Experience and learning in cross-dialect perception: Derhoticised /r/ in Glasgow

Robert Lennon

MA(Hons), MSc

University of Glasgow

College of Arts

School of Critical Studies

English Language & Linguistics

Thesis submitted to the University of Glasgow in fulfilment of the requirements for the degree of Doctor of Philosophy

Abstract

It is well known that unfamiliar accents can be difficult to understand. Previous research has investigated the effect of hearing e.g. foreign-accented speech, but relatively little research has been conducted on the effect of hearing an unfamiliar native English accent. This thesis investigates how listeners process fine phonetic detail for a phonological contrast, measuring their perceptual efficiency depending on their level of experience with the working class Glaswegian dialect.

In Glasgow, speakers are stereotypically rhotic. However, recent sociophonetic research indicates a trend towards derhoticisation (the phonetic erosion of postvocalic /r/ in working class Glaswegian speech). The potential for misperception exists when listeners hear minimal pairs such as *hut/hurt*, when spoken by working class speakers who realise postvocalic /r/ as an acoustically ambiguous variant, with delayed tongue tip gesture and early tongue body gesture. This makes derhoticised /r/ perceptually very similar to the preceding open back vowel in both *hut* and *hurt*, leading to difficulty when listeners try to distinguish between /CʌC/ and /CʌrC/ words in general.

This thesis begins with a novel dynamic acoustic analysis of the key cues for rhoticity in Glaswegian for such words, demonstrating that minimal pairs such as *hut/hurt* are acoustically very similar in the Glaswegian working class accent, but remain distinct for middle class speakers.

A suite of listening experiments is then described. Experiment 1 was conceived and run as a pilot study to this work, but new analysis of the data assessed the influence of long-term learning, using Signal Detection Theory as a key analytical tool. This showed strong effects of listener familiarity in both sensitivity and response bias.

Experiment 2 tested listeners' ability to learn the distinction between *hut* and *hurt* word types, with Response Time and Signal Detection analyses finding that listeners least familiar with the accent very quickly matched the response patterns of listeners with an intermediate level of experience. However, neither listener group was able to match the performance of listeners most familiar with the accent, who were native to Glasgow, suggesting that acoustic phonetic detail plays an important role in perception, with interesting interactions with listener experience.

Finally, for Experiment 3 the overarching purpose shifts from offline to online perception. The results of the acoustic analysis showed that dynamics are key, so Experiment 3, a Mouse Tracking experiment, allows for the measurement of dynamic perceptual responses – in a listening context which is more difficult – for the most experienced listeners. It yielded results in terms of Response Time, and two sets of measures capturing cursor trajectories, Area Under the Curve, and Discrete Cosine Transformation. Taken together, the results of these analyses reveal that even for the most experienced listeners (Glaswegians), the phonetically ambiguous tokens present perceptual challenges when hearing working class Glaswegian *hut* and *hurt* words, and also demonstrated that challenging listening conditions lead to processing costs, even for the 'easiest' stimuli; i.e. when hearing talkers and accents randomised together.

This thesis examines a single difficult phonological contrast, with the simplicity of the linguistic scope affording an extremely in-depth analysis. Not only did this provide a clear insight into the perception of the contrast itself, but the depth of analysis allows for a more sophisticated discussion of the results, potentially speaking to wider theoretical standpoints. The results have implications for theories of speech perception, as they may be explained by some general principles which underlie exemplar theories and Bayesian inference. They also constitute valuable acoustic and perceptual contributions to the ongoing research into the complex and changing nature of postvocalic /r/ in Scotland.

Contents

List of Tables

List of Figures

Acknowledgements

Firstly, I owe a huge debt of gratitude to my two supervisors, Rachel Smith and Jane Stuart-Smith. They have been with me throughout the last five years, listening to my ramblings about experimental ideas, and coming up with many (mostly better) suggestions of their own. They have been supportive in the extreme, always ready to encourage me to push my limits when I needed it. But mainly I owe them so much coffee...

Huge thanks are also due to all the folks in Glasgow University Laboratory of Phonetics: Julia Moreno, James Parnell-Mooney, Duncan Robertson, Vijay Solanki, Ewa Wanat, and Fabienne Westerberg. They have graciously put up with me pestering them about R and LaTeX, both questions and answers (all right, Vijay, mostly questions...!), and of course going on about Star Trek. A very special mention must go to my good friend and labmate Farhana Shaukat Alam, who is greatly missed by us all.

I must also thank the tireless admin staff in the University of Glasgow's School of Critical Studies (especially Alison Bennett and Kiran Faisal), Dr Rachael Jack in Glasgow's School of Psychology for allowing me to use their Experiment Subject Pool, Professor Francis Nolan at the University of Cambridge, who very kindly allowed me to use his lab for my first two experiments, and my five anonymous Glaswegian speakers, without whom I would have had literally no data.

Finally, I couldn't have done any of this without my family, who have been an unwavering source of support on this journey.

This work was supported by the Economic and Social Research Council, grant no. ES/J500136/1.

Declaration

I declare that, except where explicit reference is made to the contribution of others, this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signature ________________________________

Printed Name Robert Lennon

For Laura and Zoe.

Part I

Introduction

Chapter 1

Literature review

1.1 Introduction

Speech perception is a complex but crucial process for understanding each other, and it mostly operates in a surprisingly efficient manner. However, it is well known that speech can sometimes be misunderstood. In particular, perception can be more difficult when the listener is confronted with certain challenges, potentially interfering with their understanding of the speaker's intended message.

Challenges to perception may include a lack of familiarity with certain characteristics of the speech signal, or the presence of multiple speakers. For example, if the listener is very familiar with the speaker or the accent they are hearing, and if there is only one speaker in a quiet room, it is logical that they would find it relatively easy to correctly and rapidly process the speech they hear. However, the addition of challenges such as hearing an accent with unfamiliar phonetic features, hearing an unfamiliar speaker, or processing the speech of more than one person, can adversely affect the listener's ability to accurately receive the intended message.

This thesis directly deals with the first of these challenges – accent familiarity – and in doing so it indirectly deals with some issues surrounding the others – speaker familiarity and processing of multiple talkers. Thus, the focus of this thesis is the detailed measurement of listeners' perception of one particular fine-grained, dialect-specific phonetic realisation of a phonological contrast, in order to assess the impact of these challenges.

The accent feature which will be used in this investigation is postvocalic /r/ in the traditionally rhotic dialect of Glasgow, the largest city in Scotland. In a process known as derhoticisation, postvocalic /r/ is becoming weaker over time in the working class Glaswegian sociolect, giving rise to some interesting perceptual consequences. For example, if a speaker who uses such a variant produces the word *hut*, followed by the word *hurt*, this has the potential to cause confusion, as the /r/ in *hurt* may be barely audible to the listener. The potential for confusion may be even greater if the listener has little experience of the Glaswegian accent. It is hoped that the investigation described in this thesis can shed light on how a listener's perception is affected by their familiarity with an accent, as well as their ability to learn and potentially adapt to phonetic detail. Also under investigation is the role of more challenging listening conditions in the perception of this feature. The themes are wide ranging, so this chapter provides the key theoretical background, together with relevant information for the production and perception of rhoticity, providing the basis for the research questions for this thesis.

Section 1.2 covers some of the theoretical positions in speech perception, moving from a statement of the position of abstractionist theories, through discussions of exemplar, then hybrid, theories, and finishing with recent approaches that are based on Bayesian inference. Section 1.3 is a review of a number of studies which investigate how listeners perceive different dialects. Section 1.4 then introduces rhoticity, providing a general overview, then narrows down to a description of how the phenomenon is changing over time in Glasgow. Finally, section 1.5 is a discussion of the way in which people perceive rhoticity, including the findings of a number of perceptual studies which have tested listeners' sensitivity to this complex aspect of speech depending on their level of experience, as well as a specific examination of the perception of rhoticity in non-rhotic accents, and concludes with studies which examine the perception of rhoticity in Scotland.

1.2 Theoretical approaches to speech perception

This section will provide a general overview of some of the theoretical approaches to answering the question of how we perceive and understand speech. In general, theories have moved away from models which propose a primarily abstract organisation of perception, towards a more nuanced approach. Nevertheless, in order to provide context for the discussion, the first section will briefly discuss abstractionist theories of speech perception, then move to exemplar theories, then hybrid approaches, finishing with a short discussion of Bayesian modelling and its influence on speech perception theory.

1.2.1 Abstractionist theories

Abstractionist approaches to speech perception typically claim that when a listener hears speech, it is abstracted away from the acoustic input signal, which can be highly variable depending on speaker or accent, or distorted depending on listening environment (formalised in Chomsky & Halle 1968). Such theories typically involve a reduction of information, which is the 'decoding of specific episodes (tokens) into canonical representations' (Goldinger 1996: 1166), converting the signal into discrete categories (e.g. McClelland & Elman 1986; Studdert-Kennedy 1976), or phoneme strings (Halle 1985, cited in Smith 2013: 6).

Goldinger notes that support for such theories is often motivated by the fact that speech is variable in the extreme (1996: 1167). The fact that we can perceive speech relatively easily in the face of such variability makes it tempting to assume that people perceive categories, with natural variation being akin to noise in the signal. In order to cope with this variation, many abstractionist theories propose that normalisation occurs when a listener encounters speech sounds, such that they 'translate' what they hear back into a set of higher-level categories (e.g. Brown & Carr 1993; Carr, Brown & Charalambous 1989; Green, Kuhl, Meltzoff & Stevens 1991; Jackson & Morton 1984).

In discussing their results for bimodal perception of faces and voices, Green et al. (1991) write that there were two main approaches taken by speech perception theorists, when it comes to explaining bimodality phenomena such as the McGurk effect (McGurk & MacDonald 1976). These were actually two main approaches taken by abstractionists, but the paper was published before the emergence of competing approaches such as exemplar theory. They describe proponents of motor theory (e.g. Fowler 1986; Fowler & Rosenblum 1991; Liberman & Mattingly 1985, 1989), who write that listeners derive, or directly perceive, the articulatory movements that produced the speech signal that they heard (1991: 534). The second approach described by Green et al. is the mapping of 'different metrics derived separately from the auditory and visual modalities onto underlying phonetic representations prototypes' (1991: 534), citing studies which demonstrate the use of prototypes in phonetic categorisation (Kuhl 1991; Miller, Connine, Schermer & Kluender 1983; Samuel 1982). Motor theory was later challenged by papers such as Ohala's 'Speech perception is hearing sounds, not tongues' (1996), which argued that the primacy of such theories over others cannot yet be justified, as other theoretical positions have not yet had the time to be fully tested.

Theories which promoted an abstracting of variation in the speech signal to discrete categories seemed to be an attractive explanation for the human ability to perceive speech in a relatively effortless fashion. However, such theories began to fall short when presented with mounting evidence from studies which highlighted the importance of variation to the listener. Such studies will be discussed in the next section.

1.2.2 Exemplar theories

An alternative approach to speech perception then developed out of influential work reported through the 1990s (e.g. Mullennix & Pisoni 1990; Mullennix, Pisoni & Martin 1989; Nygaard, Sommers & Pisoni 1994; Palmeri, Goldinger & Pisoni 1993; Goldinger 1996, 1998, 2000), and is now known as Exemplar theory (or Episodic theory, and occasionally Instance theory). Key to this approach is the assumption that every perceived occurrence of speech is recorded and stored by the listener, and over time categories of occurrences emerge from the way the information is stored. Every time the listener hears speech they match each of the constituent parts of the incoming signal (phonetic features, phonemes, words, etc.) against their set of stored representations, which consists of all their previously stored exemplars. In this matching process, the categorisation of each part of the incoming signal is determined by goodness of match to the stored exemplars, enabling the listener to make a probabilistic judgement about e.g. which phoneme was perceived, or intended to be produced by the talker. Therefore there is no need to store abstract forms, and categories arise simply because of the distributions of the stored exemplars, encoded in terms of values along parameters of phonetic space (Smith 2013: 7).

Johnson (1997; cited in Evans and Iverson 2004) wrote that listeners may be able to perform speaker normalisation if they use their stored exemplars of speech to evaluate words produced by similar talkers. This would be a counter proposal against abstractionist theories which, as described above, explained speaker normalisation in terms of the listener accessing invariant information and effectively stripping away variation, then moving the abstracted signal into a higher level in the perception process. Indeed, coping with accent variation could be seen as an extreme form of the same process, with the present speaker's acoustic signal being compared to all other signals from other speakers. This would result in ‘accent normalisation’ without the stripping away of variation (Nygaard & Pisoni 1998, cited in Evans and Iverson 2004, 2007).

In a ‘same-different’ classification task, Cole et al. (1974) found that listeners responded faster to sounds heard in the same voice as the previous stimulus, and this effect held whether the sound was the same or different. This lends further support to the theory that aspects of a speaker's voice are stored along with the phonemes the listener hears, and these aspects contribute to the message being understood.

Smith (2013) writes that there has been considerable debate about whether the representations of each exemplar are ‘holistic episodes, of flexible size and structure’, or whether they are more comparable to particular types of linguistic unit

(2013: 235). This consideration highlights a major issue with a purely exemplar based approach. If one assumes that *everything* the listener has ever heard is stored, then how does the listener organise this into meaningful 'bits' of data, rather than simply a stream of sound? The next section will outline a number of approaches that have attempted to provide an answer.

1.2.3 Hybrid theories

More recently, hybrid models have become popular, allowing for both the storage of exemplars, as well as more abstract categories as described by traditional models. Such models advocate the preservation of talker-specific and other indexical information, but that abstract, symbolic representations still exist (Schacter & Church 1992, cited in McQueen 2005: 264). Pierrehumbert (2002, 2006) sets out the case that allophonic details are 'systematically associated with words' (2002: 19), but that categories are still important for the organisation of the incoming signal into meaningful 'bits' of sound.

Smith (2015) argues that a broader view of speaker-specific phonetics should be taken, from the available phonetic and perceptual evidence. Subtle differences exist in the variability between speakers, for example in the way they mark prosodic boundaries, among other features. These are termed 'prosodic signatures' (2015: 5). A small number of studies examine the way that listeners use their knowledge of speaker-specific phonetic detail in order to facilitate recognition of tokens related in some dimension; e.g. place or manner. Listeners can learn many aspects of speaker-specific phonetic detail, and the patterns of the transfer of category characteristics are principled, not arbitrary. Smith also writes that the existence of rich hierarchical (prosodic, grammatical) structures in the Polysp model (Hawkins & Smith 2001; Hawkins 2003, 2010) 'improves the process of pattern-matching between signal and memories' (2015: 24). Hybrid models therefore represent a more promising conceptual approach to modelling individual variation, and the effect this has on the perception of speech.

In order to allow for adequate interpretation by the listener of the fine details of speech, Goldinger (2007) revises the pure episodic approach of Goldinger (1996) into an approach that is hybrid in that it assumes that 'abstract knowledge is imposed upon each encountered stimulus'. Therefore, 'each stored exemplar is actually a product of perceptual input combined with prior knowledge, the precise balance likely affected by many factors' (Goldinger 2007: 50). Goldinger's new position strongly advocates some sort of 'combined' model of speech perception. Various positions have been taken in order to promote an integrated approach to the co-existence of abstract categories and exemplars. These positions allow for phonetic features (Nielsen 2011; Kraljic & Samuel 2006), variation of allophones in different positions in the word or syllable (Dahan & Mead 2010; Smith & Hawkins 2012), and phonemes (McQueen, Cutler & Norris 2006). This means that a wide range of hierarchical links between categories and detail have been proposed.

