specific coefficient for each participant.

[9] The variance inflation factor was checked to ensure that there were no collinearities in the predictors. Also, all plausible interactions were tested for and none were found to be significant. In the table, the coefficients which compare each of the participants to participant 1 are not shown in full. This factor is only included to hold constant any effects of individual variation, as described in the previous footnote.

[10] Odds are the probability of an event occurring over the probability of that event not occurring. The coefficients represent odds, but on a logarithmic scale which allows for linear modelling.

- Pitch-shifting –there was a significant effect of whether or not the target was one of those which had been pitch-shifted to create the sense of melody. If the token had been pitch-shifted, it was less likely to be categorised as *bed*. As mentioned earlier, the pitch-shifting led to higher formant values. Since an increase in F1 is associated with vowel opening, it is not surprising that the pitch-shifted vowels sound inherently more *bad*-like. This result also explains why the raw figures shown at the beginning of this section do not strongly reflect the effect of condition. There was an inherent bias against the hypothesis that participants would hear more *bed* in the music condition, because this condition contained all of the pitch-shifted stimuli.
- Token – predictably, more *bed*-like tokens (lower continuum numbers) are more likely to be perceived as *bed*.
- Order – the later on in the experimental condition, the more likely the participants are to respond *bed*. This could have occurred if participants had a tendency to try and give an even number of responses for each of the words, and if the continuum was slightly biased towards *bad* overall, leading them to respond *bed* more in later trials. This effect, as with all others, is held constant in the model and so does not affect the result for condition.
- Preceding – as expected on the basis of the pilot study, there is a contrast effect such that the more *bad*-like the preceding token, the higher the likelihood of responding *bed*. That is, people had a tendency to compare the target with its predecessor and respond according to the relative difference. The magnitude of this effect is very small compared to the magnitude of the effect of the actual token number, showing that people were mainly responding according to the sound of the target vowel.

Table 6.2: *Output of the Logistic Regression Model Analysing the Log Odds of Categorising a Stimulus as* bed*, as a Function of the Independent Variables of the Perception Experiment*

Variable	Coefficient	Standard Error	Wald *z*-statistic	*p*-value
Intercept	3.85051	0.65859	5.85	0.0000
Condition=no-music	-0.80497	0.22288	-3.61	0.0003
Pitch-shifted=yes	-0.94275	0.26211	-3.60	0.0003
Token	-1.65673	0.08902	-18.61	0.0000
Order	0.05975	0.01580	3.78	0.0002
Preceding	0.19781	0.04123	4.80	0.0000
Participant=2	3.03843	0.78401	3.88	0.0001
...				
Participant=37	1.93041	0.81992	2.35	0.0186

The pitch-shifting of stimuli amounts to quite an important flaw in my design of this experiment, and though the effect of condition is clear in the logistic regression, the result does not come across as simply as it could have. Fortunately, there is a portion of the results which is totally unaffected by these complications. Recall that all participants were played an ambiguous stimulus (continuum steps 3-5) at the very beginning of the experiment. Half of the participants began with the music condition, and half with no-music, creating a balanced between-participants design study for this first trial. In the music condition the first stimulus played is a non-pitch-shifted token, so the stimuli are acoustically identical across conditions. Essentially, by assessing participants' perception of the very first word heard, we cut out effects of pitch-shifting, and also any 'tuning in' effects that might be caused by hearing various versions of *bed* and *bad* 40 times in a row.

Of the 37 participants, only eight were presented first with continuum step 3 and six of these eight were in the no-music condition. The other 29 participants, however, are well balanced across token number and condition. For continuum step 4, nine occurrences were with music, and ten without. For continuum step 5, the ten occurrences were equally balanced across the music and no-music conditions. Since the participants who first heard the continuum step 4 or 5 are balanced for condition, their results can be grouped together to test for any difference between conditions, and to see whether or not people were more likely to categorise the stimulus as *bed* when they had just been exposed to a few seconds of instrumental pop music. Figure 6.4 shows the responses for

these first trials. In the music condition, 11 out of the 14 participants responded *bed*. In the no-music condition, only 2 out of 15 respondents perceived the word as *bed*. A Fishers Exact test found this difference between conditions to be highly significant (p=.0007).[11]

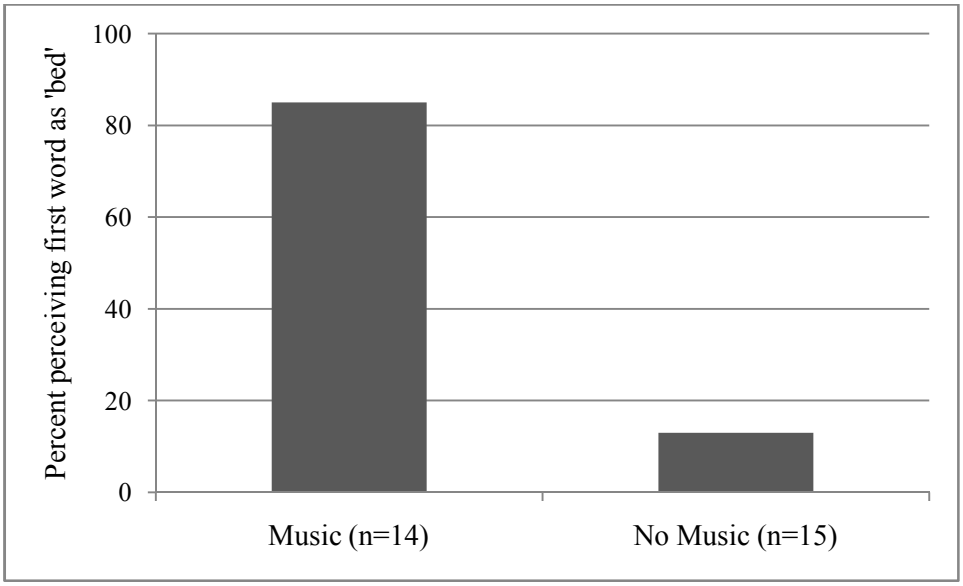

Figure 6.4: Percent of "bed" responses to first stimulus of the experiment, according to condition, for stimuli in the middle of the *bed-bad* continuum (steps 4 and 5).

This result is quite striking, showing that across participants, the boundary between DRESS and TRAP shifted to a more open position when the context for perception was musical. While this result is based on a small subset of the data collected, the finding is transparent and robust. It is unaffected by the issues caused by the pitch-shifting of stimuli in the rest of the experimental trials, and it is also unaffected by the contrast effect which occurs after having been exposed to stimuli from either end of the continuum.

6.2.4. Discussion

The hypothesis was supported. *Bed* was perceived more often when people were told they would hear singing, and there was a musical accompaniment, than when they were told they would hear speech, and there was no music. This means that while participants expected to hear a typical NZ-like raised TRAP in *bad* in the no-music context, they did

[11] This pattern seemed to be upheld by the eight people that had continuum step 3 in their first trial. One of the two people in the music condition categorised the stimulus as *bed*, compared to one out of six when there was no music.

not expect this realisation for *bad* in singing, where *bed* was a more context-appropriate interpretation.

These results can be interpreted in two ways. They are of great interest even without considering how they relate to exemplar theories of speech perception. Most sociolinguists would agree that a listener must adapt the way they perceive in order to understand different dialects of their language. If a New Zealander is talking to a Scot, for example, they will perceive the speech of their interlocutor with some reference to whatever exposure they have had to Scottish English. In the experiment, the listener perceives the ambiguous stimuli with reference to what they know about vowel sounds in singing as opposed to speech. Not only do singers use a more open DRESS and TRAP in singing than speech, but also, their listenership knows that this is the case. They use that knowledge to understand lyrics in pop music, in the same way that they use their knowledge of Scottish English to understand the speech of a Scot. Listeners shift their perceptual frame according to the context.

Exemplar theories can model this intuitive concept of a perceptual frame in a more detailed way. The person in the example above might label their interlocutor as 'a Scottish person', causing memories of that speaker's words, along with their phonetic detail, to be indexed as Scottish. In the same way, we can also assume that when people listen to music, they label the person singing as 'a singer' or perhaps a 'pop singer', a 'punk singer' or an 'indie singer'. Whatever the nuances of the labelling, words heard in the context of music are all likely to be labelled as 'sung'. When listening to music, exemplars labelled as 'sung' become active. Based on this idea, the results of the perception experiment can be described in detailed terms, describing not only the shifting of a perceptual frame, but *how* that might occur.