1.2.4 Bayesian inference

Bayesian approaches to speech perception are an important and relatively recent direction of research (e.g. Scharenborg, Norris, ten Bosch & McQueen, 2005; Norris & McQueen, 2008; Clayards, Tanenhaus, Aslin & Jacobs, 2008; Feldman, Griffiths, & Morgan, 2009; Kleinschmidt & Jaeger 2015; Norris, McQueen & Cutler 2016). The Bayesian approach to linguistic phenomena requires certain computational analyses to be applied to the data, thus allowing judgements to be made about linguistic patterns. Bayesian inference can either be used for specific computational modelling of speech perception or as a general informing framework. The latter is the approach taken in this thesis. This section summarises some recent work which has been influential in driving forward the Bayesian approach to modelling speech perception.

Smith (2013) defines the role of Bayesian inference in speech perception theory, writing that 'decisions can be made based on knowledge or expectation in combination with evidence.' Under this approach, 'hypotheses are associated with prior probability distributions...which give probabilities of any hypothesis being true', prior to any current evidence for a hypothesis (2013: 5). This is contrasted with posterior probability distributions, which 'reflect the probability of a hypothesis being true given the current [i.e. previously known] evidence' (2013: 5). A word which has many and/or highly probable neighbours may be harder to recognise than one with fewer neighbours. Bayesian inference may be a good explanation for the perceptual magnet effect.

Kleinschmidt, Weatherholtz & Jaeger (2018) examine the relationship between the social and linguistic aspects of speech perception, claiming that the link between them is very strong. They investigate the question of how social perception and speech perception interact. They emphasise that 'speech perception [seems to be] sensitive to talkers' social identity', such that '[listeners] interpret acoustic cues differently based on a talker's perceived socio-indexical features, such as regional origin (Hay & Drager, 2010; Niedzielski, 1999), gender (Strand, 1999; Johnson, Strand, & D'Imperio, 1999), age (Walker & Hay, 2011), and individual identity' (2018: 3). Because of this, they claim that speech perception is '*conditioned* on socio-indexical features' (2018: 3). This idea is formalised and computationally implemented in the model of the *ideal adapter* (Kleinschmidt & Jaeger 2015), a theoretical listener who overcomes the problem of lack of invariance by conditioning speech perception on socio-indexical cues. This model allows for the link between the listener's recognition of the linguistic units (words, phones, etc.), and their knowledge of the probabilistic co-occurrence of linguistic units, socio-indexical features, and acoustic cues. The ideal adapter is based on Bayesian inference, in that the listener *probabilistically infers* the likelihood of each possible linguistic unit, given their prior knowledge of the distribution in question. Smith interprets their view of adaptation as 'an update in the listener's talker- or situation-specific beliefs about the linguistic generative model' (2015: 14).

In the work of Kleinschmidt and colleagues, the central point of the ideal adapter model is the fact that a listener has a set of prior beliefs about linguistic patterns and distributions, and that they are updated when new information arises. This new information may be that there is a new speaker, or a new listening environment, or a number of other factors – for example, speaker A will likely have different cue distributions for Voice Onset Time (VOT) in /b/ and /p/ than speaker B, and they are each likely to have different /b, p/ distributions from those of the listener's perceived 'typical talker'. An ideal adapter learns an *internal model* of talker variability, which includes factors such as 'socio-indexical features that are informative about the cue distributions that an unfamiliar talker might produce'. Such knowledge supports socially-conditioned linguistic inference, because listeners can change their expected linguistic cue distributions, depending upon the available information about the talker, such as age, gender, region of origin, etc. (Kleinschmidt & Jaeger, 2015: 176). Moreover, 'the same probabilistic knowledge supports inferences about a talker's socio-indexical group membership' (2015: 180-1). That is, it is possible to infer information about an unfamiliar talker such as age, sex, and regional origin, based on their cue distributions alone.

Using cue distributions conditioned by a range of socio-indexical variables such as sex, age, and dialect, Kleinschmidt et al. showed that their Bayesian framework 'can be used to make *social*, as well as linguistic, inferences' (2018: 4).

The paper makes the conclusion that from this framework, a number of implications arise for both psycholinguists and sociolinguists. For psycholinguists, the framework appears to provide some explanation for the recent work in exemplar theory suggesting that listeners do not give equal weighting to every single occurrence of a phonetic feature that they hear, but that there is at least some structure imposed on the variation that they experience. This consideration also speaks to hybrid models, which go even further than exemplar theory in claiming that there is structure, into which exemplars can be organised. According to the predictions of the ideal adapter, this structure is 'adapted to the statistics of their experience and their task goals' (2015). One implication for sociolinguists is that if sociolinguistic perception can be treated as inference under uncertainty, then the ideal observer can 'provide a link between patterns of socially-conditioned speech production and listeners' socio-linguistic perceptions'.

One argument in their paper is the possibility of a deep, yet largely unexplored, connection between sociolinguistics and psycholinguistics. This is due to the claim that their 'results suggest that social inferences can be supported by the same statistical knowledge that is necessary for robust speech perception in the face of talker variability (Kleinschmidt & Jaeger, 2015)'. Finally, their approach seems to provide a 'formal, quantitative framework for connecting variation in the world to listeners' internal models'.

These sections have provided a very brief discussion of the main approaches theorists have taken to address the challenge of how we perceive speech. These standpoints will be referred to (where relevant) in the discussions of each of the experimental chapters, and again in the general discussion in Chapter 6.

1.3 Speech perception in dialects

When a listener hears an unfamiliar accent, various factors can be affected, such as the ease with which they perceive the speech, and judgements about the geographical region the speaker is from. First, this section will review articles which have studied the way that listeners categorise accents, then it will look at how different dialects affect the comprehensibility and intelligibility of speech. Finally, listeners' adaptation to unfamiliar accents will be discussed.

1.3.1 Accent categorisation

Listeners' accuracy in categorising accents can tell us about the role of experience in accent perception, as well as the effect of certain salient dialect markers, among other factors. In an accent categorization experiment in which listeners were asked to identify the regional origin of a speaker in the Netherlands and in the UK, Van Bezooijen and Gooskens (1999) found that Dutch listeners identified the region of the speaker with 60% accuracy, and the specific province with 40% accuracy. In the same study, British listeners identified the region of the speaker with 88% accuracy, and the area with 52% accuracy. These results show that listeners were generally good at making broad judgements about where the speaker was from, but performed less well when more detail was required. They found that pronunciation varied in listeners' ratings, and that prosodic features play a minor role in identification. Clopper and Pisoni (2004) conducted a similar study, which tested whether listeners could correctly identify various American accents, by noting where they thought the speaker was from using a map on an interactive touchscreen. Their results showed significantly poorer accuracy than Van Bezooijen and Gooskens' study, at only 30% accuracy for categorising speakers into dialect groups. This was noted to be similar to the accuracy level reported by Garrett, Coupland and Williams (1999) in their perceptual study of Welsh dialects.

It may be thought that dialect categorization could be related to stereotypical social judgements about the speakers of that dialect. Clopper, Rohrbeck and Wagner researched the effect that an autism spectrum disorder (ASD) has on a person's perception of different dialects of American English (2012). People with high-functioning autism are known to exhibit difficulties in the social functions of language, but typically display intact perceptual processing. Clopper et al. found that while participants with an ASD could classify speakers into dialect groups with a similar success rate as typically developing participants, they performed less well in a language attitudes task, by not associating appropriate social stereotypes to different dialect groups (2012: 752). The results of Clopper et al.'s study might be taken as evidence that dialect categorisation is not directly linked to social judgements, but the assignment of a particular attitude or judgement about an accent is a social, and possibly cultural factor, separate from any categorisation based on phonology.

As well as categorisation, social judgements of geographical areas have been found to affect listeners' perception of phonetic differences between similar accents of English. Hay and Drager (2010) found that, when primed by the presence of stuffed toys in the experiment room which represented either the concept of New Zealand (kiwi) or Australia (koala), listeners from New Zealand perceived different vowel qualities in the same stimuli, depending on the prime. A similar result was found by Hay, Nolan and Drager (2006), in which the experimental prime was either the word ‘Australian’ or ‘New Zealand’, handwritten in the corner of the participant's answer sheet.

1.3.2 Intelligibility and comprehensibility

Intelligibility and comprehensibility are both said to suffer as a result of a listener's inexperience with an unfamiliar accent. Intelligibility is defined as the sentence-level ‘success-rate’ of the transmission of the intended message, whereas comprehensibility is usually measured at word level (often with reaction-time data), because it is a function of the cognitive effort required to correctly identify the word. In a study concentrating on the effect of a foreign accent, Munro and Derwing (1995) found that native English listeners experienced a greater processing cost when evaluating the comprehensibility of sentences in Mandarin-accented English speech, than when listening to similar sentences produced by native English speakers. However, their results did not indicate that the degree of 'accentedness' played a significant role in the processing load, suggesting that the presence of a strong accent is not necessarily a barrier to communication (1995: 302).

In a cross-dialect study, Floccia, Goslin, Girard and Konopczynski (2006) presented native French speakers from East-central France with stimuli (high-frequency disyllabic words and pseudowords in carrier sentences) spoken in an unfamiliar dialect from Southern France. The results showed that there was a processing cost associated with the unfamiliar dialect, when subjects were asked to identify the target words. Furthermore, the processing cost reduced as the number of syllables presented to the subject increased, and from this they cautiously suggest that full adaptation to the unfamiliar accent occurs (2006: 1289). Floccia, Butler, Goslin and Ellis (2009) followed this up by testing the intelligibility and comprehensibility of accented speech. Their paper describes an interesting, and possibly counterintuitive, pair of results. They found that increased exposure to the unfamiliar dialect aided intelligibility, that is, correct identification of the words. However, reaction time data showed that the exposure did not aid comprehensibility. A similar failure in short-term adaptation was also reported in Dutch listeners by Adank and McQueen (2007), who found that comprehension of isolated words in an unfamiliar dialect of Dutch did not improve after short-term exposure. Taken together, these results could suggest a number of possibilities. First, listeners may use the longer utterances in the intelligibility tasks as a way of more easily adapting to a recently encountered accent, as there is simply a greater number of distinctive segments than in the word-level tasks. If this were the case, then listeners would be using a combination of features to 'tune in' to the accent, so would therefore be able to more easily predict the kind of segments that are likely to arise in the rest of the utterance, thus enabling the perception of the message in the signal. Short-term learning (often based on reaction time data) of only one distinctive phoneme could suffer as a result of the very fact that there is only one of them in a comprehensibility task. Second, because word-level comprehensibility tasks focus attention on the phonetic detail of the segment, and information at segment level may be stored in a different way, this may explain why the level of short-term learning was found to be different in the above studies.

The effect of differing linguistic experience between accent groups should also be considered. In an experiment in which listeners from both Glasgow and London were asked to assign a truth value to sentences such as 'Tomato soup is a liquid' and

'Tomato soup is people' presented in noise, Adank, Evans, Stuart-Smith and Scott (2009) found that the comprehension of native and non-native accents did not pattern equally across the two listener groups. The listeners from London, who spoke with a SSBE accent, were slower at responding to Glaswegian stimuli than to SSBE stimuli, but the Glaswegian listeners were equally fast in responding to stimuli in both accents. These results suggest that native accent aids speech comprehension, but also that increased familiarity with a non-native accent has an important role to play. The authors claim that because the Glaswegian listeners were relatively familiar with the SSBE accent, due to the influence of an English-based media and the cosmopolitan nature of Glasgow University, they had an advantage over the SSBE listeners in this respect (2009: 19-21). In a similar finding, Sumner and Samuel (2009: 493) found that listeners who have experience with more than one dialect are more flexible when it comes to processing different forms of a word, than those who have much more experience with one dialect over the other.

The studies discussed above have hinted at listener adaptation – listeners seem capable of adjusting, maybe even within the short timeframe of an experiment. The next section will discuss what is known about adaptation, and how it can be investigated.

1.3.3 Perceptual adaptation

We have already looked at a number of studies which measure the effects of long-term exposure on intelligibility and comprehension of unfamiliar accented speech, and potentially ambiguous speech sounds. However, another dimension to accent familiarity is the fact that listeners can adapt to unfamiliar speech surprisingly quickly (although this may depend upon a number of factors). Researchers may use a number of methods for investigating short-term listener adaptation, but there is a general tendency to a common methodology, with some variation: Pretest, to gather information on the listener's existing performance, Exposure, to build up a short-term learning of the feature/accent in question, and Posttest, to measure how much, if at all, the listener has improved in their performance. This section details a number of studies, and the methodologies they use.

Maye et al. (2008) used an Exposure-Test design to investigate the mechanism by which listeners adjust their perception of speech produced with a slightly different accent to one they are already familiar with. In two sessions on different days, listeners heard two near-identical versions of the same story, a 20-minute section of the film 'The Wizard of Oz', spoken in the same voice in the standard North American accent. The only difference between the two versions was that the voice in the second story was acoustically altered so that vowels were lowered, e.g. *witch* was produced as *wetch*, and *west* was produced as *wast*. Immediately following each of the exposure sessions was a lexical decision task, where listeners were asked to indicate whether each single-word stimulus was a word or a nonword. The results of this experiment showed that listeners adapted their interpretation of what could be classed as a real word, with more e.g. *wetch* and *wast* stimuli being accepted as real after the lowered-vowel-accent story. Furthermore, the listeners' interpretation of the vowels generalized to include words which had not been presented in the exposure phase.

A second experiment in the same study had an identical design in which the exposure story was the same as the other experiment, i.e. had lowered vowels, but the front vowels of the nonword stimuli in the lexical decision task were raised, not lowered, (e.g. *witch* became *weetch*). This time, listeners did not increase their endorsement rates for raised front vowels after hearing the lowered vowels passage. Taken together, both experiments showed that listeners shifted their interpretation of phonemes in a novel accent only in the direction of the acoustic shift. In other words, listeners did not simply relax their category boundaries when hearing an unfamiliar accent, but followed the specific direction of the difference.

While the methodology of the above study was to simply play a passage of speech to the listeners, many researchers ask the participants to complete a task, to keep their attention focused on the audio signal. Adank et al. (2009) found positive effects of long-term exposure in Glaswegian listeners hearing SSBE speech, with SSBE listeners incurring a processing cost when hearing Glaswegian accent. In an earlier study, Adank and McQueen (2007) conducted a short-term experiment for the effects of unfamiliar accent on processing individual words, this time in Dutch. The exposure task in this study was an animacy decision task, where listeners had to decide whether individual nouns presented in the unfamiliar accent described animate objects, e.g. *hoen* 'hen' and *stoel* 'chair'. It therefore acted as a distractor task, while maintaining the listeners' attention on the content. The authors found that listeners were adversely affected by hearing the unfamiliar accent: i.e. their response times were longer than when they heard the familiar accent as the stimuli. However, even though there were 20 minutes of exposure to an unfamiliar regional accent, there was no effect of short-term learning on the listeners' response times in the animacy decision task, which represented a failure to adapt to individual words in the unfamiliar accent.

In a foreign-accent study, Japanese listeners were successfully trained to perceive differences between English /r/ and /l/, following a training period which involved a minimal-pair identification task (Logan, Lively & Pisoni 1991). This appears to demonstrate that adults' perceptual phonetic categories are plastic, so they can adapt given the appropriate stimuli. Later retesting showed that the participants retained their recently-acquired categories, months after the initial training period (Pisoni et al. 1994). In another study, Japanese participants were successfully trained to produce more accurate /r/ and /l/ in English, after short-term exposure to real recordings of native GenAmerican words (Bradlow et al. 1997). These participants also retained their new categories of /r/ and /l/ production months later (Bradlow et al. 1999). Bradlow & Bent (2008) later found that listeners achieved talker-independent adaptation to Chinese-accented speech, after exposure to multiple talkers of Chinese-accented English.

Norris, McQueen & Cutler (2003) investigated the effect of short-term learning of, and adaptation to, an ambiguous segment in Dutch. They edited the speech of a Dutch talker so that the final sound in her [f]-final words (e.g. *witlof* ‘chicory’) was ambiguous to the listener group, who performed a lexical decision task on isolated words, but her [s]-final words (e.g. *naaldbos* ‘pine forest’) were unchanged. A second group of listeners heard the opposite: the same ambiguous final sound in [s]-final words, with the [f]-final words unchanged. The ambiguous sound was modified so it was a fricative, acoustically halfway between [f] and [s]. The listeners in the first group learned to interpret the ambiguous sound as [f], and the second group learned to interpret the same sound as [s]. During their test phase, the listeners who had heard the ambiguous sound in [f]-final words tended to label fricatives on a continuum as [f], with the opposite pattern for the other group (a tendency to label them as [s]). These results demonstrate listeners using lexical knowledge to categorise ambiguous sounds. They also highlight the fact that learning took place when exposure to the ambiguous fricatives was only 20 target words, spread over a list of 100 words and 100 nonwords. Investigating a similar question, Clarke & Luce (2005) found that listeners shifted their mean categorization boundary to a shorter VOT, after exposure to English sentences in which syllable-initial /t/ and /d/ were modified to have short-lag /t/s and prevoiced /d/s. However, a follow up study showed that the same modified /t/-/d/ stimuli did not promote a shift in categorization in a different place of articulation, specifically /g/ and /k/.

To test the retention of this learning effect over time, Eisner & McQueen (2006) replicated the study by Norris et al. (2003), but added a second post-test after 12 hours. Another difference was that the ambiguous words were embedded in a story, rather than presented as isolated words. For one group, the 12-hr delay was from morning to late evening (day group), and for the other group the 12-hr delay was overnight (night group). The authors of this study found the same effect as Norris et al. (2003) in the immediate post-test, thus replicating the effect but with a different task (exposure to a story instead of a lexical decision task). Furthermore, this learning effect was retained after 12 hours; i.e. there was no change in categorization of the ambiguous sound along the [f]-[s] continuum between the immediate post-test, and the 12-hr post-test, for either the day group or the night group. Eisner & McQueen interpret these results as evidence for high stability in a listener's lexically driven perceptual adjustments in response to talker idiosyncrasies, adding to Kraljic & Samuel's (2005) evidence from a similar experimental design that learning effects are reliable after a 25 minute interval, unless the unambiguous tokens that come from the voice of the exposure talker are presented to the listener.

Eisner, Melinger & Weber (2013) used an exposure-test design with a control group to investigate whether listeners could learn final-/d/ devoicing, a well-known feature of Dutch-accented English. They presented one group of native English listeners with words with devoiced alveolar stops (e.g. seed [si:t^h]), in an initial exposure phase which consisted of words, pseudowords, and filler items as stimuli, and instructions to indicate whether they thought the stimuli were real English words, by pressing yes or no buttons. The control group did not hear the target words, but additional filler items in their place. The test phase for all groups consisted of a similar lexical decision task, but the stimuli were auditory primes presented with visual target words and pseudowords. The results of this study showed that the experimental group successfully learned to interpret the devoiced items as real words.

Floccia et al. (2006) found no effect of speaker-specific adaptation after a set of short-term exposure experiments. Floccia et al. (2009) took this work further by investigating the interaction between intelligibility and comprehensibility, which are two different measures of the cost of processing accented speech. The study used short-term word-spotting tasks, and found that accent changes cause a temporary perturbation in reaction times, but that listeners did not habituate to the new accent, as even though they displayed adaptation with their improved accuracy, the delay did not decrease.

Short-term learning experiments are useful when investigating adaptation to unusual pronunciations of certain words and phonemes, but this methodology can also be applied when investigating unusual pronunciations of segments that indicate morphological structure. Barden & Hawkins (2013) investigated the effect of exposure on a listener's ability to adapt to atypical pronunciation of a morphological feature of English, specifically the /ri:/ prefix in words like re-build. In order to test for listener adaptation, they recorded paragraphs in the exposure phase such that all /ri:/-prefix words were produced with an atypical pronunciation of the vowel, [rɪ]. A control group heard sentences which were identical, except they would hear the typical pronunciation of the prefix, i.e. [ri:]. They found that once the listeners in the test group had been exposed to a 19-minute listening task they scored higher than the control group (who had heard the typical pronunciations in exposure) in an intelligibility-in-noise task that included the atypical pronunciations. In the task the listeners were asked to report what they heard by typing the sentences after hearing them. Crucially, all listeners heard novel keywords in the test phase; that is, /ri:/-prefix words had not already appeared in the exposure phase. The authors therefore conclude that listeners in the test group had learned an association between the atypical [rɪ] pronunciations and the orthographic re-prefix. This study highlights listeners' ability to quickly adapt to a speaker's unusual pronunciation of a morphological feature of English, and to then generalize from a small amount of speech data to the speaker's pronunciation of the entire morphological class.

The use of short-term learning experiments, which employ various different methodologies and examine an even greater range of linguistic features, can tell us a great deal about the mechanisms of listener adaptation to an unfamiliar accent. Maye et al. (2008) write that top-down knowledge may be helpful for enabling listeners to quickly note the lexical intent of an utterance and remap the vowel space. Presumably, this also applies to other classes of speech sound, for example fricatives (Norris, McQueen and Cutler 2003) and VOT (Clarke and Luce 2005; Yu, Abrego-Collier and Sonderegger 2013). However, Floccia et al. (2006) did not find significant adaptation to speakers of various regional accents, when using lexical decision tasks. When suggesting possible explanations for Floccia et al.'s results compared with their own, Adank and McQueen (2007) suggest that effects of accent familiarity may be stronger in animacy decision than in lexical decision. Given the varied success in the results of the studies described above, it could be assumed that listeners can adapt to some changes in speech, but not others. It is possible that there are certain categories of sounds that they can adapt to, but it is equally possible that the differences in experimental design between the different studies could have an effect on their results.

Tomé Lourido and Evans (2015) looked at the effect of a small adjustment in a speaker's production, on their perception of that feature. In order to investigate ‘learners' potential to build new phonetic representations or modify existing ones’ (2015: 1), they investigated the speech of people who were raised as Spanish monolinguals, but at a later stage chose to change to primarily speaking Galician, for a variety of ideological or cultural reasons. These speakers, known as *neofalantes*, were (mostly) found to be able to produce a distinction in their mid vowels (/e/-/ɛ/ and /ɔ/-/o/) when they spoke Galician, where no such distinction exists in Spanish. However, in an experiment testing their perception of natural Galician stimuli, they were unable to reliably detect these vowel distinctions. Not only does this suggest that production changes and perception changes are independent processes for the *neofalantes* (or, at least, *not fully integrated* processes), but the authors also assert that they may be behaving like learners of a second language, processing their newly dominant language, Galician, through categories which exist in their former dominant language, Spanish (2015: 4). Tomé Lourido and Evans suggest that this finding may be attributed to the difficulty of processing certain phonetic contrasts in a second language (Best, 1995, Flege, 1995; cited in Lourido and Evans, 2015).

This effect could also happen when a person moves to a new dialect area. When this happens, the person can adapt their perception to accommodate to unfamiliar features. Evans and Iverson (2004) presented listeners from both Northern and Southern England with synthetic vowel stimuli embedded within carrier sentences. The listeners were asked to report the best matches for vowel quality, when synthetic vowel stimuli were presented within the carrier sentence ‘I'm asking you to say the word [ ] please’, produced in either a Northern or a Southern English accent. The findings suggested that listeners were able to make alterations in their best exemplar representations of vowel qualities, in order to adjust for when they heard a non-native dialect in the carrier sentence. Evans and Iverson suggest that their results are consistent with motor theory, in that listeners perceive speech based on the terms of their own articulatory gestures (2004: 359), but that auditory targets may play a role in a listener's modification of their accent when they move to a new area: they may have to change their notions of best exemplars, in order to successfully modify their accent. In a subsequent experiment, Evans and Iverson (2007) tested subjects with Northern English accents, but this time the focus was on plasticity in both perception and production of vowels. They found that speakers adjusted both their vowel perception and production after attending university, though the results were complex.

This section has discussed a wide range of speech perception experiments which have looked at the role of experience and learning. The specific focus of this thesis is the perception of rhoticity, so section 1.5 will look at studies which have investigated how /r/ is perceived. However, since rhoticity can be a particularly complicated aspect of language, the next section (1.4) is dedicated to explaining what rhoticity is and how it can vary between articulatory variants, through change over time, and due to social factors.

1.4 Rhoticity

The study of perception and listener experience in this thesis uses the phenomenon of rhoticity, to measure how listeners' perception of fine phonetic detail can vary across certain factors. Rhoticity is an ideal testing ground for such a study, given the range of observed variants as well as the continual observation of rhotic variants showing sociolinguistic stratification, in English in general (e.g. Labov 1986), and in Scottish English in particular (e.g. Lawson, Scobbie & Stuart-Smith 2014).

1.4.1 Auditory and articulatory properties of /r/

A rhotic accent is one in which postvocalic /r/, as in words like *car* and *card*, is pronounced (Wells 1982). Postvocalic /r/ may be realised as one of a number of different variants, and Maddieson and Ladefoged state that across the world's languages there is a core membership of the class of rhotics which display a ‘single or repeated brief contact between the tongue and a point on the upper surface of the vocal tract’ (1996: 182), which is a general description of trills, taps and flaps. Aside from this ‘core class’, rhotics can vary in terms of their place and manner of articulation, and whether or not they are voiced. Therefore, most rhotic variants display acoustic features which are distinct from those of other rhotic sounds. Even so, some authors have attempted to unify the acoustic features of all /r/ allophones, in order to more precisely define what it is that makes a rhotic sound ‘rhotic’. Lindau (1985) attempts to answer this question by bringing together many different kinds of /r/-realisations from various languages, and puts forward the notion that rhotics display a ‘family resemblance’ of articulatory and acoustic parameters. These parameters can be shared by two rhotic sounds; for example, a voiced alveolar trill shares the fact that it displays a pulse pattern with both a voiceless alveolar trill [r̥] and a voiced uvular trill [R]. It also shares its place of articulation with an alveolar tap [ɾ] and an alveolar approximant [ɹ]. Uvulars all have a high third formant, and alveolar approximants all have a low third formant. In this way, Lindau's family resemblance of rhotics can be built into a network of similarity parameters which links all rhotics together. One possible criticism of this theory of family resemblance is that it is conceivable for any two speech sounds to be linked, simply because they share a certain parameter. If an alveolar approximant [ɹ] can be linked to both a voiced uvular fricative [ʁ] and a voiced alveolar trill [r] because they all share the parameter of being sonorant, it may be possible to link an alveolar approximant to the class of vowels, which also display the parameter of sonorance, or to an alveolar plosive, which shares its place of articulation. Indeed, Maddieson and Ladefoged point out that whether or not two sounds are ‘rhotic’, they are likely to have similar spectral properties if they have similar constriction locations (1996: 245). They also suggest that the overall unity of the class of rhotics may simply be due to the historical connections between the subgroups of trills, approximants, and so on, and that they are all represented by the letter ‘r’ (1996: 245).

The approximant rhotic variants, postalveolar and retroflex, found in many varieties of English, including American and Scottish English, have formants like those found in vowels. The existence of formant frequencies in sonorant approximants means we can apply analysis techniques similar to those already used for analysis of vowels (even though, as Plug and Ogden point out, rhotics are ‘commonly treated as complex segments, with a consonantal component and a vocalic component’ (2003: 160)). In acoustic theory for vowels, the third formant is predicted to be relatively lower, and close to the second formant, if there are constrictions in the lower pharyngeal region, or in the post-alveolar or palatal region (Lindau, 1985). Alveolar approximants have constrictions in both the lower pharynx and the palate, as does retroflex [ɻ] in the same American accent (Lindau, 1985). By using an x-ray camera in conjunction with a microphone, Delattre and Freeman observed that a lowering of F3 towards F2 directly correlated with a narrowing of the constriction at the palato-velar region (1968: 50).

Rhotic (or ‘r-coloured’) vowels always have a lowered third formant (Maddieson and Ladefoged, 1996: 313), even though they may be produced with different articulatory configurations. Indeed, Lindau claims that American speakers use ‘all available articulatory mechanisms’ in order to produce an approximant /r/ with a low F3 (1985: 165). Although the existence of a lowered third formant is said to be important for a strong percept of rhoticity in rhotic approximants (e.g. Delattre and Freeman, 1968: 46; Lindau, 1985: 163, 165; Maddieson and Ladefoged, 1996: 244; Espy-Wilson, Boyce, Jackson, Narayanan, and Alwan, 2000), Heselwood and colleagues conducted a series of studies which used formant manipulation to find that the attenuation, or even removal, of the third formant of alveolar approximants, actually increases the perception of rhoticity (e.g. Heselwood, 2009; Heselwood, Plug, and Tickle, 2010; Heselwood and Plug, 2011). They found that it was not simply the lowering of F3 towards F2 that causes stronger rhoticity, but the creation of a single ‘strong perceptual peak’ of resonance around the region of these formants that strengthens audible rhoticity.

An additional aspect of ‘auditory’ retroflex approximants, is that they may be produced using either tip up (retroflex) or tip down (bunched) articulations. Zhou, Espy-Wilson, Tiede and Boyce (2007) conducted an MRI comparison of bunched and retroflex articulations, with the intention of finding whether they displayed unique acoustic signatures. While the frequencies of the first three formants were very similar between bunched and retroflex /r/, there was a distinct difference in the spectral gaps between the fourth and fifth formants of each of the two articulations they measured in their study. Zhou et al. (2007) found the difference between F4 and F5 in retroflex /r/ to be around 1400Hz, but only around 700Hz in bunched /r/: the authors claim that this difference is due to dissimilarities in vocal tract shapes because of the different tongue configurations. These configurations create cavities immediately behind the palatal constriction, which are of different size and shape in bunched /r/ and retroflex /r/.

In summary, the key acoustic cues for approximant /r/s are lowered F3, such that F3 and F2 being very close together increases the percept of rhoticity. F4 and F5 may give additional information about tongue configuration, but is not as important for an overall rhotic percept as the lower formants.

1.4.2 Social factors

This section provides the essential sociolinguistic background for the phonetic feature and experimental design adopted in this thesis.

Historically, Scotland has always retained a strong sense of Scottish identity, despite political fluctuations over the past few centuries (Braber and Butterfint 2008). Macaulay claims that a part of this identity comes from having a form of speech which remains distinct from that of England, as well as cultural attitudes that mirror the sense of linguistic separation (2005). The effects of this linguistic separation have been studied in locations surrounding the border between Scotland and England, and Watt and colleagues found that a sense of group membership arising from a national identity at either side of the border can cause some variation in a speaker's lexical and phonological performance, when speaking to either ‘group’ (e.g. Watt, Llamas and Johnson 2010). This sense of Scottish identity is evident in Glasgow, Scotland's largest conurbation, and the linguistic connection with this identity may show in an individual's speech in different ways according to their socioeconomic status. Braber and Butterfint write that what constitutes a local or geographical identity may be open to questioning, but being born and living in Glasgow may contribute to this identity in an individual in a major way (2008).