As the introductory music played in the first seconds of the music condition, memories of previous experience with music were activated, more or less strongly according to the similarity of the remembered and currently on-going music styles. These memories of music in turn activated associated memories of the singing, and the phonetics thereof, to which they were linked. Participants also knew they were going to hear either *bed* or *bad*, so exemplars of these words and of DRESS and TRAP vowels more generally, would also have been active. The phonetic form of the activated exemplars in the music condition would have had higher F1s than those activated for participants in the no-music condition, who were expecting to hear a New Zealander speaking. The incoming vowel sounds were then perceived in a way which was appropriate to the participants'

expectations in each case. Thus, listeners shift their phonetic space according to context just as speakers do. The implications of this perception experiment in relation to the vowel production study will be discussed below. However, there are several limitations of this experiment which should be noted first.

For a start, it is an entirely artificial scenario. Being told that you are going to hear one of two words, in isolation, and that you are to choose which one you hear, is not an ordinary language activity. All kinds of non-linguistic and meta-linguistic strategies might be used. However, this issue is not specific to this study, but is rather a problem for a range of methodologies for exploring perception. Another limitation is the lack of investigation into differences between participants according to sex, age, and musical taste. This was considered to be out of the scope of the thesis, and beyond the purpose of this experiment, which was to determine whether the difference between a musical and non-musical context could affect vowel perception. Another limitation is the fact that the differences between singing and speech for DRESS and TRAP are on the F1 dimension, which is affected at least in part by a preference for sonority. It would be particularly convincing if this experiment could be replicated using an ambiguity which relied on an F2 difference. However, that does not affect the validity of the present finding, since it is more to do with the reasons for the differences between singing and speaking than to do with their normativity.

This study could be repeated and adapted in several ways. The most obvious extension is to use different vowels. It would be possible, for example, to examine the perception of NEAR/SQUARE words. Listeners might be more likely to perceive [iə] as *hair* in non-musical contexts, since the pronunciation of SQUARE words with [iə] is very rare in singing. Another adaptation would be to use different styles of background music. For instance, a punk music background versus a pop music background could influence the perception of a vowel towards an urban London English interpretation in the former, and an AmE interpretation in the latter.

In relation to the theories of speech perception presented in the first part of this chapter, the results are difficult to account for in models which posit speaker normalisation to deal with variability. The results suggest that variation in the speech signal is used by listeners, not filtered out. While I have chosen to discuss the results in terms of exemplar theory, I cannot claim to have discounted the possibility of other interpretations. A model which posits abstract prototypes with some kind of interface with contextual knowledge could potentially be applied. Some mixed theories, for

example, allow multiple prototypes, without the stipulation that individual episodic memories are referred to in speech perception. Such a theory could model abstract representations of a prototypical spoken DRESS and a prototypical sung DRESS. While an exemplar framework seems to account for this result well, I am also open to other potential interpretations.

An exemplar perspective on speech (and singing) production

In this chapter, we have seen how listeners use their experience with variation to resolve ambiguities in the speech signal, drawing on memories of previously heard and understood speech from similar speakers engaged in similar types of verbal activity. The processes which allow listeners to understand speech despite its variability may be quite similar to those used by speakers in their construction of style. At the beginning of this chapter, I highlighted the fact that the singers analysed in the vowel production study are also listeners. Thus, the results of this perception experiment can shed light on the question of why singers modify their pronunciation when they sing.

Pierrehumbert (2001) provides an extension of the exemplar approach to discuss speech production, in which similar mechanisms as those used in speech perception apply. In the uttering of a word, the phonetic target is based on stored exemplars, and it is influenced by context. Pierrehumbert (2001, p. 7) states that "social and stylistic factors may select for different parts of the exemplar cloud in different situations". In singing production, exemplars indexed as 'sung' would be highly activated, and the phonetic target of vowel production would be influenced by this subset of exemplars. The strong association between exemplars of 'pop singing' and AmE-influenced phonetic forms would cause singers to reproduce those forms when singing pop music.

The exemplar approach to speech production may also provide a new way to conceptualise how the responsive and initiative dimensions of style (Bell, 2001) might operate for a speaker. While speaking, exemplars indexed to people similar to oneself or one's interlocutor are likely to be most strongly activated – exemplars indexed to the label 'New Zealander', for example. The resulting pattern of activation causes NZE phonetic forms to be the most automatic (responsive) target for speech production in that context. To use an AmE alternative would constitute an initiative act – a stylisation; it would require the speaker to turn their attention to 'Americans', thus *initiating* activation of the relevant exemplars rather than *responding* to the situationally determined activation pattern.

In the pop singing contextual frame, the relevant group is not so much 'New Zealanders' as 'pop singers', or maybe 'New Zealand pop singers' (these labels should not be taken as fixed; labels relating to specific artists, to music genres, and various stances would all be simultaneously activated). In this situation, a very different set of exemplars would be activated. It is these 'sung' exemplars which influence the phonetics of the singing production most strongly, and which the singer *responds* to most automatically. To sing in a NZE accent (or a Sheffield accent for Arctic Monkeys), would require the singer to turn their attention to 'New Zealanders talking' (again, a variety of labels would be activated), which would *initiate* activation of NZE variants, allowing the singer to stylise NZE, even though it is not the norm in the context of singing.

Chapter 7 Conclusions

This chapter draws together the results of the two empirical studies described above and integrates them into an approach to popular singing that incorporates both social constructionist and exemplar theory perspectives. The results of the vowel production study show how singing and speaking are phonetically different for three NZ popular singers. These differences occur across the full range of vowels studied. The singers who held positive views about NZ-accented singing did not show many signs of this opinion in their singing styles, except through avoidance of features that are stereotyped as Americanisms in the context of pop singing. Also, there was some evidence that singers are able to use NZE features when they make a conscious effort to do so, but this requires a more initiative construction of style than does the use of American-influenced forms.

The perception study provided complementary evidence for the claim that American-influenced phonetic forms are normative, by showing that such forms are expected by listeners in the context of pop music. Different listeners presented with the same stimulus word perceived it differently according to whether or not it was expected to be sung. Thus, the normative singing style is not just something singers perform; it is also what listeners expect.

7.1. A Memory Perspective on the Sociolinguistics of Style

Having reviewed both sociolinguistic theories of style and exemplar approaches to language perception, it is now possible to consider how compatible they are with one another. There is a growing body of literature which sees the amalgamation of the two perspectives as providing a promising avenue for the conceptualisation of social and stylistic variation (Drager, 2009; Foulkes & Docherty, 2006; Hay & Drager, 2007). The main commonality between the two approaches has come about through the adoption in exemplar theory of the concept of indexicality as it is used in sociolinguistics. In very general terms, both social constructionist and exemplar approaches have in common a focus on the multi-dimensional nature of language variation, which can be represented by multi-dimensional constructs such as indexical fields and exemplar clouds.

In social constructionist terms, a raised TRAP vowel, for example, may be described as *indexing* New Zealand-ness. Exemplar theorists would say that the phonetic memories of raised TRAP are *indexed* to 'New Zealander' and become activated when the label

'New Zealander' itself is activated. The difference is that exemplar theory attempts to address how indexicality might operate in the mind of the language user, as the overlapping of neural structures that are activated by both sides of a given sign-referent pairing. When [bɛd] is heard for *bad*, memories (patterns of neural activation) of the word activate other memories associated with that phonetic form. This web of associations between memories of phonetic form and socio-contextual memories constitutes the indexical field. The indexical field shifts according to the degree of activation of these associations, so that the social meaning of a given phonetic string can change depending on any aspect of context that the listener is paying attention to at a given moment. Linguistic style in speech and singing is very different because singing and speech differ so dramatically in terms of such a wide range of contextual elements. More subtle stylistic variation may occur when the context shifts on only a few dimensions.

7.2. Identity in Singing

Turning to the identity processes involved in singing accents, the title of Trudgill's seminal article is still highly relevant; singing in popular music involves *acts of conflicting identity*. The sheer contextual specificity of language use in popular music means that the dimensions of identity which attract the most attention are those to do with the music itself. The other dimensions of self, including that of national identity for example, are less salient in this context. In an exemplar account, this is because exemplars of spoken language are less activated than exemplars of singing, which do not reliably index nationality.

When those selves associated with the singer's normal speaking persona stay in the background, there is no conflict – the singer simply sings how popular singers sing. But sometimes, these other selves attract the singer's attention in the context of singing, especially when a salient variable has quite different pronunciations in singing and speech. For example, sometimes a variant in the normative singing accent evokes negative social meanings for the singer, as in the example about Apache Indian country singers recounted in 2.3.1. For those singers, the association of vowel diphthongisation with 'redneck' had received enough metapragmatic attention that it remained relevant even in the context of singing, where other variables would take on different social meanings associated more with musical style.

In this situation, a conflict of identity may arise, and the singer is faced with a decision. They may evaluate the indexical meaning of the 'offending' variant in one of two ways. On the one hand, they can say 'this is just how singing is done, so it is not a betrayal of my identity'. Alternatively, they may think (taking the Apache example) 'this sounds redneck, I don't want to sound redneck, I cannot pronounce the word in that way'. If the latter line of reasoning occurs, then the singer may engage in an ingroup referee design, where the referee is their own spoken voice.

In New Zealand popular music, there have been relatively few models of NZE singing accents available for emulation, but this scenario appears to be changing as more singers adopt local variants. The use of such NZE forms may make the dimension of identity more salient in the pop singing contextual frame for these singers, by attracting attention. The more variation there is between the pronunciation of different NZ musicians, the less new singers will be able to ignore their place identity – they must make decisions between competing variants. This is perhaps what makes hip-hop such an interesting area for the study of stylistic processes, both linguistic and non-linguistic. There is a lot of variability in pronunciation forms, which is either the result of, or the cause of (or both) an application of the value assigned to 'keeping it real'. Hip-hop artists have to negotiate between their identity as a member of the global hip-hop community, and their 'ontologically authentic' identity, which includes their localised place identity, as Australian, for example. Bands like Arctic Monkeys may be part of a movement which will have the same result for anti-mainstream pop music forms over the coming years. However, there is little evidence of this in the singing of the singers analysed in the vowel production study.

Both Dylan and John are good candidates for styling a non-normative style. Dylan even stated that because his music is not mainstream pop, he could get away with singing in a NZ accent. But in the end he chose not to. In the NZ scene at the moment, it seems that the normative force of the AmE-influenced model is still strong. But if a range of high profile artists started using NZE in their singing, this situation could change quickly.

New pronunciations lay down new memories. The way these new memories interact with musical genre is very important. If NZ accents are only associated with alt-indie music, for example, in the same way that working-class London accents have been associated with punk, then it may become especially context *in*appropriate for a mainstream pop singer to use NZE. Rather than aligning them solely with their NZ identity, it would align them with a specific genre. Thus, it is important to understand

the way genre-specific identities are constructed, not just in terms of language, but in terms of an artist's image and in the style of the music itself. The notion of conflicting identities is once again apparent. A pop singer may have to decide whether it is more important to assert their New Zealand identity, or to assert their identity as a commercial pop artist.

If successfully performed, 'own accent singing' may index sincerity and authenticity, but if the accent is perceived as contrived it will attract a negative response. Singers must therefore negotiate a path between imitation and innovation in order to be both accessible and distinctive. They do this through their music, their image, the way they move, and the pronunciation they use when they sing.

7.3. Future Directions: The Rise of Sung NZE

In chapter 6, I introduced the idea that people base their speech on socio-contextually appropriate examples of remembered speech. From this base, they have the ability to break with norms in order to do identity work. Trudgill's (1983) application of the acts of identity framework is relevant to an analysis of popular singing, but it is the American-influenced singing style which should be viewed as normative and automatic. Acts of identity occur when singers break away from this American norm.

In the two years since beginning this research, I have discovered a number of new artists who use NZE in their singing. I have heard several artists using a fronting GOAT vowel in a consistent fashion (though perhaps this is only because that variable is especially salient to me now), and stretches of singing that use NZE realisations of less salient variables. A high profile example of this is in a recent song by indie-electropop artist Ladyhawke. Her pronunciation of the line "bang bang bang on the wall, from dusk till dawn" is delivered in a strongly NZE style, containing raised TRAP vowels, and rounded, raised and retracted LOT and THOUGHT. While this song does not represent Ladyhawke's overall pronunciation style, this particular line has now been heard by tens of thousands of young New Zealanders, and it may have made an impression.

It takes a certain degree of effort and practice to translate a NZE accent into the sung context, along with a willingness to embrace the fact that it will sound unusual, and that some people might *cringe*. When singers such as Ladyhawke do break the established norms in this way the indexical field will shift slightly, with associations forming between NZE phonetic forms and popular singing contexts. These new indexical associations could become more stable at the community level if a number of NZE-

accented singers achieve prominence. Some young singers-to-be may grow up singing along to this handful of NZE-accented pop songs. Their stored exemplars will be quite different, the habit to AmE pronunciation in singing will be diluted and new indexical ties will be formed. Sung NZE may develop into an established sociolinguistic resource, just like any other. It will certainly index localness, and it may also index non-commercialism and be associated with certain musical styles.

Once there is a contextually well-defined precedent for NZ accented singing it will be more appropriate and conventional to sing with NZE vowels in that context. However, singing inherent effects such as a preference for sonorous vowels will remain, and in some cases conflict with the acts of NZ identity making. The evolution of this process is in its infancy, but may progress rapidly. The developments which occur over the coming years are thus an ideal site to study both the linguistic construction of identity, and processes of language perception and representation.

Because study of popular singing deals with an already recorded form of language, there is also significant scope for establishing how patterns of pronunciation have changed over time, across regions and between musical genres. For example, the current state of Australian hip-hop as being pronounced largely in an AusE style may represent the end of a process of localisation. It would be very interesting to analyse the 15-20 years of hip-hop which preceded this state. In the same way, the acoustic analyses presented in this thesis may provide a baseline for the changes in NZ singing accents. It may be that in ten years' time there will be far greater variability, with NZE being sung in a wholesale way for some genres, and being mixed into the pronunciation of others.

I hope that in this thesis I have convincingly put forward an argument for the study of singing accents as a way to examine how people construct different styles, adapting their phonetic space in quite dramatic ways according to contextual changes. The phonetics of singing provides fertile ground for the study of performativity, and for studying the tension between existing norms and the ongoing desire to do identity work by deploying existing semiotic resources in new ways. This thesis has also suggested, following recent work in sociophonetics, that the study of singing accents would benefit from an approach which unifies sociolinguistic theory and psycholinguistic theory to explain context-dependent style. A theory of social variation in language behaviour can benefit from the perspectives offered by both social constructionist theories of style and exemplar accounts of language representation.

References

Agha, A. (2003). The social life of cultural value. *Language and Communication, 23*, 231-273.

Alim, H. S. (2002). Street-conscious Copula Variation in the Hip Hop Nation. *American Speech, 77*(3), 288-304.

Austin, S. F. (2007). Jaw Opening in Novice and Experienced Classically Trained Singers. *Journal of Voice, 21*(1), 72-79.

Bakhtin, M. (1981 [1935]). Discourse in the novel. In C. Emerson & M. Holquist (Eds.), *The Dialogic Imagination* (pp. 259-422). Austin: University of Texas Press.

Bayard, D. (1995). *Kiwitalk: Sociolinguistics and New Zealand Society*. Palmerston North: Dunmore Press.

Beal, J. C. (2009). "You're Not from New York City, You're from Rotherham": Dialect and Identity in British Indie Music. *Journal of English Linguistics, 37*(3), 223-240.

Bell, A. (1984). Style as Audience Design. *Language in Society, 13*(2), 145-204.

Bell, A. (1999). Styling the other to define the self: A study in New Zealand identity making. *Journal of Sociolinguistics, 3*(4), 523-541.

Bell, A. (2001). Back in style: reworking audience design. In P. Eckert & J. R. Rickford (Eds.), *Style and Sociolinguistic Variation* (pp. 139-169). Cambridge: Cambridge University Press.