Glasgow has a complex linguistic environment, with Aitken (1979) noting the existence of a ‘bipolar continuum’ in Glaswegian speech. The notion of a continuum is a useful way of describing the coexistence of the two varieties of Scots and Scottish Standard English (SSE) in one geographical location, together with the style-shifting between both varieties which is commonly observed in many

Glaswegian speakers. Over the last few decades, strong network ties in working class communities in Glasgow may have become stronger, possibly in response to Glasgow's stagnant economy in the mid twentieth century having an impact on overall social mobility. This led to certain linguistic class markers becoming more sharply defined (Stuart-Smith, Timmins & Tweedie 2007; Chambers 2008: 52). Wells' (1982) seminal account of English accents reported Scottish English to be conservative with respect to maintaining rhoticity, particularly Glaswegian vernacular in terms of many Scots features. We shall see that several studies contemporary with the writing of Wells have uncovered increasing derhoticisation in working class Central Belt English, especially Glaswegian. According to Wells, many people may claim that ‘authentic Scots’ has died out, yet working class Glaswegian speech displays many features which could be considered characteristic of Scots rather than Standard English, including lexical, syntactic, morphological and phonological features (1982: 395). He also reports that because Scottish English is generally rhotic it is ‘strikingly conservative’ (1982: 407).

Many speakers in Glasgow have the capacity to shift their style along lexical, grammatical, and phonological continua, and these choices are made for various reasons, such as subject, interlocutor, and formality of setting. However it is Stuart-Smith's choice of the term ‘style-drifting’ (1999: 204), instead of style-shifting, which may best demonstrate the subtle, yet complex and fluid nature of the relationship between Scots and SSE, on an individual speaker level. In other words, it is not simply a matter of linguistic choice for a Glaswegian speaker to style-shift, but the choice to do so may very often be associated with identity and location. Macafee suggests that, for many Glaswegian Scots speakers, a sense of discomfort overrides the tendency to code-switch, such that a speaker would not feel comfortable speaking fully standard English, or feel able to maintain it for a long time (1983: 23). Additionally, she writes that even urban middle class speakers are likely to use local dialect forms in order to express a sense of place (1983: 24).

One example of the complexity of Glaswegian style-shifting is the variable nature of vowel choice for many speakers of Glaswegian Scots. Vowel quality can vary in almost any of the classic lexical sets, often with even more variation within these (see Stuart-Smith 1999; Wells 1982). For example, the vowel in the word ‘both’, which in RP is /əʊ/, is /o/ in SSE and /e/ in Glasgow Scots. In practice however, many Scots speakers alternate between /e/ and /o/, indicating style-drifting between Scots and SSE forms (Stuart-Smith 1999). Stuart-Smith (2003: 117) writes that alternation in the vowel systems of many Glaswegian Scots speakers may be a result of historical dialect mixture between Scots and English, but that the alternation between vowel quality is not regular for each set, with lexical frequency playing a key role in the realisation. Indeed, Macafee notes that code-switching in Glasgow is a matter of the frequency and type of dialect items, rather than drastic switches between fully standard and fully dialectal styles (1983: 24).

Despite possibly having a greater sense of Scottish identity, some consonantal features of working class Glaswegian speech are changing towards realisations found primarily in working class London speech. Stuart-Smith, Timmins and Tweedie (2007) suggested that the influence of the UK-wide broadcast media is helping to exert a change on speech in working class communities. The existence of strong network ties in such communities helps to spread these changes in a much faster way than linguistic changes happen in middle class communities. This partly explains why middle class Glaswegian speech is maintaining more traditionally Scottish features, despite having more opportunities for contact with English English speakers and weaker network ties (Stuart-Smith et al. 2007: 222).

While many Glaswegians have the capacity for style-shifting, some variants are so socially marked that a speaker will be unlikely to adopt it. For working class speakers, Stuart-Smith et al. (2007) report evidence of a negative ideology around the association between middle class speakers and conservative pronunciations. An interview transcript showed that, when shown a card with ‘loch’ written on it, the young working class informants produced it with [k] as the final segment. When asked, they confirmed that they knew it was ‘supposed to’ be pronounced with [x], but ridiculed those who did so, associating the pronunciation with residents of Bearsden, a middle class suburb of Glasgow (Stuart-Smith et al. 2007: 253). This suggests that working class speakers are readily adopting innovative forms such as [k] for [x], despite the fact that these forms are traditionally associated with English English speakers. In a similar way, we shall see later that a middle class speaker would be unlikely to produce a working class feature like derhoticised /r/ in normal speech, unless making a choice to do so for stylistic reasons, or making social comment specifically about the feature. Macaulay (1976) wrote that Glaswegians' speech varies along a continuum closely correlated with social differences, and that Glaswegians indicate their membership in a particular social class by the way that they speak. This seems to be explicitly borne out in the evidence gathered by Stuart-Smith et al. (2007).

It is possible that as well as the shift towards a loss of rhoticity in working class Glaswegian, the trend towards hyper-rhoticity will continue in middle class speakers in the city. An interesting situation may then result: two accents in the same city, one fully rhotic and one non-rhotic. We could imagine that this situation may continue for some time yet in Central Scotland, as Lass writes about the coexistence ‘for hundreds of years’ of rhotic and non-rhotic states in England, often in the same lects (1997: 287). In England in the eighteenth and nineteenth centuries, authors with prescriptive sensibilities wrote that only the ‘vulgar’ and the ‘lower classes’ would vocalise /r/ (Mugglestone 2003) (this stigmatisation of certain accent features still exists, for example ‘h-dropping’, and rhoticity in South West England), and it is possible that the kind of resistance was a contributor to the length of the process of /r/ loss. However, the influence of the internet and broadcast media in the present day could mean that changes in rhoticity will not progress as slowly in Scotland as they did in England.

Lawson et al. (2008: 103) and Stuart-Smith et al. (2014: 4) follow Romaine (1978) in writing that derhoticisation originated as a change from below the level of consciousness, as it is a subtle, system-internal change (see Labov 1994, 2001). An example of a change from above in postvocalic /r/ is reported by Dalmasso, who finds that a reduction in allophony towards approximant [ɹ] in Amsterdam Dutch is motivated by adoption of features from another dialect area of the Netherlands, and is therefore a system-external change (2012: 62). The fact that the increase in rhoticity in Glaswegian middle class speakers and the decrease in rhoticity in working class speakers are likely to be changes from below, underlines the fact that differences between the communities are very highly marked, with separate system-internal processes at play in each case. The gentrification of certain areas of Glasgow may be increasing the segregation between working class and middle class areas (e.g. Paton 2009), and stigmatising working class communities. This may help to increase the perception of differences between socioeconomic groups, and this may be partly seen in the increasing linguistic difference between these groups.

1.4.3 Rhoticity and derhoticisation in Scotland

Over the last two hundred years, there has been a gradual, yet persistent, loss of postvocalic /r/ in Standard Southern British English (SSBE), leading to its present-day status as a non-rhotic accent. As this change progressed non-rhoticity was socially stigmatized (e.g. Jones, 1989; Mugglestone, 2003; Lass, 1997), and was often accompanied by derogatory comments about the education of the speakers. It is therefore reasonable to assume that those leading the change were perceived as being lower on the social scale. This sociolinguistic change seems to have continued to spread across the British Isles, and in her study of working class schoolchildren in Edinburgh, Suzanne Romaine (1978) reported the beginnings of a loss of postvocalic /r/. This is despite Wells' claim that Scottish speech is overall ‘firmly rhotic’ (1982: 410). Romaine claimed that this change was unlikely to be linked to the standard or prestige model of a non-rhotic RP, unlike the pattern of slightly weakened rhoticity found before this time in some middle class speakers. In fact, Lawson et al. write that the loss of rhoticity in Scotland is clearly a change from below (2008: 103), so it is likely to have covert prestige, and in this way it may mirror the historical change in SSBE. Romaine's chapter presents possibly the earliest study of this phenomenon in Scottish English. A few years later, Macafee (1983) reported the loss of postvocalic /r/ in Glasgow, and Johnston (1997) noted the change across Scotland's central belt, which encompasses Edinburgh on the east coast, Glasgow on the west, and the towns and conurbations surrounding the two cities.

However it was not until relatively recently that 'derhoticisation' in working class Scottish speech was investigated in depth, with studies by Stuart-Smith, Scobbie, and Lawson, employing auditory, acoustic, and most recently articulatory methods including ultrasound tongue imaging (UTI) to fully explore the phenomenon (e.g. Stuart-Smith, 2003, 2007; Lawson, Stuart-Smith, and Scobbie, 2008, 2011b, 2013; Stuart-Smith, Lawson, and Scobbie, 2014; Jauriberry, Sock, Hamm, and Pukli, 2012; Bond, 2013). Elsewhere in Scotland, Brato (2012) describes an emergent loss of rhoticity in Aberdeen English, but the pattern of loss is socially stratified in a different way to derhoticisation in the central belt. Brato finds that the loss of rhoticity is primarily in young middle class speakers, and attributes this phenomenon to contact with non-rhotic speakers originally from England. In middle class Glaswegian speech, Lennon (2012) found that rhoticity is actually increasing over time, with average F3 lowering even further towards F2, resulting in more schwar-like realizations of approximant /r/. In this accent these articulations are likely to be bunched rather than retroflex, with constriction at the palato-velar region (Lawson, Scobbie, and Stuart-Smith, 2011b, 2013; Stuart-Smith et al., 2014).

In the literature on derhoticisation in the central belt, Stuart-Smith describes one token of *card*, produced by a Glaswegian informant, as having a 'pharyngealized/uvularized vowel' (2007), and points out the slight rise and weakened amplitude of the third formant in that token, with other papers making similar observations about raised or flat vowel-like F3 (e.g. Lawson, Stuart-Smith, Scobbie, Yaeger-Dror, and Maclagan, 2010, 2011a, 2011b, 2017). Furthermore, (UTI) articulatory studies show that Scottish derhoticised /r/ shows delayed tongue tip gesture with an early tongue root retraction gesture (Lawson et al., 2008, 2010, 2011a, 2011b; Stuart-Smith et al., 2014), and so there is often a degree of constriction at the pharyngeal/uvular place of articulation. Because of this, and the fact that uvular rhotics often display raised third formants (Lindau, 1985; Maddieson and Ladefoged, 1996), Scottish derhoticised /r/ appears to share some acoustic and articulatory properties with uvulars found in other languages.

Stuart-Smith (2007) drew parallels between derhoticisation in Glasgow and the patterning of r-deletion found in Plug and Ogden's (2003) investigation of Dutch postvocalic /r/. In their parametric analysis, Plug and Ogden found that postvocalic /r/ affects the whole rhyme, by changing both the quality and the length of the preceding vowel, and Stuart-Smith found similar effects in derhoticised postvocalic /r/ rhymes in Glaswegian. In another parallel with derhoticisation in Scottish English, Dutch is also showing signs of change over time in postvocalic /r/, in that younger speakers are producing more approximant /r/ variants than older speakers, who produce more uvular trills (Dalmasso, 2012).

If we revisit Lindau's (1985) family resemblance theory, we can begin to see how derhoticised /r/ in some working class varieties of Glaswegian (and other locations across the Scottish central belt) has developed a sociolinguistic allophonic relationship (polarization) with the auditorily r-ful (bunched, low F3) variants found in middle class speakers (e.g. Lawson, Stuart-Smith and Scobbie 2014). Derhoticised /r/ characteristically has a rising or high flat third formant (Stuart-Smith, 2007; Lawson et al., 2010, 2017), an acoustic feature very much like uvular /r/ (Lindau, 1985: 161, 165). Therefore, according to Lindau's model, we can parametrically link derhoticised /r/ with bunched /r/, as they share Lindau's parameter of sonorance (Lindau, 1985: 167). Because of the ability of many Glaswegians to shift their style of speech with relative ease along the 'bipolar continuum' between Scots and Standard Scottish English, depending on factors such as the social class of the speaker and the situation, it may be that some allophonic relationship exists between these /r/ variants, as in many varieties of Dutch (e.g. Scobbie, Sebregts, and Stuart-Smith, 2009). Lass writes about the coexistence 'for hundreds of years' of rhotic and non-rhotic states in Standard Southern English, often in the same lects (1997: 287), so it is not difficult to imagine that derhoticisation, if we assume that it will continue, could be a very long process in Scottish English.

1.5 Perception of rhoticity

We have seen how rhoticity can be defined in many different ways, and that even rhotic sounds that are auditorily very similar, can be produced with very different articulatory configurations (e.g. Zhou, Espy-Wilson, Tiede, and Boyce, 2007). However, an utterance is not simply produced by a speaker, but perceived by a listener. An integral part of understanding how rhoticity functions within a language or accent is to investigate how it is perceived, both by the speakers of that language or accent, and by listeners less familiar with the variety. This section will first look at how differences in phonetic detail can have a bearing on the listener's perception of rhotic sounds, then how different articulatory configurations can affect this perception, and will conclude by examining the role of a listener's experience with a particular accent in the perception of rhotics.

1.5.1 Perception of acoustic cues

Very often in the literature, a lowered third formant, towards the second formant, is said to be important for the perception of rhoticity. Both Ladefoged (1975) and Lindau (1978) suggested, using primarily English data, that a lowered third formant is an acoustic factor common to rhotic sounds (cited in Maddieson and Ladefoged, 1996). This appears to be the case for most approximant rhotics in English and other languages. However, a lowered F3 is not apparent in other types of rhotics; voiced and voiceless uvular rhotics in Swedish, French, and German all have a high F3 (e.g. Maddieson and Ladefoged, 1996: 244; Stuart-Smith et al., 2014: 11).

Heselwood (2009; with Plug 2011) tested the effect of certain acoustic properties of /r/ in a residually rhotic variety in Northern England, on the auditory judgements of strength of rhoticity by phonetically trained listeners. It was found that when F3 lowered towards F2, an auditory judgement of rhoticity was reported by the listeners. However, when F3 was attenuated or even removed from the same stimuli, using acoustic manipulation techniques such as low pass filtering, an increase in the perceptual strength of the /r/ was reported by the listeners. Because of this finding, Heselwood and Plug suggest that the widely held assumption that a low frequency F3 is a crucial acoustic and auditory correlate of rhoticity should be refined (2011: 870). Indeed, Heselwood et al. (2010) write that it is the location of formants in auditory space rather than acoustic space that best predicts perceptual judgements of strength of rhoticity. Furthermore, they suggest that the presence of a strong F3 close to the F2 frequency range may in fact inhibit the perception of rhoticity (Heselwood, 2009; with Plug 2011). This is because of a combination of effects: the closer F3 gets to F2, the stronger the perceptual peak is around the F2 frequency range and away from F4, and therefore the more complete the auditory integration becomes around F2 (Bladon, 1983; Chistovich and Lublinskaya, 1979; Hayward, 2014). If F3 is now removed entirely, this means that the strong perceptual ‘peak’ around F2 lowers even further away from F4, thus increasing the distance between perceptual peaks in the F2 and F4 ranges.

1.5.2 Listener experience and perception of rhoticity

In this discussion of the perception of rhoticity, we have been primarily concerned with the acoustics of rhotic sounds, the various articulatory means by which they are produced, and the perceptual and sociophonetic implications of these. However, an equally important question is whether a listener even identifies a sound as rhotic at all. A listener's identification of a variant of /r/ in a particular variety of English can be greatly influenced by how much experience they have of the variety. For example, if a listener whose accent is non-rhotic hears an unfamiliar production of a postvocalic /r/, it is possible that they can misidentify the word as something else. Of course, the same may apply if a rhotic listener hears a variant from a non-rhotic accent. Sumner and Samuel (2009) found that, when General American listeners with rhotic accents heard examples of a New York non-rhotic accent, they experienced a clear and consistent processing cost associated with their lack of experience with the accent. These listeners were much less accurate when processing words like ‘baker’ with a non-standard (in America) r-less pronunciation (bak[ə]), than they were at processing the standard r-ful pronunciation (bak[ɚ]). Listeners who were raised in the r-less accent area of New York found it almost as easy to process r-less tokens as r-ful tokens. This result is interesting, because it appears that the standard, dominant r-ful pronunciation facilitates processing for both groups of listeners, regardless of their linguistic experience.