Bell, A., & Holmes, J. (1992). H-droppin': Two sociolinguistic variables in New Zealand English. *Australian Journal of Linguistics, 12*(2), 223-248.

Blom, J.-P., & Gumperz, J. (1972). Social meaning in linguistic structures: Code-switching in Norway. . In J. Gumperz & D. Hymes (Eds.), *Directions in Sociolinguistics: The Ethnography of Communication* (pp. 407-434). Oxford: Basil Blackwell.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1.04).

Borch, D. Z., & Sundberg, J. (1997). Spectral distribution of solo voice and accompaniment in pop music. *Logopedics, phoniatrics, vocology, 27*, 37-41.

Bourdieu, P. (1991). *Language and Symbolic Power*. Cambridge: Polity Press.

Burns, P. (1986). Acoustical analysis of the underlying voice differences between two groups of professional singers: opera and country and western. *Laryngoscope, 96*(5), 549-554.

Cleveland, T. F., Stone, R. E. E., Sundberg, J., & Iwarsson, J. (1997). Estimated subglottal pressure in six professional country singers. *Journal of Voice, 11*(4), 403-409.

Cleveland, T. F., Sundberg, J., & Stone, R. E. E. (2001). Long-term-average spectrum characteristics of country singers during speaking and singing. *Journal of Voice, 15*(1), 54-60.

Clopper, C. G., Pisoni, D. B., & de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America, 118*(3), 1661-1676.

Coddington, A. (2003). Stylistic Variation in Music and Speech: A Case Study. Unpublished honours dissertation. University of Auckland.

Coddington, A. (2004). *Singing as we Speak? An Exploratory Investigation of Singing Pronunciation in New Zealand Popular Music.* Unpublished master's thesis, University of Auckland.

Coleman, J. (1998). Cognitive reality and the phonological lexicon: A review. *Journal of Neurolinguistics, 11*(3), 295-320.
Coupland, N. (1980). Style-shifting in a Cardiff Work-setting. *Language in Society, 9*(1), 1-12.
Coupland, N. (1985). 'Hark, hark, the Lark': Social Motivations for Phonological Style-Shifting. *Language & Communication, 5*(3), 151-171.
Coupland, N. (2003). Sociolinguistic authenticities. *Journal of Sociolinguistics, 7*(3), 417-431.
Coupland, N. (2007). *Style: Language Variation and Identity*. Cambridge: Cambridge University Press.
Coupland, N. (2009). The mediated performance of vernaculars. *Journal of English Linguistics, 37*(3), 284-300.
Dominello, Z. (2008). "Keepin' it real, mate": A Study of Identity in Australian Hip Hop. *Griffith Working Papers in Pragmatics and Intercultural Communication, 1*(1), 40-47.
Drager, K. (2003). When I Sing I Like to Sound American. Unpublished honours dissertation. University of Canterbury.
Drager, K. (2005). *The Influence of Social Characteristics on Speech Perception.* Unpublished master's thesis, University of Canterbury, Christchurch.
Drager, K. (2009). *A Sociophonetic Ethnography of Selwyn Girls' High.* Unpublished doctoral dissertation, University of Canterbury, Christchurch.
Eckert, P. (2000). *Linguistic Variation as Social Practice*. Oxford: Blackwell.
Eckert, P. (2003). Elephants in the room. *Journal of Sociolinguistics, 7*(3), 392-397.
Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics, 12*(4), 453-476.
Einstein Music Journal. (n.d.). Retrieved November 15, 2009, from http://www.einsteinmusicjournal.co.nz/category/reviews/page/3/
Feld, S., Fox, A., Porcello, T., & Samuels, D. (2004). Vocal anthropology. In A. Duranti (Ed.), *A Companion to Linguistic Anthropology* (pp. 321-345). Oxford: Blackwell.
Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics, 34*, 409-438.
Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14*(1), 3-28.
Frith, S. (1988). *Music for Pleasure*. Cambridge: Polity Press.
Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The Identification and Discrimination of Synthetic Vowels. *Language and Speech, 5*, 171-189.
Gibson, A. (2005). Non-Prevocalic /r/ in New Zealand Hip-Hop. *New Zealand English Journal, 19*, 5-12.
Gibson, A., & Bell, A. (fc). Performing Pasifika English in New Zealand: The case of bro'Town. *English World-Wide, 31*.
Giddens, A. (1984). *The Constitution of Society: Outline of the Theory of Structuration*. Cambridge, UK: Polity Press.
Giddens, A. (1991). *Modernity and Self-Identity: Self and Society in the Late Modern Age*. Oxford: Blackwell.
Giles, H., Coupland, J., & Coupland, N. (Eds.). (1991). *Contexts of Accommodation*. Cambridge: Cambridge University Press.
Giles, H., & Powesland, P. F. (1975). *Speech Style and Social Evaluation*. London: Academic Press.
Goffman, E. (1959). *The Presentation of Self in Everyday Life*. New York: Doubleday.

Goffman, E. (1974). *Frame Analysis: An Essay on the Organization of Experience* Norwich: Fletcher & Son.

Goldinger, S. D. (1996). Words and Voices: Episodic Traces in Spoken Word Identification and Recognition Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1166-1183.

Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *Journal of the Acoustical Society of America, 102*(1), 655-658.

Halliday, M. A. K. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.

Hay, J., & Drager, K. (2007). Sociophonetics. *Annual Review of Anthropology, 36*, 89-103.

Hay, J., & Drager, K. (in press). Stuffed toys and speech perception. *Linguistics*.

Hay, J., Drager, K., & Warren, P. (in press). Exposure to one dialect affects processing of another. *Language and Speech*.

Hay, J., Maclagan, M., & Gordon, E. (2008). *New Zealand English*. Edinburgh: Edinburgh University Press.

Hay, J., Nolan, A., & Drager, K. (2006). From Fush to Feesh: Exemplar Priming in Speech Perception. *The Linguistic Review, 23*, 351-379.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*(5), 3099-3111.

Hintzman, D. (1986). "Schema Abstraction" in a Multiple-Trace Memory Model. *Psychological Review, 93*(4), 411-428.

Hoit, J. D., Jenks, C. L., Watson, P. J., & Cleveland, T. F. (1996). Respiratory function during speaking and singing in professional country singers. *Journal of Voice, 10*(1), 39-49.

Howard, D. M. (2009). Acoustics of the trained versus untrained singing voice. *Current Opinion in Otolaryngology & Head and Neck Surgery, 17*, 155-159.

Hymes, D. H. (1968). The Ethnography of Speaking. In J. A. Fishman (Ed.), *Readings in the Sociology of Language* (pp. 99-138). The Hague: Mouton.

Irvine, J. T. (2001). Style as distinctiveness: The culture and ideology of linguistic differentiation. In P. Eckert & J. R. Rickford (Eds.), *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. H. Bower (Ed.), *The psychology of learning and motivation: advances in research and theory* (Vol. 18, pp. 1-48). London: Academic Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity in phonology. *Journal of Phonetics, 34*, 485-499.

Johnson, K., Strand, E., & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics, 27*, 359-384.

Johnstone, B., & Kiesling, S. F. (2008). Indexicality and experience: Exploring the meaning of /aw/-monophthongization in Pittsburgh. *Journal of Sociolinguistics, 12*(1), 5-33.

Kiesling, S. F. (1998). Men's identities and sociolinguistic variation: The case of fraternity men. *Journal of Sociolinguistics, 2*(1), 69-99.

Kingston, J. (2007). The phonetics-phonology interface. In P. V. De Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 401-434). Cambridge: Cambridge University Press.

Kuiper, K., & Austin, J. P. M. (1990). They're off and racing now: the speech of the New Zealand race caller. In A. Bell & J. Holmes (Eds.), *New Zealand ways of speaking English* (pp. 195-220). Clevedon: Multilingual Matters.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (2001). The anatomy of style-shifting. In P. Eckert & J. R. Rickford (Eds.), *Style and Sociolinguistic Variation* (pp. 85-108). Cambridge: Cambridge University Press.

Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English*. Berlin: Mouton de Gruyter.

Ladefoged, P., & Broadbent, D. E. (1957). Information Conveyed by Vowels. *Journal of the Acoustical Society of America, 29*(1), 98-104.