A similar finding according to experience with an unfamiliar accent feature was made by Lennon (2013) who found that Southern English non-rhotic listeners in Cambridge were much less accurate than Glaswegian listeners when identifying minimal pairs of derhoticised tokens of working class Glaswegian speech as /r/, such as ‘hut’ and ‘hurt’, but they were almost as accurate as the Glaswegian listeners when identifying postvocalic /r/ in the standard (in Scottish English) middle class, hyper-rhotic productions of the same words (middle class /r/ in this study was presumably either bunched or retroflex articulation, but since no articulatory or acoustic measurements were made at the time of the study, this is currently unknown). Interestingly, the Glaswegian listeners also found it easier to identify the /r/ in the standard rhotic tokens than the non-standard derhoticised tokens, even though they had been raised in (broadly) the same accent area as the derhoticising speakers. This appears to mirror the results in Sumner and Samuel's (2009) study, in which even the native New Yorkers found it difficult to process the non-standard, non-rhotic, New York tokens. In Lennon's study, a combination of factors could have exerted an influence on the Cambridge listeners' accuracy to the rhotic stimuli, compared to the derhoticised stimuli. Since the rhotic variant is associated with the standard accent in Scotland (meaning more visibility in the media, etc.), the listeners may have had more experience of the rhotic tokens, so were more accurate in their responses. However, a much more likely cause for increased accuracy is the fact that the rhotic minimal pair stimuli were more acoustically distinct from each other in environments like ‘hut’ and ‘hurt’ than these pairs in the derhoticised variety. More work is required in this area in order to pick apart the various influences of social salience, acoustic and auditory distinction, and level of listener experience, when listeners hear Central Scottish /r/, and the experiments described in this thesis attempt to address part of this complex topic.

An interesting observation relates to the linguistic performance of the listeners themselves. Sumner and Samuel (2009) suggest that some New Yorkers, described in the study as ‘overt’ speakers of the non-rhotic dialect, store both the rhotic and non-rhotic variants in their memory as equivalent forms of the word, but other listeners – ‘covert’ speakers who are more variable in their use of the dialect – encode the non-rhotic form as a variant of an underlying r-ful form. They suggest that this implies an ability to map a wider set of inputs onto a single underlying representation. It is not possible to determine whether this pattern is present in Lennon's (2013) results, because the Glaswegian listeners were not split into groups of working class and middle class, roughly analogous in this experimental design to Sumner and Samuel's overt and covert speakers (but crucially not quite the same: there are many different kinds of social factors at play between the two cities). However, a major caveat when comparing these two studies lies with the fact that the variation of rhoticity in both social stratification and acoustic properties is very different between New York and Glasgow, and as Stuart-Smith et al. write, it is extremely difficult to separate the phonological from the social in the rhotic-derhotic continuum in Scottish English (2014: 30). Again, a more detailed study should provide a useful insight into how /r/ variants are stored in memory by different communities of speakers in Glasgow. The present work takes a step in this direction, by determining whether these variants are indeed distinguishable, and to what extent.

Overall then, both Sumner and Samuel's (2009) and Lennon's (2013) findings illustrate that listeners who have limited experience with a dialectally-associated form of postvocalic /r/ can have difficulty when processing tokens with the unfamiliar form, whether the distinction is the presence/absence of /r/ as in New York, or a hyper-rhotic/derhoticised continuum in Glasgow. However, some research has been done on how listeners perceive different forms of Scottish /r/, if they do have some experience of the Scottish linguistic environment. Carey (2010) conducted a perception test, in which both Southern English and Glaswegian listeners heard both Southern English and Glaswegian stimuli. Their task was to write down what they heard when presented with stimuli varying only according to presence of postvocalic /r/, such as ‘That's a prize for the child’ vs. ‘That surprise for the child’. Both groups of listeners had equal difficulty in recovering /r/. An unpublished undergraduate dissertation by Ashton (2011) found that listeners from Scotland's Central Belt would strongly associate bunched /r/ with middle class Edinburgh speech, and derhoticised and r-less realisations with both working class and Glaswegian speech. This result appears to be indicative of the strong social significance of different /r/ articulations, especially in Scotland. Another study which shows the importance of indexical cues in Scottish /r/ variants was by MacFarlane and Stuart-Smith (2012). Glaswegian listeners (mainly middle class) categorised approximant [ɹ] in postvocalic position as being indicative of middle class (‘Glasgow Uni’) speech, and tapped [ɾ] in the same position as indicating working class (‘General Glasgow’) speech. An unexpected result from this study was that onset /r/ did not follow a similar pattern of social judgements as postvocalic /r/, being categorised at chance level. The results of this study seem to suggest that there is more indexical information contained within postvocalic /r/ than in other positions.

1.5.3 Perception of /r/ in non-rhotic accents

The previous sections have primarily focused on the perception of /r/ in rhotic accents, but we will now look at listeners' perception of the phenomenon of linking and intrusive /r/ in non-rhotic accents. The perception of /r/ in accents which do not produce it in postvocalic position may help us to learn more about how the acoustics of the segment affect its perception in rhotic accents. In their perception study into listeners' sensitivity to intrusive /r/ in SSBE English (Standard Southern British English), Tuinman, Mitterer and Cutler (2011) found that listeners could successfully distinguish between utterance pairs containing sequences such as ‘saw (r) ice’, which contained intrusive /r/, and ‘saw rice’, where the /r/ was canonical (i.e. forming part of the second word). Native British English listeners (recruited from the University of Sussex in Brighton on the south coast of England) were heavily influenced by the duration of the /r/, where the intrusive /r/ was shorter than canonical /r/. Unlike the two non-native groups of listeners, American English and Dutch natives, the British listeners were not significantly influenced by orthography, and the two non-native groups were not as heavily influenced by the duration of the /r/ as were the British listeners. This result is interesting because canonical /r/, which is in onset position, appears to be stored by the speakers as a different ‘type’ of segment than intrusive /r/ in postvocalic position. This suggests either that these different types of /r/ have different durations in the mental representations of the speaker, or that they are produced with slightly different articulations. This possibly shows a parallel with the result in MacFarlane and Stuart-Smith (2012), above, who found that listeners attached different social judgements to tapped and alveolar /r/ in Glaswegian, depending on the word position. In another pair of studies, Hay and Maclagan (2010, 2012) conducted analyses of /r/-sandhi in New Zealand English in speakers born in the early 20th century, and found that the more often a speaker tended to produce linking /r/ and intrusive /r/, the lower the F3 tended to be. They suggested that this was because of a frequency effect: if a speaker produced more /r/s, then they are more likely to be of a more rhotic quality, i.e. with a higher perceptual peak at the F2 range (cf. Heselwood and colleagues, 2009, 2011). These studies, especially Tuinman et al. (2011), have potential implications for the perception of postvocalic /r/ in the present work, because they can tell us more about the perception of a similar segment by listeners with different levels of experience.

1.5.4 Perception of Scottish rhoticity

In this section, we will now examine some evidence about the effect that the articulatory configuration of different /r/ variants has on the perception of rhoticity, with a focus in particular on Scottish rhoticity. Delattre and Freeman (1968) describe many articulatorily distinct forms of the American English phoneme /ɹ/, but they report that in general it is said to be produced with either a bunched or a retroflex articulation, and that the auditory impression is the same for both articulations. Because of this, along with Lindau's claim that American speakers use all available articulatory methods [that they can] in order to produce a low F3 (1985), it seems reasonable to assume that retroflex and bunched /r/ are entirely auditorily equivalent. Indeed, Mielke, Baker and Archangeli (2006) report that variation between bunched and retroflex articulations in American English is speaker-specific, so we may assume that, at least in American English, an individual can arbitrarily choose either articulation. However, when testing perceptions of different /r/ allophones in American English, Twist, Baker, Mielke and Archangeli (2007) found that listeners could weakly distinguish between retroflex and bunched articulations. Furthermore, Zhou et al. (2007) found a difference in the acoustic relationship between the fourth and fifth formants between bunched and retroflex /r/ articulations.

However, Twist et al.'s (2007) result was less than conclusive, as there was a relatively high rate of non-classification (due to timeouts) and incorrect responses. They also expressed a potential concern with their methodology, in that the interpretation of null results could be related either to a negative result for the ease of distinction between the allophones, or to a problem with the power of the experiment. If the former were true, meaning a result of an overall difficulty in distinguishing between retroflex and bunched /r/, this may lead to the conclusion that the difference in perception between the two articulations may not be salient enough to have any social significance. This would be despite Zhou et al.'s acoustic distinction between the higher formants (2007). This could mean that a detailed acoustic analysis of a distinction between two sounds that are acoustically very similar, such as Zhou et al.'s analysis, may actually be perceptually indistinguishable in real terms. This may reveal a gap between the detail afforded by the acoustic analysis, and the auditory sociophonetic judgements continually made by real listeners. Therefore, this might mean that although there is articulatory variation, such variation may not be particularly important in sociophonetic terms. Indeed, Lawson et al. write that the above evidence suggests it is unlikely that the variants would become socially indexical (2011b).

Despite this, however, the recent work done by Lawson and her colleagues may suggest that there is indeed some social salience attached to the bunched articulation. Ultrasound Tongue Imaging (UTI) recordings of Scottish middle class speakers show that they are much more likely to produce /r/ with a bunched articulation, rather than retroflex (Lawson et al. 2011b; Stuart-Smith et al. 2014). The fact that middle class Scottish speakers have become increasingly rhotic, while working class speech is continuing to undergo derhoticisation (e.g. Lawson et al. 2011b, 2014; Stuart-Smith et al. 2014; Lennon 2012, 2013) underlines the social significance of a perceptually strong rhotic variant. Therefore, if Scottish middle class speakers are continuing to differentiate their form of rhoticity from working class varieties in a perceptually divergent pattern – albeit in a covert manner – perhaps a single (hyperrhotic) variant can do this more effectively in a linguistic system than an arbitrary, non-socially weighted, choice between two articulations. This may explain the prevalence of the bunched articulation, rather than both bunched and retroflex variants being used interchangeably in the same linguistic context by different speakers of the same dialect. Delattre and Freeman (1968: 64) write that bunched /r/ gives the strongest auditory impression out of all the variants they analysed, and they note that it offers the highest degree of contrast with their 'British weak /r/', an almost non-rhotic variant. Lawson and colleagues are also researching the articulatory adaptations made by speakers when mimicking different types of Central Scottish /r/. Their work is ongoing (e.g. Lawson et al. 2017), and it will be interesting to see whether bunched articulation is preferred over retroflex, in order to signal middle class speech. If bunched /r/

coexists alongside a weakly-rhotic variant, as it does in such a complex way in Glasgow, then it is likely to be a socially salient articulation.

When working class Glaswegian listeners were presented with acoustically manipulated (i.e. cross-spliced) recordings of working class Glaswegian tokens *had* and *hard* with derhoticised /r/, Bond (2013: 48-60) found that they attended to the quality of both the preceding vowel and the following plosive, in order to correctly identify the word. Interestingly, Bond notes that one of the participants in his perceptual test reported hearing all 14 of the manipulated stimuli as /r/-less, i.e. as 'had' rather than 'hard', even when the speaker intended to produce e.g. *hard* (2013: 53). One possible explanation Bond puts forward for this, is that the participant may have been attending to only the apical articulation (e.g. [d]) rather than any of the other acoustic cues in the stimuli (2013: 53). Perhaps more interestingly however, Bond hypothesises that 'this participant uses the presence [in the vocalic portion] of a rising F2 transition to the following coronal plosive as the main cue for identifying /r/-less words' (2013: 53). In fact, Bond's experimental stimuli were manipulated so that the F2 frequency of the vocalic portion in each stimulus was constant, meaning that all vocalic portions had the characteristics of the derhoticised /r/-ful tokens, as opposed to the /r/-less tokens, which often had a varying F2 frequency. This is significant for the present research, because it appears that in the absence of strong perceptual differences like a perceptually-salient spectral peak at F2/F3 that is found in middle class Glaswegian speech, listeners may use more subtle phonetic detail to make their decisions about the identity of ambiguous words.

In terms of the social salience of rhoticity, Watt, Llamas and Johnson (2010) found in their study of speech accommodation in Scottish/English border towns that the rhoticity of the Scottish interviewer was unaffected by the fact that the interviewee was non-rhotic, and vice-versa, even though many other segments in these varieties were affected. They wrote that rhoticity is both stable within the two speech communities and different between the two speech communities (2010: 285), so the feature will not be accommodated to. This may serve as evidence that rhoticity has a particularly strong social salience (Trudgill 1986).

Heselwood and Plug's (Heselwood, 2009, 2011) findings about the importance of a strong percept of rhoticity caused by a spectral peak around the F2/F3 region, could raise questions about the importance of the higher formants in the perception of rhoticity. For example, let us assume that the contrasting relationships between the fourth and fifth formants is important in listeners distinguishing between the different articulatory configurations of bunched and retroflex /r/ (e.g. Zhou et al., 2007 – see section on 'rhoticity: production', as well as the section below, for discussion of this paper). It may then seem that distinctions between sounds because of acoustic differences, arising from differences in articulation, can translate to equally salient distinctions in auditory terms. However, this analysis does not take into account the fact that auditory or perceptual judgements cannot be directly correlated with acoustic definitions as measured on a computer: humans perceive sounds in a different way. It is therefore likely that, given that a strong perceptually-salient spectral peak around the second formant results in a strong percept of rhoticity in alveolar /r/ variants (e.g. Heselwood, 2009), the stronger this peak becomes (i.e. the closer the formants get), the less important for the listener the higher formants then become in the distinction between articulatory variants. It is possible that articulatory variants such as bunched and retroflex then become more difficult to distinguish, because of the influence of the spectral peak around the second and third formants.

A problem that may arise in acoustic analyses of formants in some forms of rhoticity is the intensity of the formant traces in spectrograms, especially if the recordings are not of a perfect studio standard. Also, F3 may show variable patterning, making classification difficult. Because of issues like these, it may be difficult to determine which acoustic cues listeners are attending to when they perceive rhoticity. An example of this is found in the literature on Scottish English. In the derhoticised form of /r/ produced by some Glaswegian speakers, the acoustic quality can be hard to classify, or even to measure. Stuart-Smith and colleagues (e.g. 2014) found that derhoticised /r/ showed, variously, either rising or falling F3, which may depend upon the articulatory configuration of the front cavity in each case, and often they found that a drop in spectral intensity made the measurements hard to make. In such cases, perceptual cues may therefore be difficult to investigate without a fuller understanding of the articulatory variation which is present in Glaswegian /r/.

This section has explored a number of studies that have tested, in different ways, how rhoticity is perceived by different listener groups. This is vital for the current investigation, as we now have a greater understanding of the factors that may lead to misperception, and the role that increased experience with an accent plays in the perception of a phonetic feature such as rhoticity.

1.6 Summary and research questions

This chapter has laid out some key theoretical background for this thesis, specifically an overview of the different theoretical approaches to speech perception. Next we looked at a number of experimental studies which have informed these theoretical positions. Then we focused on the key feature, the production and perception of rhoticity, in general and in Scottish English. This provides the core motivation for the central three research questions presented here, and tackled with the evidence of the three experiments.