Le Page, R. B., & Tabouret-Keller, A. (1985). *Acts of Identity: Creole-based Approaches to Language and Ethnicity*. Cambridge: Cambridge University Press.

Lee, E. (1982). The foremost medium - the voice In G. Vulliamy & E. Lee (Eds.), *Pop, Rock and Ethnic Music in School* (pp. 102-126). Cambridge: Cambridge University Press.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*, 71-102.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86.

Milkshake. (2007, September 18). *Singing in your accent? Annoying?* Retrieved November 4, 2009, from http://mog.com/milkshake/blog/112189

Morrissey, F. A. (2008). Liverpool to Louisiana in one lyrical line: Style choice in British rock, pop and folk singing. In M. A. Locher (Ed.), *Standards and norms in the English language* (pp. 195-218). Berlin: Mouton de Gruyter.

Myers-Scotton, C. (1993). *Social motivations for codeswitching*. Oxford: Clarendon.

MySpace. (n.d.-a). *Broken Heartbreakers*. Retrieved April 29, 2010, from http://www.myspace.com/brokenheartbreakers

MySpace. (n.d.-b). *Dylan Storey*. Retrieved April 29, 2010, from http://www.myspace.com/dylanstorey

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18*(1), 62-85.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357-395.

O'Hanlon, R. (2006). Australian Hip Hop: A Sociolinguistic Investigation. *Australian Journal of Linguistics, 26*(2), 193-209.

Ochs, E. (1992). Indexing gender. In A. Duranti & C. Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 335-358). Cambridge: Cambridge University Press.

Peirce, C. S., Hartshorne, C., & Weiss, P. (1935). *Collected papers of Charles Sanders Peirce, Volumes 5-6*. Cambridge, Mass.: Harvard University Press.

Pennycook, A. (2003). Global Englishes, Rip Slyme, and performativity. *Journal of Sociolinguistics, 7*(4), 513-533.

Pennycook, A., & Mitchell, T. (2009). Hip hop as dusty foot philosophy: Engaging locality. In H. S. Alim, A. Ibrahim & A. Pennycook (Eds.), *Global linguistic flows: hip hop cultures, youth identities, and the politics of language* (pp. 25-42). New York and London: Routledge.

Peterson, G. E., & Barney, H. L. (1952). Control Methods Used in a Study of the Vowels. *Journal of the Acoustical Society of America, 24*(2), 175-184.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. P. J. H. Bybee (Ed.), *Frequency Effects and Emergent Grammar*. Amsterdam: John Benjamins.

Potter, J. (1998). *Vocal Authority: Singing Style and Ideology*. Cambridge: Cambridge University Press.

Prince, E. F. (1987). Sarah Gorby, Yiddish folksinger: a case study of dialect shift. *International Journal of the Sociology of Language, 67*, 83-116.

Prince, E. F. (1988). Accommodation Theory and Dialect Shift: A Case Study from Yiddish. *Language and Communication, 8*(3/4), 307-320.

R Development Core Team. (2009). R: A language and environment for statistical computing. Vienna, Austria.

Rampton, B. (1995). *Crossing*. London: Longman.

Rampton, B. (1999). Styling the other: Introduction. *Journal of Sociolinguistics, 3*(4), 421-427.

Rampton, B. (2006). *Language in Late Modernity: Interaction in an Urban School*. Cambridge: Cambridge University Press.

Recording Industry Association of New Zealand. (n.d.). *The official New Zealand music chart*. Retrieved December 10 2009, from http://www.rianz.org.nz/rianz /chart.asp

Schilling-Estes, N. (1998). Investigating "self-conscious" speech: The performance register in Ocracoke English. *Language in Society, 27*, 53-83.

Schilling-Estes, N. (2004). Constructing ethnicity in interaction. *Journal of Sociolinguistics, 8*(2), 163-195.

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language and Communication, 23*, 193-229.

Simpson, P. (1999). Language, culture and identity: With (another) look at accents in pop and rock singing. *Multilingua, 18*(4), 343-367.

Smith, G. (2005). *Singing Australian: The history of folk and country music*. North Melbourne: Pluto Press.

Stone, R. E. E., Cleveland, T. F., & Sundberg, J. (1999). Formant Frequencies in Country Singers' Speech and Singing. *Journal of Voice, 13*(2), 161-167.

Stone, R. E. E., Cleveland, T. F., Sundberg, J., & Prokop, J. (2003). Aerodynamic and acoustical measures of speech, operatic, and broadway vocal styles in a professional female singer. *Journal of Voice, 17*(3), 283-297.

Sundberg, J. (1987). *The Science of the Singing Voice*. Dekalb: Northern Illinois University Press.

Sundberg, J. (2009). Articulatory Configuration and Pitch in a Classically Trained Soprano Singer. *Journal of Voice, 23*(5), 546-551.

Sundberg, J., & Skoog, J. (1997). Dependence of Jaw Opening on Pitch and Vowel in Singers. *Journal of Voice, 13*(2), 161-167.

Thomas, E. (2002). Instrumental Phonetics. In J. K. Chambers, P. Trudgill & N. Schilling-Estes (Eds.), *Handbook of Language Variation and Change*. Cambridge: Blackwell Publishers.

Trudgill, P. (1980). Acts of conflicting identity: A sociolinguistic look at British pop songs. In M. V. de Silva (Ed.), *Festschrift for R. B. Le Page. York Papers in Linguistics, 9*.

Trudgill, P. (1983). *On Dialect: Social and Geographical Perspectives*. Oxford: Blackwell.

Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.

Warren, P., Hay, J., & Thomas, B. (2007). The loci of sound change effects in recognition and perception. In J. Cole & J. I. Hualde (Eds.), *Laboratory phonology 9* (pp. 87-112). Berlin: Walter de Gruyter.

Wells, J. C. (1982). *Accents of English*. Cambridge: Cambridge University Press.
Wenger, E. (1998). *Communities of Practice*. Cambridge: Cambridge University Press.
Wikipedia. (n.d.). *Small Faces*. Retrieved August 5, 2009, from http://en.wikipedia.org/wiki/Small_Faces
Wray, A. (1999). Singers on the Trail of 'Authentic' Early Modern English: The Puzzling Case of /æ:/ and /ɛ:/. *Transactions of the Philological Society 97*(2), 185-211.
Yaeger-Dror, M. (1991). Linguistic evidence for social psychological attitudes: Hyperaccommodation or (r) by singers from a Mizrahi background. *Language and Communication, 11*(4), 309-331.

Appendix A Ethics Forms for Production Study

AUT UNIVERSITY
TE WĀNANGA ARONUI O TĀMAKI MAKAU RAU

MEMORANDUM

Auckland University of Technology Ethics Committee (AUTEC)

To: Allan Bell
From: **Madeline Banda** Executive Secretary, AUTEC
Date: 7 November 2008
Subject: Ethics Application Number 08/218 **Vowel differences between singing, reciting and speaking.**

Dear Allan

Thank you for providing written evidence as requested. I am pleased to advise that it satisfies the points raised by the Auckland University of Technology Ethics Committee (AUTEC) at their meeting on 13 October 2008 and that I have approved your ethics application. This delegated approval is made in accordance with section 5.3.2.3 of AUTEC's *Applying for Ethics Approval: Guidelines and Procedures* and is subject to endorsement at AUTEC's meeting on 8 December 2008.

Your ethics application is approved for a period of three years until 7 November 2011.

I advise that as part of the ethics approval process, you are required to submit the following to AUTEC:

- A brief annual progress report using form EA2, which is available online through http://www.aut.ac.nz/about/ethics. When necessary this form may also be used to request an extension of the approval at least one month prior to its expiry on 7 November 2011;
- A brief report on the status of the project using form EA3, which is available online through http://www.aut.ac.nz/about/ethics. This report is to be submitted either when the approval expires on 7 November 2011 or on completion of the project, whichever comes sooner;

It is a condition of approval that AUTEC is notified of any adverse events or if the research does not commence. AUTEC approval needs to be sought for any alteration to the research, including any alteration of or addition to any documents that are provided to participants. You are reminded that, as applicant, you are responsible for ensuring that research undertaken under this approval occurs within the parameters outlined in the approved application.

Please note that AUTEC grants ethical approval only. If you require management approval from an institution or organisation for your research, then you will need to make the arrangements necessary to obtain this.