The first research question for this thesis is:
'What is the role of experience in the perception of fine phonetic detail for a contrast?'

We have seen that rhoticity is changing over time, and that it can be perceived in different ways according to articulatory configurations. We have also seen that listeners may be able to differentiate between a more specific range of variants if they have more experience with a particular dialect. On a sociolinguistic level, rhoticity is changing in Glasgow and central Scotland, with a diverging pattern between the working class and middle class speech communities in the region. While middle class speakers are increasing their rhoticity with more bunched articulations, working class speech is undergoing a process of derhoticisation, possibly because of the influence of an England-based UK national media. Derhoticisation is perceived differently depending on the listener's experience of the Glaswegian accent, and this perception will now be assessed further. We already know from the Masters study (Lennon 2013) that less familiar listeners are less accurate, and slower, when distinguishing similar minimal pairs which differ by derhoticised /r/, as produced by working class Glaswegians. We will be able to provide a much more nuanced answer to Research Question 1 with a more sophisticated analysis of the data from the Masters – Experiment 1 – using signal detection analysis.

This leads to the second research question, which is:
'How does experience relate to the learning of ambiguous fine phonetic detail for a contrast?'

Experiment 2 follows on from the findings of Experiment 1, by testing participants' responses to the derhoticised /r/, both before and after hearing a short sample of working class Glaswegian. This is done by presenting listeners with acoustically manipulated recordings of real speech, in order to measure the effect on perception of the relationship between F2 and F3. The participants are all native English speakers from different locations in the UK, representing different levels of familiarity with Glaswegian.

The final two research questions are:
'How do experienced listeners process ambiguous fine phonetic detail for a contrast?'

and:
'Do harder listening conditions affect the online perception of ambiguous fine pho-

netic detail for a contrast?'

These will both be addressed by Experiment 3, which uses a novel analysis method to measure the degree of attraction to similar words in minimal pairs. The participants are native Glaswegian listeners, who have the most experience with the accent under investigation. This final experiment will tell us about what happens *as listeners hear* ambiguous words, in other words, online processing. It will also inform the wider theoretical literature by examining the effects of challenging listening conditions, by presenting listeners with stimuli from more than one talker.

Chapter 2 of this thesis provides key information about acoustic cues for the perceptual experiments by describing a detailed acoustic analysis of the word types under investigation. Chapter 3 begins with an overview of signal detection theory, then describes the methodology and results of the detailed analysis of Experiment 1. Chapter 4 describes the methodology and results of Experiment 2, and Chapter 5 does the same for Experiment 3. Finally, Chapter 6 provides a summary and general discussion of all the results in the previous chapters, making conclusions about the implications of the observed patterns of data.

Chapter 2

Acoustic analysis

2.1 Introduction

This chapter presents an acoustic analysis of the stimuli used in my unpublished Masters dissertation, which will henceforth be referred to as Lennon (2013). This analysis was carried out in order to interpret the results of Lennon (2013, re-analysed in this thesis as Experiment 1 in Chapter 3), and in order to feed into the design of Experiments 2 and 3 (Chapters 4 and 5). It provides new information, based on a small but carefully controlled speech sample, about the acoustic manifestation of rhoticity in both working-class and middle-class Glaswegian speech.

There is substantial recent research into the articulation of /r/ in Glasgow. Using Ultrasound Tongue Imaging (UTI), Lawson and colleagues (e.g. Lawson, Scobbie & Stuart-Smith 2011b), found that derhoticising working class speakers display a retracted tongue root configuration (causing a degree of pharyngealisation) in combination with a post-voicing tip-up gesture, leading to a vowel-like quality. They also found that hyper-rhotic middle class Scottish speakers use a bunched tongue configuration similar to the American English shape described by Delattre and Freeman (1968).

Acoustic work on Scottish /r/, however, has been more limited. When analysing working class /r/ in Glasgow, Stuart-Smith (2007) found that in /Car/, /CarC/ and /CaC/ words (e.g. *car*, *heart*, *cat*), those with /r/ tended to have a longer rime, and had lower F2 and higher F3 throughout the rime, possibly reflecting uvularization. There have been no detailed acoustic studies of Scottish middle class rhoticity (though it was briefly discussed in Lawson, Scobbie & Stuart-Smith 2014); however there have been acoustic analyses of hyper-rhoticity in other varieties. The proximity of F3 to F2 in approximant /r/ variants is noted by some authors to be important for a strong percept of rhoticity (e.g. Ladefoged, 2003; Lindau, 1985). Heselwood, Plug and colleagues have observed that the most important feature for rhoticity is a strong perceptual peak around the F2 region, whether that arises through a combination of F2+F3 or, as they found in experiments using low-pass filtered speech, absence of F3 entirely (Heselwood, 2009; Heselwood and Plug, 2011).

The following analysis explores the acoustic underpinnings of the perceptual results of Lennon (2013) by examining the formant trajectories in the stimuli used in the experiment. The aim of this analysis is to examine the acoustic contrasts between V and Vr words (e.g. *hut/hurt*) for middle class and working class varieties in Glasgow, in order to better understand the basis for listeners' capabilities and difficulties in distinguishing between these minimal pairs.

2.2 Method

2.2.1 Recordings

The recordings analysed here were made for Lennon (2013). Two pairs of native Glaswegian males (2xMC, 2xWC; 22-25 years) were recorded in a sound-attenuated booth, using lightweight Beyerdynamic TG H74c Condenser headset microphones, at a sampling rate of 44.1kHz. Each pair of speakers was recorded separately. Each pair took part in a collaborative word-finding task: this meant that speech was as naturalistic as possible, while still ensuring that the desired set of target words was produced by each speaker. In the task, the speakers produced connected speech. The target word tokens were excised from this; the excision did not lead to any noticeable artefacts.

/i/	/ʌ/
bead/beard	*bud/bird*
feed/feared	*hut/hurt*
weed/weird	*thud/third*

Table 2.1: Minimal pairs used by Lennon (2013)

The words analysed here are 6 sets of minimal pairs, listed in Table 2.1. Each word was produced between 1 and 3 times by each of the four speakers (average 2.35 repetitions per speaker), totalling 113 tokens. The stimuli were representative of both middle class and working class Glaswegian speech, because the speakers, who were from Bearsden and Maryhill (a middle class area and a working class area, respectively, of Glasgow), spoke with the accents typical of those areas.

2.2.2 Segmentation & formant analysis

Following Stuart-Smith (2007) and Plug & Ogden (2003), the acoustic analysis focused on the entire periodic portion of each word. For completeness, formants F1 through F5 were analysed.

Segmentation was carried out in Praat (2006). Figure 2.1 gives an example of a segmented token. For each word, four acoustic segments were labelled: |o| (onset), |v| (vocalic), |s| (silence, i.e. period with low/no energy), and |c| (coda). Boundaries were placed at zero crossings, always at positive-going deflections from zero, and were later refined by running a Praat script (2006).

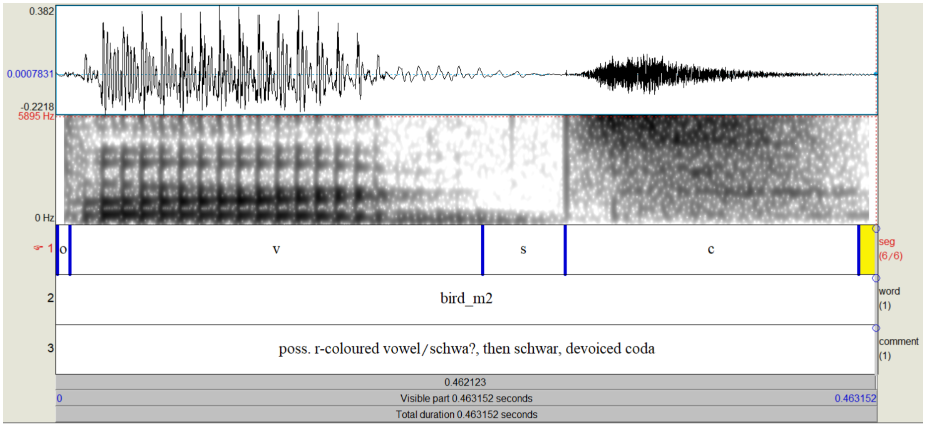

0.382
0.0007831
-0.2218
5895 Hz
0 Hz
1
o
v
s
c
seg
(6/6)
2
bird_m2
word
(1)
3
poss. r-coloured vowel/schwa?, then schwar, devoiced coda
comment
(1)
0.462123
0
Visible part 0.463152 seconds
0.463152
Total duration 0.463152 seconds

Figure 2.1: Waveform, spectrogram and segmentation for *bird*, spoken by middle-class speaker 2. See text for details.

The start of |o| (onset consonant) was set as follows:

- voiceless fricatives: start of aperiodic noise
- voiced stops: start of burst
- /w/: start of voicing bar

The boundary between |o| (onset) and |v| (vocalic) was set as follows:

- after voiceless fricatives: at the start of periodicity
- after voiced stops: at the start of the period where amplitude of the periodic portion was at or close to maximum (Figure 2.2);
- after /w/: at the point at which the formants had finished changing frequency due to the /w/

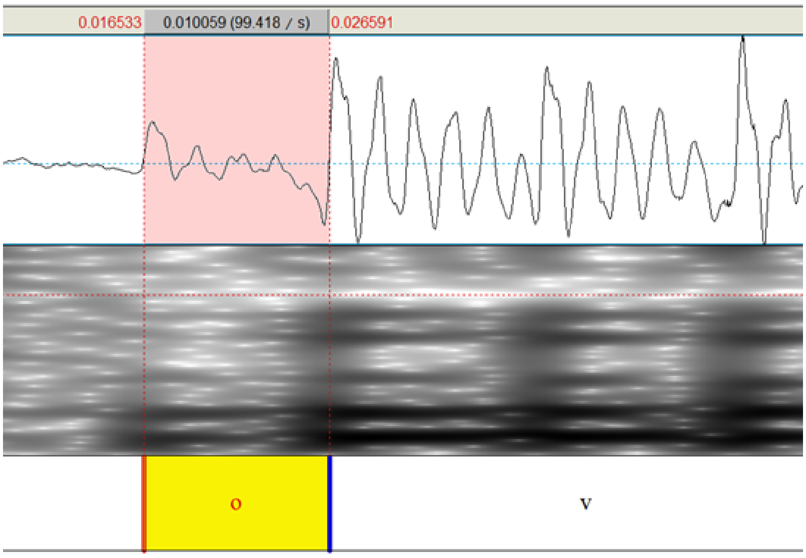


Figure 2.2: Location of |o|-|v| boundary for the token of *bird* shown in Figure 2.1

The boundary between |v| (vocalic) and |s| (silence: low/no energy) was set at the point where the formant energy appeared to finish on the spectrogram (window length: 0.007s, dynamic range: 70dB, view range: 6kHz). Where F1 appeared to continue right through into the |c| (coda) segment, even when there was little evidence of periodicity in the waveform, the end of the last remaining higher formant (normally F2) was used to judge the point at which the |s| segment began (Figure 2.3).

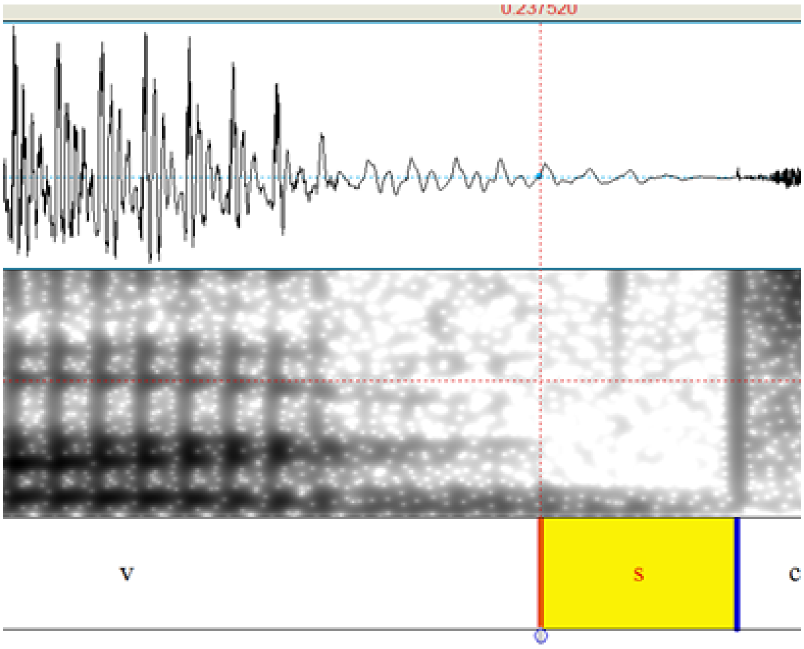


Figure 2.3: Location of |v|-|s| boundary for the token of bird shown in Figure 2.1

The boundary between |s| (silence: low/no energy) and |c| (coda consonant) was placed at the start of the burst (Figure 2.4). Finally, the end of the |c| (coda) segment was marked at the point where the energy ended. The durations of the four intervals were extracted via a Praat script.

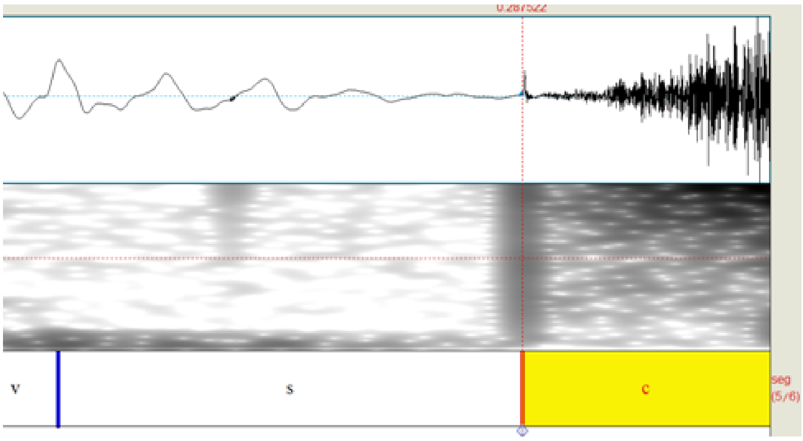


Figure 2.4: Location of |s|-|c| boundary for the token of bird shown in Figure 2.1

Formant tracks for F1-F5 were then extracted and manually corrected using the Python-based program Formant Editor (Sóskuthy 2014). Formant Editor first generates formant tracks in Praat, then a GUI (graphical user interface) allows the user to manually adjust the formant tracks (by clicking and dragging) to follow the path that is judged to be closest to the 'real' position of the formant. An example of uncorrected and corrected tracks is shown in Figure 2.5 (left) and Figure 2.6 (right).