When communicating with us about this application, we ask that you use the application number and study title to enable us to provide you with prompt service. Should you have any further enquiries regarding this matter, you are welcome to contact Charles Grinter, Ethics Coordinator, by email at charles.grinter@aut.ac.nz or by telephone on 921 9999 at extension 8860.

On behalf of the AUTEC and myself, I wish you success with your research and look forward to reading about it in your reports.

Yours sincerely

Madeline Banda
Executive Secretary
Auckland University of Technology Ethics Committee

Cc: Andy Gibson andy.gibson@aut.ac.nz

Participant Information Sheet

Date Information Sheet Produced:

26 September 2008

Project Title

Singing, Reciting, Speaking

An Invitation

My name is Andy Gibson. I wish to invite you to be involved in my research. Remember that your participation in this experiment is totally voluntary and you can withdraw at any time if you want to.

What is the purpose of this research?

I am running this experiment to learn about the differences and similarities between singing and speaking. This experiment is part of my study for the MPhil degree and the results will also be presented at conferences and may be published in academic journals.

How was I chosen for this invitation?

You were chosen for this invitation because you expressed interest in being involved after we discussed the project. In order to be involved in this research you need to be an active musician who writes, performs and records original songs. You also need to have spent most of your life living in New Zealand. If you have spent more than five years living in another country you will not be eligible to participate in the research.

What will happen in this research?

The experiment will take about 45 minutes. You will be recorded singing a song you have written (you can record some accompaniment before singing the vocal), reciting the words to that song, reading a list of words and singing that list of words along to a backing track. After this, we will have an informal discussion about music and singing, this conversation will also be recorded.

How will my privacy be protected?

In order to discuss the findings of the project in relation to your public profile and musical style, we would like to be able to use your name. However, we will only do this if you give your explicit consent for us to do so. If you would prefer for your identity to be confidential, that is fine.

We would also like to be able to play short excerpts (of up to 10 seconds) of the recordings at academic conferences. Once again, we will only do this if you give your consent.

What will happen to the recordings and data?

Data will be kept for a period of at least six years. It may be kept for a longer period to allow for further statistical analysis and comparison with future experiments using similar methodologies. When the data is no longer required, all electronic files kept on CD will be destroyed. Any other physical records will also be destroyed securely.

How do I agree to participate in this research?

If you would still like to participate in this research, please fill out the consent form provided with this sheet.

Will I receive feedback on the results of this research?

Andy will send you a summary of the findings in early 2009.

What do I do if I have concerns about this research?

Any concerns regarding the nature of this project should be notified in the first instance to the Project Researcher, Andy Gibson, andy.gibson@aut.ac.nz, 921-9999 ext 8476

Concerns regarding the conduct of the research should be notified to the Executive Secretary, AUTEC, Madeline Banda, madeline.banda@aut.ac.nz, 921 9999 ext 8044

Whom do I contact for further information about this research?

Researcher Contact Details: Andy Gibson. Email: andy.gibson@aut.ac.nz Ph. 09-921-9999 ext 8476

Project Supervisor Contact Details: Allan Bell. Email: allan.bell@aut.ac.nz Ph. 09-921-9683

Approved by the Auckland University of Technology Ethics Committee on 7 Nov 2008, AUTEC Reference number 08/128.

Consent Form

Project title: ***Singing, Reciting, Speaking***
Project Supervisor: ***Allan Bell***
Researcher: ***Andy Gibson***

- I have read and understood the information provided about this research project in the Information Sheet dated 26 September 2008.
- I have had an opportunity to ask questions and to have them answered.
- I **do / do not** (please circle) give consent for my name to be used in the presentation of results at conferences and in publications.
- I understand that my voice will be recorded. I **do / do not** (please circle) give consent for short excerpts (up to 10 seconds) of these recordings to be played at conferences.
- I understand that I may withdraw myself or any information that I have provided for this project at any time prior to completion of data collection, without being disadvantaged in any way.
- If I withdraw, I understand that all relevant information, including recordings, will be destroyed.
- I agree to take part in this research.

Participant's signature: .....

Participant's name: .....

Date:

Approved by the Auckland University of Technology Ethics Committee on 7 Nov 2008, AUTEC Reference number 08/218.

Appendix B Interview Schedule

- Who are some of your favourite bands/artists?
- What plans do you have for your music?
- Who do you consider your audience to be? New Zealand based or international?
- Have you thought about the accent you sing in?
- Do you feel like your singing accent is different to your speaking accent?
- Are there any examples that stand out?
- How important is it to you that your persona as a singer reflects your normal day-to-day persona?
- Do you have any other comments on the way words are pronounced in singing?

Appendix C Additional Tables and Figure

Table C.1: *Mean F1 and F2 Values for Dylan's Vowels in Reciting and Interview Data, with the Results of a Wilcoxon Rank Sum Test on the Differences*

	F1				F2			
	rec. (n)	int. (n)	diff.	p	rec.	int.	diff.	p
DRESS	412 (8)	406 (10)	6	NS	1841	1856	-15	NS
TRAP	471 (10)	547 (12)	-76	.	1772	1780	-8	NS
THOUGHT	388 (14)	410 (12)	-22	.	838	904	-66	.
LOT	492 (8)	550 (8)	-58	NS	1103	1155	-52	NS
START	680 (1)	739 (2)	-59	NA	1306	1414	-108	NA
GOOSE (a)	382 (2)	432 (2)	-50	NA	1530	1508	22	NA
GOOSE (b)	321	310	11	NA	1866	1662	204	NA
GOOSE (b-a)	-61	-123	62	NA	336	154	182	NA
GOAT (a)	507 (15)	576 (9)	-69	**	1256	1328	-72	*
GOAT (b)	413	359	54	.	1412	1612	-200	***
GOAT (b-a)	-95	-217	122	***	156	285	-129	*

(*** for p<.001; ** for p<.01; * for p<.05; . for p<.1; NS for not significant; NA for cases where there are too few pairs to warrant a test for significance)

Table C.2: *Mean F1 and F2 Values for Andrew's Vowels in Reciting and Interview Data, with the Results of a Wilcoxon Rank Sum Test on the Differences*

	F1				F2			
	rec. (n)	int. (n)	diff.	p	rec.	int.	diff.	p
DRESS	369 (10)	385 (12)	-16	NS	1862	1874	-12	NS
TRAP	560 (9)	537 (15)	23	NS	1861	1820	41	NS
THOUGHT	396 (12)	395 (15)	1	NS	859	833	26	NS
LOT	520 (4)	612 (4)	-92	NS	1030	986	44	NS
START	682 (3)	695 (6)	-13	NS	1358	1293	65	.
GOOSE (a)	437 (2)	454 (3)	-17	NA	1231	1388	-157	NA
GOOSE (b)	334	367	-33	NA	1502	1532	-30	NA
GOOSE (b-a)	-103	-87	-16	NA	271	144	127	NA
GOAT (a)	573 (13)	618 (9)	-45	*	1189	1194	-5	NS
GOAT (b)	409	428	-19	NS	1382	1410	-28	NS
GOAT (b-a)	-164	-190	26	NS	193	217	-24	NS

(*** for p<.001; ** for p<.01; * for p<.05; . for p<.1; NS for not significant; NA for cases where there are too few pairs to warrant a test for significance)

Table C.3: *Mean F1 and F2 Values for John's Vowels in Reciting and Interview Data, with the Results of a Wilcoxon Rank Sum Test on the Differences*

	F1				F2			
	rec. (n)	int. (n)	diff.	p	rec.	int.	diff.	p
DRESS	413 (13)	367 (8)	46	*	2059	2184	-125	NS
TRAP	528 (15)	486 (4)	42	NS	1808	2010	-202	*
THOUGHT	384 (13)	421 (8)	-37	NS	853	832	21	NS
LOT	490 (18)	519 (10)	-29	NS	1028	1043	-15	NS
START	656 (4)	708 (6)	-52	NS	1413	1385	28	NS
GOOSE (a)	344 (2)	427 (5)	-83	.	1639	1553	86	NS
GOOSE (b)	241	354	-113	.	1881	1806	75	NS
GOOSE (b-a)	-103	-73	-30	NS	243	253	-10	NS
GOAT (a)	424 (10)	545 (13)	-121	***	1349	1441	-92	NS
GOAT (b)	321	447	-126	**	1618	1692	-74	NS
GOAT (b-a)	-103	-98	-5	NS	269	250	19	NS

(*** for p<.001; ** for p<.01; * for p<.05; . for p<.1; NS for not significant; NA for cases where there are too few pairs to warrant a test for significance)

Table C.4: *Tests for Spearman Correlations between F1 and f0, for all Singers, Showing Correlation Coefficient (ρ) with Significance Level in Brackets*

	Dylan		Andrew		John	
	Sing: ρ (p)	Speak	Sing	Speak	Sing	Speak
DRESS	.14 (NS)	.30 (NS)	.17 (NS)	-.13 (NS)	.54 (.)	-.34 (NS)
TRAP	.79 (***)	.71 (***)	.30 (NS)	-.1 (NS)	.03 (NS)	-.27 (NS)
THOUGHT	.48 (.)	.48 (*)	.18 (NS)	-.15 (NS)	.40 (NS)	.24 (NS)
LOT	.45 (NS)	.31 (NS)	.50 (NS)	.12 (NS)	-.24 (NS)	.42 (*)
START	NA	1 (NS)	1 (NS)	.37 (NS)	.4 (NS)	.73 (*)
GOOSE (a)	-.80 (NS)	.80 (NS)	1 (NS)	-.05 (NS)	-1 (NS)	.60 (NS)
GOOSE (b)	.40 (NS)	-.40 (NS)	1 (NS)	.7 (NS)	.50 (NS)	.54 (NS)
GOAT (a)	.56 (*)	.28 (NS)	.60 (*)	.35 (NS)	.72 (*)	-.07 (NS)
GOAT (b)	.67 (**)	-.20 (NS)	.67 (**)	.29 (NS)	.25 (NS)	.32 (NS)
PRICE (a)	.66 (***)	-.01 (NS)	.11 (NS)	.26 (NS)	-.22 (NS)	.22 (NS)
PRICE (b)	.24 (.)	-.20 (NS)	-.11 (NS)	-.59 (.)	-.20 (NS)	-.06 (NS)

(*** for p<.001; ** for p<.01; * for p<.05; . for p<.1; NS for not significant; NA for cases where there are too few pairs to warrant a test for significance)

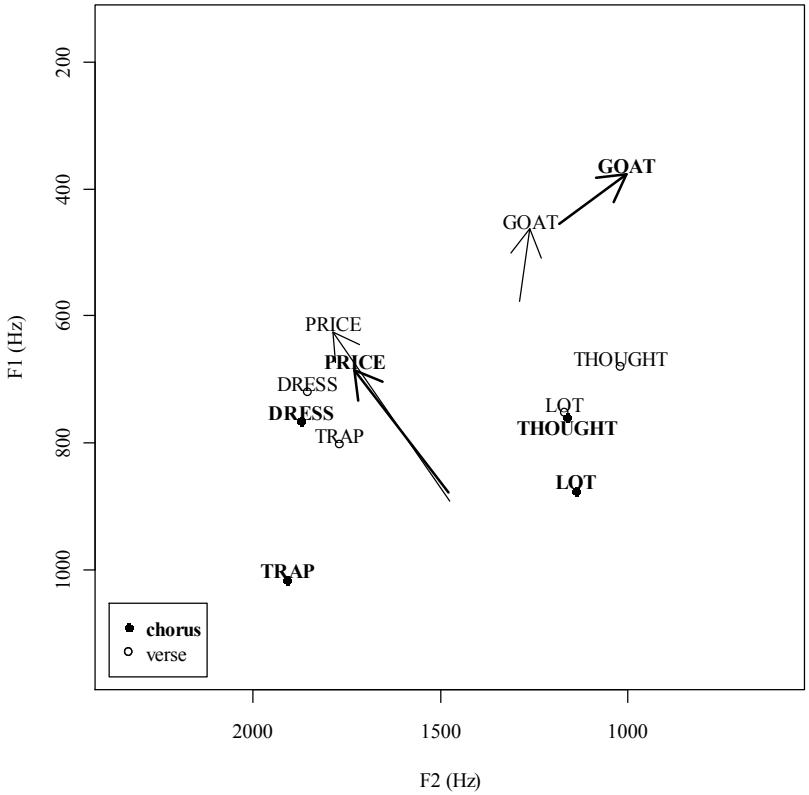

Figure C.1: Comparison of sung pronunciation in choruses and verses for Dylan

Appendix D Lyrics

All lyrics are reproduced here with the authors' kind permission. All rights to the lyrics are reserved by the songwriters.

'Sold it all away'

Lyrics copyright Dylan Storey, from the album *Out of the Soup.*

Hunger is sharper than a Tiger's tooth
and love is a blind lizard on the roof.
My thoughts are a spider creeping down the wall
and I'll follow my baby to the final call.

Throw me a lemon and I'll catch a peach.
Drop me a question with one in the breach.
Wonder is heavy as an old man's sigh
and I'm watching the morning with an Eagle's eye.

So I know I might have sold it all away.
So I know I might have sold it all away.

So wrap me in plastic and I'll save my breath.
I'll swallow a potion to postpone my death.
I'm drinking my weight five times in wine
and I'll float like a feather over space and time.

So I know I might have sold it all away.
So I know I might have sold it all away.

Looking for something is like falling over,
running a race in which the winner must die.
Luck is a landing in a three leaf clover.
Never rely on a good time staying, love is alive in the morning by evening must die.

So I know I might have sold it all away.
So I know I might have sold it all away.

Could love survive a full attack by reason?
Pushing and pulling on the sanity line.
You try to keep warm but your hands keep freezing,
never rely on a good time staying, hope is alive in the morning by evening must die.

‘Gloria’

There's something in the way her hair hangs and falls like your heart does
Slowing down speeding up as she does
I didn't want to know she owns the street that you live on
Run run nails that she paints on

Gloria does anyone love Gloria
It's so cold but she'll burn tonight.
Gloria when are you coming home?

There's something in the way her right hand man falls back when she says so
Working relationship, he knows
they appreciate clean sheets and the fish bowl by the bed
The extra mile for the extra visit

Gloria does anyone love Gloria
It's so cold but she'll burn tonight.
Gloria when are you coming home?

There was something in the way I drove that night that magnetised her to my blue car
What have we here, a young one in a blue car
Banging down my window
She was banging down my window

Gloria does anyone love Gloria
It's so cold but she'll burn tonight.
Gloria when are you coming home?

‘Wintersun’

Lyrics copyright John Howell.

I think it's time my time begun
Think it's time I felt the winter sun
I think it's time for me to ease up some
Think it's time I felt the winter sun

Many thoughts I've thought have weighed a ton
Many fights I’ve fought I haven't won
Many dreams I've dreamed have never come
Think I've earnt myself some winter sun

Go on, go on, go on
Go on, go on, go on

Everybody's waiting for the one
Thinking that that day will never come
When it does your whole world is undone
Start again beneath the winter sun.

Start again beneath the winter sun.
Start again beneath the winter sun.

Hey, go lightly on yourself

Go on, go on, go on
Go on, go on, go on

‘Ballad of Archibald Baxter’

Lyrics copyright John Howell.

I’m not gonna fight, I’m not gonna fight
I won’t be your soldier, I won’t lose my soul

So take me downtown, you’re not gonna break me down

I’m not gonna kill, I’m not gonna kill
I won’t kill my brother

So take me downtown, you’re not gonna break me down

I’m fighting for love, I’m fighting for love
You say I’m a coward but that isn’t so

So take me downtown, you’re not gonna break me down

‘Mi Corazon’

Lyrics copyright John Howell.

You used to walk beside me, take me by the hand
When I felt the shadow you would hold me

Listen to my story, ask about my day
Then one day the shadow came and took you away

Leaving me
Leaving me

Mi corazon es la piedra

I can hear you calling, calling out my name
I’ll always try to find you, I’ll keep turning

I see you when I’m dreaming, I wish I could stay
But morning always comes along and takes you away

Leaving me
Leaving me

Mi corazon es la piedra

‘Tell that Boy’

Lyrics copyright John Howell.