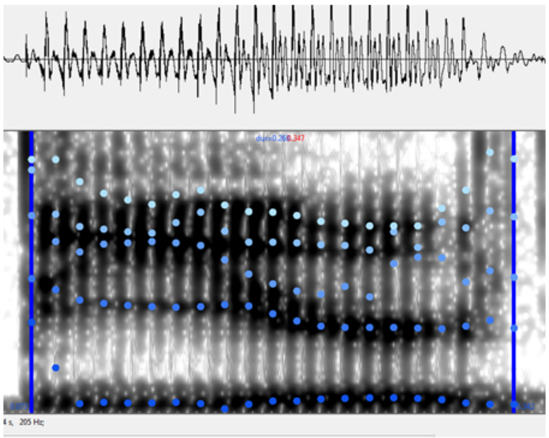


Figure 2.5: Token of *beard* before formant correction

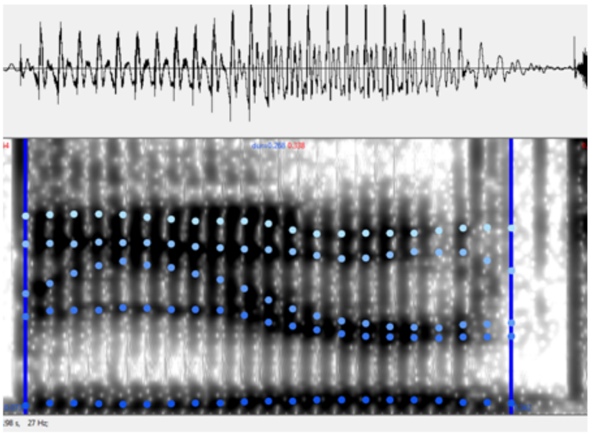
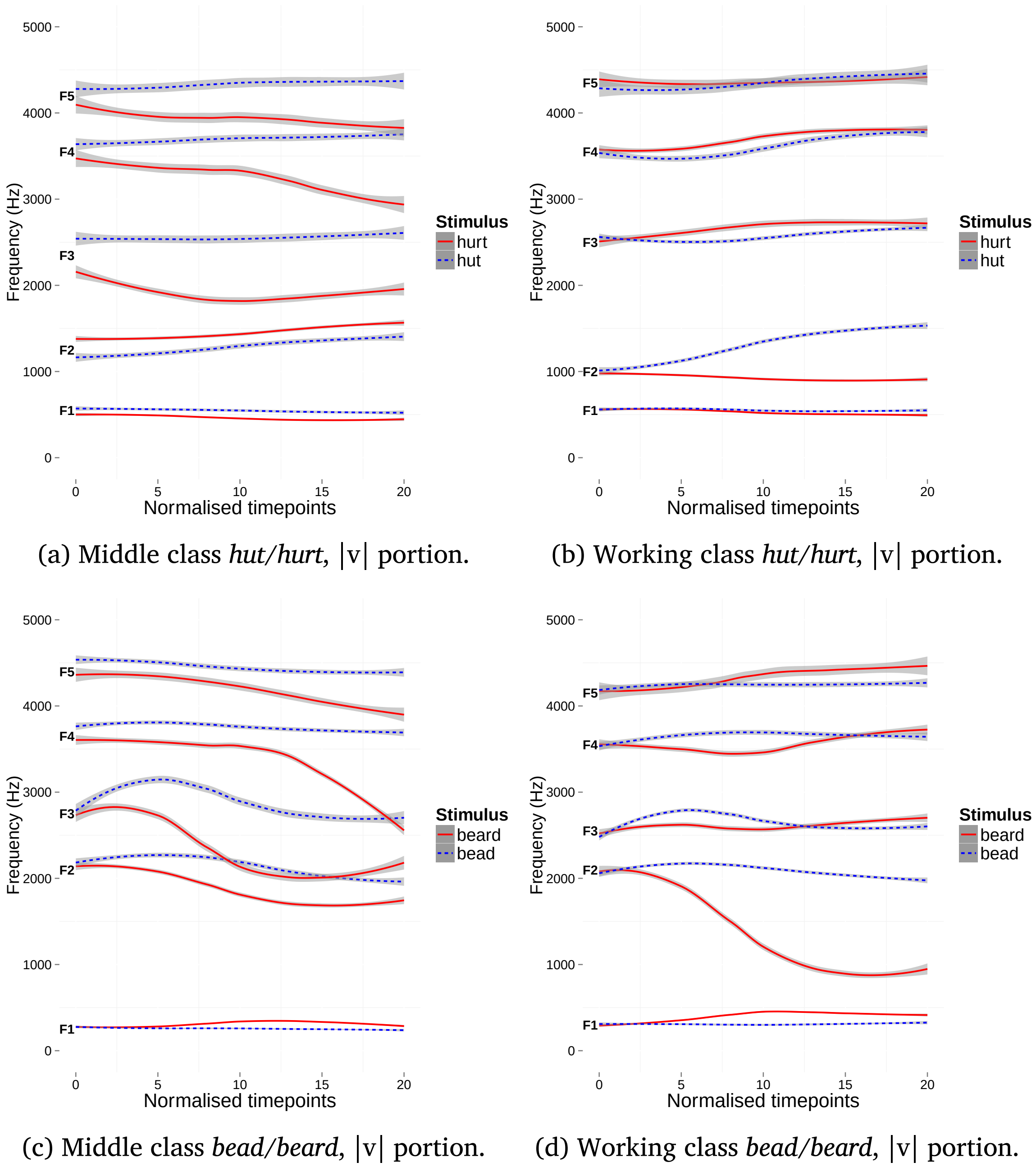


Figure 2.6: Token of *beard* after formant correction

Once all formant correction and tagging was completed, the information was output as a csv file, which was then used as the input to the statistical program R (R Development Core Team, 2013), for analysis and visualization.

2.3 Results

The results of the formant tracks for all five formants are shown in Figure 2.7. The tracks are shown for the |v| portion of each word. They have been time-normalised so that the shape of trajectories can be compared, ignoring differences in duration.



(a) Middle class *hut/hurt*, |v| portion.

(b) Working class *hut/hurt*, |v| portion.

(c) Middle class *bead/beard*, |v| portion.

(d) Working class *bead/beard*, |v| portion.

Figure 2.7: Formant tracks F1-F5 for all stimuli, averaged across Class and Vowel. Left panels: MC speakers; Right panels: WC speakers. Top panels: words with /ʌ(r)/; bottom panels: words with /i(r)/. /r/ stimuli are represented by solid lines, /r/-less by dotted lines. E.g. *'hut'* represents all *bud, hut, thud* stimuli. Plotted in R using ggplot2's stat_smooth function to draw formant tracks. Shaded ribbons: 95% C.I.

All five formants show clear differences relating to coda structure, vowel quality, and social class. These were confirmed by Linear Mixed Effects regression modelling using lme4 in R. Five sets of models were run, one for each formant. The dependent variable was the frequency of the formant. Variation across the time course of the track was assessed by modelling measurement_number (defined as normalised timepoint number, from 0 to 20) as a fixed factor, with interactions for the factors of interest, which were Social class, Coda structure, Vowel quality, and Duration. Thus for example a significant interaction of Social class with Measurement number for F2 would indicate that the evolution of the F2 trajectory over time differed significantly between working class and middle class speakers. This approach follows that of Stuart-Smith (2016).

Initial modelling included main effects for Measurement number, Social class, Coda structure, Vowel quality, and Duration, and all three-way interactions involving measurement number. Four-way interactions could not be included because of sample size. Model comparison and pruning was carried out using the R package lmerTest's step() function, which carries out backwards stepwise regression to prune non-significant predictors. The function aims to reduce the number of terms in the model by first performing automatic backward elimination of random-effects terms, followed by backward elimination of fixed-effects terms (Kuznetsova et al. 2017). For each random-effects term a reduced model is built without it, then the resulting model is compared with the full model by performing a likelihood ratio test. If the largest p value out of all the models is higher than the alpha level then that random effect is removed from the model. Backward elimination of fixed-effects terms takes place once all random-effects terms have been tested, and the procedure is similar to that for the random-effects terms. Starting with the model's highest-order interaction effects, the effect with the highest p value is removed. If the excluded term's p value is greater than the alpha level, the term is removed, then testing continues in a stepwise manner with lower-order interactions, retaining all terms where the p value is less than the alpha level.

After running step(), the optimal model for all five formants contained significant interactions ($p < 0.01$) for measurement_number*class*vowel, and for F2-F5 also for measurement_number*class*coda ($p < 0.001$). The statistical results are not presented further here, but observations of differences in the descriptive tracks in the sections below are also those which were found to be statistically significant in planned comparisons ($p < 0.05$).

2.3.1 Coda structure

The factor 'coda structure' relates to the presence or absence of /r/. For middle class speakers, the presence or absence of /r/ is signalled by a difference in all formants. In contrast, for working-class speakers, words with and without /r/ differ in only some formants. The most extreme case is shown in Figure 2.7b, where the only significant difference between working class stimuli with and without /r/ is rising F2 in *hut*.

2.3.2 Vowel quality

The main difference between stimuli with a back vowel (e.g. *hut/hurt*) and those with a front vowel (e.g. *bead/beard*) is the starting frequency of F2 in all graphs, which is much higher before a front vowel, as expected (compare Figures 2.7a & 2.7b with Figures 2.7c & 2.7d). Then, in *beard* stimuli F2 drops for the /r/, especially for working class speakers, accompanied by a slightly rising F3, whereas for middle class speakers, F3 drops for the /r/ (but with a more complex trajectory than the change in the working class F3). In all /i/ stimuli, the contrast between /r/ stimuli and /r/-less stimuli is clear. For middle class speakers, F3 is lower for /ʌ/ than /i/. For working class speakers, the /ʌ/-/i/ difference is not found, and F3 is also higher overall than for middle class speakers. The measurement_number*class*vowel interaction for F4 shows that F4 is slightly lower for WC /ʌ/ than MC across the vocalic portion. The same interaction for F5 shows that F5 is not as low for working class /ʌ/ words as MC, across the vocalic portion. Finally, the class*vowel interaction for F5 shows that F5 rises in working class speakers for /ʌ/ tokens. It must be noted that the overall formant pattern of /i/ stimuli is such that more differences are maintained *overall* throughout the vocalic portion, than in /ʌ/ stimuli, which has the potential to influence perception.

2.3.3 Speaker class

The most striking pattern is that F2 and F3 become very close in all middle class /r/ stimuli towards the end of the vocalic portion (solid lines in Figures 2.7a & 2.7c), clearly showing their hyper-rhoticity. Conversely, the equivalent F2 and F3 tracks are much further apart for the working class speakers (solid lines in Figures 2.7b & 2.7d), showing derhoticisation (similar to the large F2-F3 difference in Stuart-Smith, 2007). Overall, the differences in these stimuli are further highlighted by the interactions described above.

2.3.4 Proximity of pairs of formants

Previous research has suggested that the difference in frequency between certain pairs of formants is what underlies the percept of rhoticity. Many authors observe that the proximity of F2 and F3 creates a strong percept of rhoticity (e.g. Heselwood 2009). Zhou et al. (2007, 2008) further suggest that in bunched /r/ in American English, F4 and F5 are close together (about 700 Hz apart, compared to 1400 Hz for retroflex /r/).

These patterns are mirrored in the middle class tokens, which have, around the mid-region of the vocalic portion, F2-F3 differences around 350 Hz on average, and F4-F5 differences around 700 Hz (Table 2.2). The working class tokens show a different pattern, with much more equally-spaced formants. There are large F2-F3 differences (over 1500 Hz), and F4-F5 differences for working class stimuli are closer to the range proposed for bunched /r/ by Zhou et al.. However, articulatory research has not shown evidence of bunching for working class /r/ (e.g. Lawson et al. 2014), so this acoustic pattern presumably has a different cause.

	MC /r/		WC /r/	
	beard	*hurt*	*beard*	*hurt*
F2	1728	1475	1009	901
F3	2044	1845	2604	2725
F2-F3 diff.	316	370	1595	1824
F4	3435	3235	3561	3773
F5	4139	3929	4480	4433
F4-F5 diff.	704	694	919	660

Table 2.2: Average higher formant values for all /r/ stimuli (unit: Hz), taken from normalised timepoints 10-15.

2.3.5 Duration of vocalic portions

Table 2.3 shows average durations of the segmented |v| (vocalic) portions of the stimuli. Like the rime durations described by Stuart-Smith (2007), the vocalic durations in working class /r/ stimuli are longer than in their /r/-less counterparts, especially for *bead* & *beard* (208 vs 308 ms) and to a smaller extent for *hut* & *hurt* (238 vs 273 ms). The middle class speakers also show longer durations in words with /r/ than those without /r/, in addition to the strong spectral cues discussed above. The statistical modelling for formants also found a significant interaction of measurement_no*class*duration for all formants except for F5, showing that middle class and working class speakers also show different formant trajectories according to the duration of the vocalic portion. This is presumably because they are using different articulatory gestures (e.g. Lawson et al. 2011b; 2014) with temporal patterns.

	Middle class	Working class
bead	174	208
hut	175	238
beard	253	308
hurt	216	273

Table 2.3: Average duration of vocalic portion V(r) for all tokens, by type (ms).

Results summary

The main finding is that by far the most acoustically similar word types are minimal pairs *bud/bird*, *hut/hurt* and *thud/third*, produced by working class speakers. That is, these pairs show the fewest significant differences between the formant trajectories of words with and without /r/, even though duration and F2 trajectory do distinguish these words to some extent. In contrast, middle class speakers show much more extensive differences between /r/-ful and /r/-less words. They are acoustically hyper-rhotic, primarily because of the proximity of F2 and F3 in /r/ words, and they also show higher formant characteristics similar to bunched /r/.

2.4 Discussion

This acoustic analysis examined in detail the formant characteristics of both working class and middle class postvocalic /r/ variants in Glasgow, providing a detailed description of the acoustic contrasts between V and Vr words. The finding that the most acoustically similar word types are the minimal pairs in the /ʌ/ vowel environment produced by working class speakers, supports previous work on derhoticisation in Glasgow, e.g. Stuart-Smith (2007). The potential for misperception in these pairs is therefore likely to be very high. The acoustically hyper-rhotic middle class speakers are much less likely to encounter misperception when producing the same minimal pairs, as they are acoustically much more distinct. The secondary finding that the middle class speakers' higher formant characteristics resemble those of bunched /r/ (e.g. Zhou et al. 2007; 2008) is interesting, but without articulatory analysis this cannot be taken as evidence for tongue configuration.

Though admittedly small-scale, this is the first systematic acoustic analysis which considers the impact of presence/absence of /r/, vowel quality and social class, on Scottish English formant trajectories. Furthermore, it contributes to our understanding of the acoustics of weak /r/ variants in Scotland and beyond.

This acoustic analysis is important support for the primary research on the perception of the stimuli as conducted in Lennon (2013) and refined in Chapter 3. This analysis allowed the subsequent design of Experiments 2 and 3 to proceed armed with detailed information about the acoustic features that vary over the timecourse of the words under investigation. Experiment 1, described in the next chapter, benefits from this information, and expands on the knowledge gained here.

Part II

Experiments

Chapter 3

Experiment 1: Long-term familiarity

3.1 Introduction

It is evident from the acoustic analysis in the previous chapter that some working class word contrasts are extremely acoustically similar, meaning that the nature of derhoticised /r/ variants can potentially lead to misperception, for example in the minimal pair *hut/hurt*. In contrast, because middle class speakers are displaying an increase in rhoticity (as seen in Lennon 2012, and confirmed in the previous chapter), this misperception would not be expected. It was the potential for misperception in the working class derhoticised words that motivated the Masters research topic, and this would become the pilot for the present work. The experiment conducted in the Masters study created a rich resource for different analysis methods, which could not be explored in the confines of a short Masters dissertation. A comprehensive analysis of the data from that experiment was essential to enable the development of the present thesis, as the appropriate direction of the investigation could only be properly decided upon once the results of Experiment 1 were fully understood. The analyses reported in this chapter are new – conceived and conducted during the doctoral research period, not the Masters – justifying their inclusion in this thesis. Because there is so much extra analysis of the Masters experiment in the thesis, it will be termed 'Experiment 1' from now on.

The purpose of the present chapter is a fine-tuning of the results of Lennon (2013), allowing a much clearer picture to develop from the data which was collected. With this in mind, the research question for this chapter can be stated as:

'What is the role of experience in the perception of fine phonetic detail for a contrast?'