Tell that boy there’s a world outside
Tell that boy the sun will surely rise tomorrow
Tell that boy if she doesn’t feel it it’s not love
Tell that boy she doesn’t have to be the one

Tell that boy no-one’s got it in for ya
Tell that boy to breathe two three four

Appendix E Tracklist for CD

Songs analysed.

The audio material on tracks 1-18 of this CD is included thanks to the kind permission of the three artists. All rights to the lyrics, music, and recordings are reserved by the songwriters. Please do not copy or distribute these recordings.

Dylan Storey: 'Sold it all away'
1. Full recording
2. Unaccompanied sung vocal
3. Reciting of lyrics

Andrew Keoghan: 'Gloria'
4. Full recording
5. Unaccompanied sung vocal
6. Reciting of lyrics

John Guy Howell: 'Wintersun'
7. Full recording
8. Guitar and vocal recording made for this study
9. Unaccompanied sung vocal
10. Reciting of lyrics

John Guy Howell: 'Mi Corazon'
11. Full recording
12. Unaccompanied sung vocal
13. Reciting of lyrics

John Guy Howell: 'Tell that Boy'
14. Full recording
15. Unaccompanied sung vocal
16. Reciting of lyrics

John Guy Howell: 'Ballad of Archibald Baxter'
17. Unaccompanied sung vocal
18. Reciting of lyrics

Materials for perception experiment.

19. *bed-bad* continuum
20. *bed-bad* continuum after pitch-shifting
21. Example of music condition
22. Example of no-music condition

Appendix F Response Sheet for Perception Experiment

SINGING

Instructions:

- You will hear a New Zealander ***singing*** either 'bed' or 'bad'.
- The word will be repeated twice, and then there is a short break for you to circle the word you heard.
- There are no right or wrong answers here, so just go with your first intuition and circle whichever word you heard.

1.	bed	bad
2.	bed	bad
3.	bed	bad
4.	bed	bad
5.	bed	bad
6.	bed	bad
7.	bed	bad
8.	bed	bad
9.	bed	bad
10.	bed	bad
11.	bed	bad
12.	bed	bad
13.	bed	bad
14.	bed	bad
15.	bed	bad
16.	bed	bad
17.	bed	bad
18.	bed	bad
19.	bed	bad
20.	bed	bad

Appendix G Ethics Forms for Perception Experiment

AUT UNIVERSITY
TE WĀNANGA ARONUI O TAMAKI MAKAU RAU

MEMORANDUM

Auckland University of Technology Ethics Committee (AUTEC)

To: Allan Bell
From: **Madeline Banda** Executive Secretary, AUTEC
Date: 4 April 2008
Subject: Ethics Application Number 08/41 **The perception of sung and spoken vowels in New Zealand English.**

Dear Allan

Thank you for providing written evidence as requested. I am pleased to advise that it satisfies the points raised by the Auckland University of Technology Ethics Committee (AUTEC) at their meeting on 10 March 2008 and that and I have approved your ethics application. This delegated approval is made in accordance with section 5.3.2.3 of AUTEC's *Applying for Ethics Approval: Guidelines and Procedures* and is subject to endorsement at AUTEC's meeting on 12 May 2008.

Your ethics application is approved for a period of three years until 4 April 2011.

I advise that as part of the ethics approval process, you are required to submit the following to AUTEC:

- A brief annual progress report using form EA2, which is available online through http://www.aut.ac.nz/about/ethics. When necessary this form may also be used to request an extension of the approval at least one month prior to its expiry on 4 April 2011;
- A brief report on the status of the project using form EA3, which is available online through http://www.aut.ac.nz/about/ethics. This report is to be submitted either when the approval expires on 4 April 2011 or on completion of the project, whichever comes sooner;

It is a condition of approval that AUTEC is notified of any adverse events or if the research does not commence. AUTEC approval needs to be sought for any alteration to the research, including any alteration of or addition to any documents that are provided to participants. You are reminded that, as applicant, you are responsible for ensuring that research undertaken under this approval occurs within the parameters outlined in the approved application.

Please note that AUTEC grants ethical approval only. If you require management approval from an institution or organisation for your research, then you will need to make the arrangements necessary to obtain this.

When communicating with us about this application, we ask that you use the application number and study title to enable us to provide you with prompt service. Should you have any further enquiries regarding this matter, you are welcome to contact Charles Grinter, Ethics Coordinator, by email at charles.grinter@aut.ac.nz or by telephone on 921 9999 at extension 8860.

On behalf of the AUTEC and myself, I wish you success with your research and look forward to reading about it in your reports.

Yours sincerely

Madeline Banda
Executive Secretary
Auckland University of Technology Ethics Committee

Cc: Andy Gibson andy.gibson@aut.nz

Participant Information Sheet

Date Information Sheet Produced:

28 April 2008

Project Title

Perception of Words

An Invitation

My name is Andy Gibson. I wish to invite you to be involved in my research. Remember that your participation in this experiment is totally voluntary and you can withdraw at any time if you want to.

What is the purpose of this research?

I am running this experiment to learn about how people perceive language. This experiment is part of my study for the MPhil degree and the results will also be presented at conferences and may be published in academic journals.

How was I chosen for this invitation?

You were chosen for this invitation because you responded to one of the advertisements placed around the university campus or to an email from me or a mutual friend. The only selection criteria involved in this research is that you need to have spent most of your life living in New Zealand. If you have spent more than five years living in another country you will not be eligible to participate in the research.

What will happen in this research?

The experiment will take about 15 minutes. You will be played a series of words through headphones and you will circle on a piece of paper which word you heard. There are no right or wrong answers. I am just interested in your first reaction. I will also record you reading out a short paragraph.

How will my privacy be protected?

Only the primary researcher will know your identity. The data you provide will be stored with an arbitrary ID number and not your name to protect your privacy. Only the researchers involved in this project will have access to the data that is collected.

How do I agree to participate in this research?

If you would still like to participate in this research, please fill out the consent form provided with this sheet.

Will I receive feedback on the results of this research?

Results of the experiment will be available in July 2008. If you would like to learn more about these results you can contact Andy (see contact details below).

What do I do if I have concerns about this research?

Any concerns regarding the nature of this project should be notified in the first instance to the Project Researcher, Andy Gibson, *andy.gibson@aut.ac.nz*, 921-9999 ext 8476

Concerns regarding the conduct of the research should be notified to the Executive Secretary, AUTEC, Madeline Banda, *madeline.banda@aut.ac.nz*, 921 9999 ext 8044.

Whom do I contact for further information about this research?

Researcher Contact Details:

Andy Gibson. Email: andy.gibson@aut.ac.nz Ph. 09-921-9999 ext 8476

Project Supervisor Contact Details:

Allan Bell. Email: allan.bell@aut.ac.nz Ph. 09-921-9683

Approved by the Auckland University of Technology Ethics Committee on 4 April 2008, AUTEC Reference number *08/41*.

Consent Form

Project title: ***Perception of Words***

Project Supervisor: ***Allan Bell***

Researcher: ***Andy Gibson***

- ○ I have read and understood the information provided about this research project in the Information Sheet dated 28 April 2008.
- ○ I have had an opportunity to ask questions and to have them answered.
- ○ I understand that presentation of results at conferences and in publications will protect my privacy and that my identity will not be revealed.
- ○ I understand that my voice will be recorded and that only the researchers working on this project will have access to this recording. It will not be played to any other people.
- ○ I understand that I may withdraw myself or any information that I have provided for this project at any time prior to completion of data collection, without being disadvantaged in any way.
- ○ If I withdraw, I understand that all relevant information, including recordings, will be destroyed.
- ○ I agree to take part in this research.

Participant's signature: .....

Participant's name: .....

Date:

Approved by the Auckland University of Technology Ethics Committee on 4 April 2008. ***AUTEC Reference number*** 08/41

Participant Information:

1. *Age (please circle):* **under 20** **20-29** **30-39** **40-49** **50-59** **60+**
2. *Occupation:* _____
3. *Ethnicity:* _____
4. *Total years spent living* ***outside*** *New Zealand:* **0-5** **6-10** **>10**
5. *Favourite kinds of music:* _____
6. *Are you fully fluent in any other languages?* **Y** / **N**

   If so, which languages: _____