This chapter briefly summarises the method and findings of Lennon (2013), then introduces the new analysis method which was chosen. It then goes on to describe and discuss the results of this analysis.

3.1.1 Summary of Lennon (2013)

Before the analyses can be described it is necessary to give a brief summary of the procedure and initial outcomes of Experiment 1, as far as they were reported in Lennon (2013). A perceptual experiment investigated whether a listener's amount of exposure to Glaswegian accents affects their ability to identify the word the speaker intended to produce, given a choice from words in minimal pairs with and without postvocalic /r/. In the experiment there were two tasks, a two-alternative-forced-choice (2AFC) design and a strength rating task, and listeners had different levels of long-term experience with the Glaswegian accent.

Participants: Listener groups were designed so that participants were from three accent groups, with varying levels of experience of the Glaswegian accent. There were 62 subjects in three groups:

1. Glasgow: Raised in Glasgow, living in Glasgow (n = 21);
2. Intermediate: Raised in England, living in Glasgow, attending Glasgow University (mean residence in Glasgow = 3.6 years, n = 21);
3. Cambridge: Raised in South East England, attending Cambridge University (little/no experience of Glaswegian, n = 20).

Materials: Materials were the six word pairs shown in Table 2.1, recorded from the Glaswegian speakers described in Chapter 2. There were 144 word tokens in total (12 words x 4 speakers x 3 repetitions).

Tasks and procedure: There were two tasks, presented individually to participants using the perceptual testing software DMDX (Forster & Forster 2003). The first was a 2AFC task, in which participants were asked to choose which word they thought they heard, out of minimal pairs such as hut/hurt. In the second task, they were asked to rate the 'strength' of the /r/ sound in each stimulus, on a scale from 1 to 5.

These tasks yielded four sets of data: accuracy scores and response times from the 2AFC task, and ratings and response times from the /r/-strength rating task. All were analysed with linear mixed effects modelling in the statistical analysis program R.

Results from the 2AFC task showed that the presence of the working class derhoticised /r/ variant caused perceptual ambiguity for all listener groups, while the middle class variants (often realised as a schwar [ɚ]) elicited much more accurate and faster responses. There were large differences between listener groups: Glaswegian listeners were the most accurate and the fastest, while English listeners in Cambridge were by far the least accurate and the slowest. The English listeners in Glasgow displayed an intermediate, yet complex pattern, which showed an intriguing effect of experience with derhoticised /r/, which is discussed further in 3.2 below.

In the strength rating task, the middle class stimuli elicited a strong /r/ rating from all listener groups, following predictions. However, there was more variation in responses to the derhoticised /r/ stimuli, with listeners in Cambridge giving much weaker ratings for the derhoticised /r/ tokens than listeners in both groups resident in Glasgow. These results show that the native Glaswegian listeners were the most accurate at recognising the derhoticised variant as /r/, with the English listeners in Glasgow showing a pattern which is almost as accurate, while the Cambridge listeners found it harder to identify derhoticised tokens as /r/-ful. Taken together, the initial results from Lennon (2013) support the hypothesis that long-term familiarity with an accent's fine-grained phonetic detail aids comprehension.

These results add to the understanding of a significant on-going change in Scottish speech, while contributing to the complex question of what defines rhoticity, from the perspective of both production and perception. However, the experiment did produce complex results, which required further analysis in order to be understood fully. The rest of this chapter describes the new analysis that was completed during the doctoral research period for the data collected in Experiment 1. The method of investigation, signal detection analysis, is described in detail in the following section.

3.2 Signal detection analysis: overview

The first new analysis conducted on the data from Experiment 1 was a full signal detection analysis. This was done because the pattern of responses between the listener groups was complex, with a ‘crossover’ pattern found in both accuracy and response time. The nature of this ‘crossover’ can be seen in Figures 3.1 and 3.2, which are taken from Lennon (2013), and show the accuracy and response time, respectively.

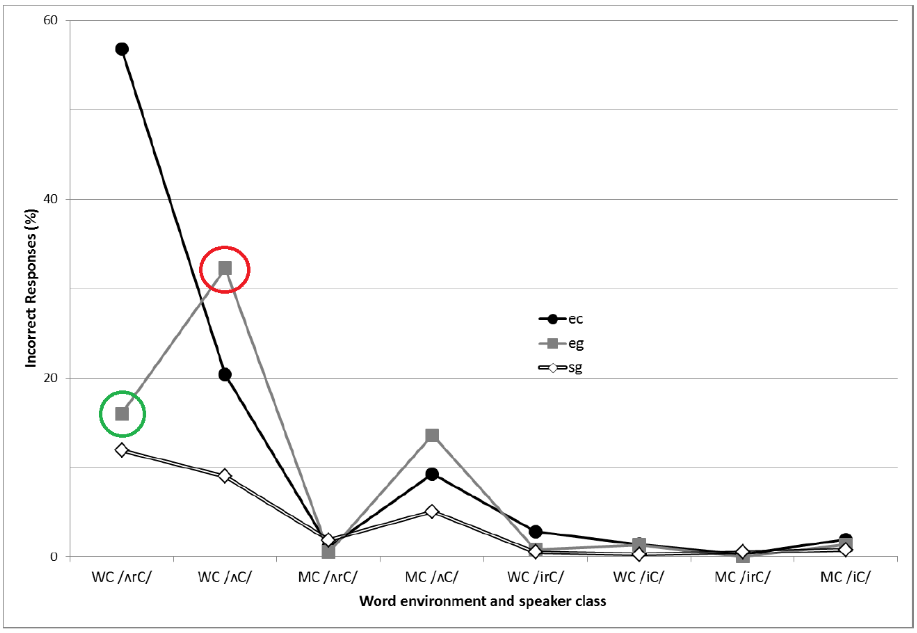


Figure 3.1: Incorrect responses by Listener group (ec: Cambridge, eg: Intermediate, sg: Glasgow), Vowel/Coda, and Class (Lennon 2013: 16). See text for explanation of green and red circles.

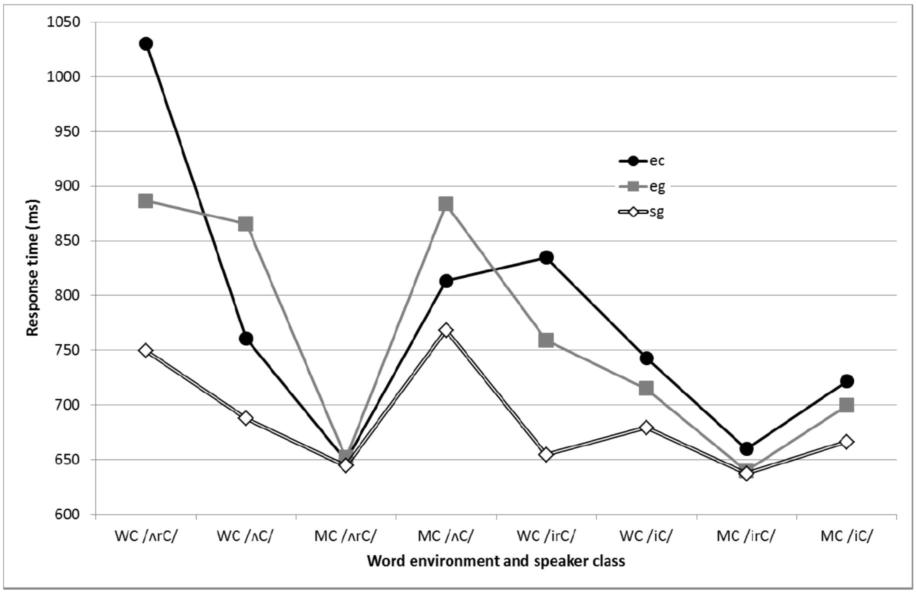


Figure 3.2: Response time by Listener group (ec: Cambridge, eg: Intermediate, sg: Glasgow), Vowel/Coda, and Class (Lennon 2013: 19)

The Intermediate listener group (termed '**eg**' in Lennon 2013, to indicate English listeners living in Glasgow) appeared to be efficient (i.e. *low* inaccuracy, green circle in Figure 3.1) at identifying working class *hurt* words (which are hypothesised to be unfamiliar to non-native Glaswegians). However, they were relatively inefficient when recognising the phonemically ‘simpler’ /r/-less words like *hut* (i.e. *higher* inaccuracy, red circle in Figure 3.1). An effect of perceptual hypercorrection by the Intermediate group was hypothesised. That is, this group's experience with Glaswegian speech might have taught them that words like ‘hurt’ can be produced with derhoticisation. They might therefore over-report hearing ‘hurt’, when presented with intended ‘hut’ as well as intended ‘hurt’. However this effect could not be confirmed without further in-depth analysis of the response data. Therefore, a full analysis was conducted.

Signal detection analysis allows response patterns to be analysed in terms of bias and sensitivity separately. This is useful for exploring the perceptual hypercorrection account which could involve listeners becoming biased to respond ‘hurt’, without necessarily becoming more sensitive to the phonetic details of the distinction. The following section will explain the theory behind signal detection analysis, and the section afterwards will look at how it was used in this thesis.

3.2.1 Introduction

Signal Detection Theory (SDT) (e.g. MacMillan & Creelman 2005; Creelman & MacMillan 1979; Heeger 1998) as a method of analysis is widely used in the fields of psychology and medical training, and is a useful way of determining the performance of a participant, or group of participants, in a perceptual experiment, or in detecting features on e.g. x-rays.

In SDT, there are two main aspects of a person's responses that can be measured: sensitivity to differences between stimuli or states, and bias towards one response or another. These aspects will be explained once we have seen how participants' responses can be classified in a way that allows for the calculation of sensitivity and bias.

Classification of responses

Figure 3.3 shows the underlying distribution of responses to a fictional experiment where participants must choose whether they judge a given visual stimulus to be a familiar (Old) face or an unfamiliar (New) face. The bottom graph shows the probability distribution of familiarity values for when the person sees the Old stimuli. If the participant correctly judges a previously seen ‘Old’ face to be familiar, they will press the appropriate button, e.g. marked ‘Old’. These types of response correlate with ‘Hits’, on the right side of this distribution curve. If, however, the participant incorrectly judges an Old face to be unfamiliar (e.g. by pressing the ‘New’ button), they will have missed the target: these responses are represented by 'Misses', on the left of the diagram.

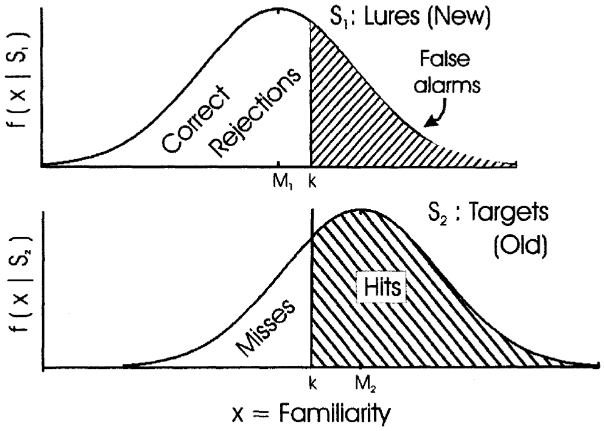


Figure 3.3: (MacMillan & Creelman 2005: 17).

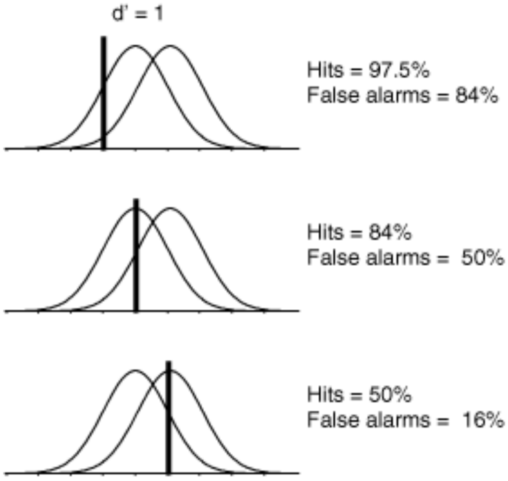


Figure 3.4: (Heeger 1998).

Conversely, the top graph of Figure 3.3 shows the probability distribution of familiarity values for the New (unfamiliar) set of stimuli. If a participant correctly judges the face to be New, they will have provided a 'Correct Rejection' of the lure; these responses are represented on the left of the top graph. Finally, if the participant is successfully lured into believing that a New stimulus is Old, they will provide a 'False Alarm'; these incorrect responses are on the right of the distribution curve.

By looking at the two distributions in Figure 3.3 it can be seen that the judge whose responses are shown is moderately successful in their judgement of familiar and unfamiliar, as there is a higher proportion of Hits and Correct Rejections than their counterparts on each graph. However, the key to SDT is visualising these proportions together, in the same 'decision space', as can be seen in Figure 3.4, from Heeger (1998). This diagram shows three hypothetical decision spaces, which are defined as the distributions for both stimuli overlapping on the same graph, with the alignment determined by the location of the criterion line, k, which is also seen in Figure 3.3. Decision space graphs such as these are the most common way of representing and describing data in SDT, as all parameters can easily be seen at the same time, on the same graph. The graphs in Figure 3.4 show that when the proportions of Hits, Misses, False Alarms and Correct Rejections vary, this moves k to a different location, ultimately affecting the response bias (discussed later).

If the number of Hits, Misses, False Alarms and Correct Rejections are known, it is possible to calculate proportions of these parameters in relation to others; specifically, the Hit rate, H, is obtained by simply dividing the number of Hits by the combined number of Hits and Misses:

$$H = \frac{Hits}{Hits + Misses} \tag{3.1}$$

The same equation applies to determining the False Alarm rate, *F*:

$$F = \frac{FalseAlarms}{FalseAlarms + CorrectRejections} \tag{3.2}$$

Once *H* and *F* are known, it is then possible to perform many other calculations, beginning with sensitivity to differences in the task.

3.2.2 Sensitivity

A person's ability to discriminate between two states, 'Yes' or 'No', 'Old' or 'New', 'Stimulus 1' or 'Stimulus 2', is described using a measure known as *d′* (pronounced 'dee-prime'): this is the Sensitivity Index, and it shows how efficiently a participant can make their choice about the stimuli with which they are presented. The formula is:

$$d' = z(H) - z(F) \tag{3.3}$$

This calculation is the difference between the Hit rate and the False Alarm rate, once they have been transformed to z-scores (i.e. units of standard deviation). It is equivalent to the distance in standard deviations between M_1 and M_2 on the graphs in Figure 3.3 (M_1 and M_2 are the statistical means – i.e. the locations of the peaks of the normal distribution curves). As another illustration of this measure, all the graphs in Figure 3.4 have the same *d′* value of 1: it is clear that the two distributions are the same distance apart in each of the three graphs, even though there are different proportions of responses, meaning that the sensitivity index is equal in each case. Finally, Figure 3.5 shows fictional data of medical trainees detecting a tumour on an x-ray. In the baseline for all three trainees (a), *d′* is 1, but increases to around 2 in (b), (c) and (d). The difference between distributions is easy to see, because each trainee's response distributions become further separated as they improved their sensitivity to difference in the stimuli after training.

While *d′* can be universally applied to perceptual experiments, it must sometimes be adjusted for the task in hand. For example, when discussing 'Old/New' 2AFC designs, Macmillan and Creelman (2005: 168) state that '2AFC is [an] easier task' than a 'Yes-No' design, in which the observer is asked to report the presence of a stimulus. The reason for this is that 'the observer estimates the familiarity of

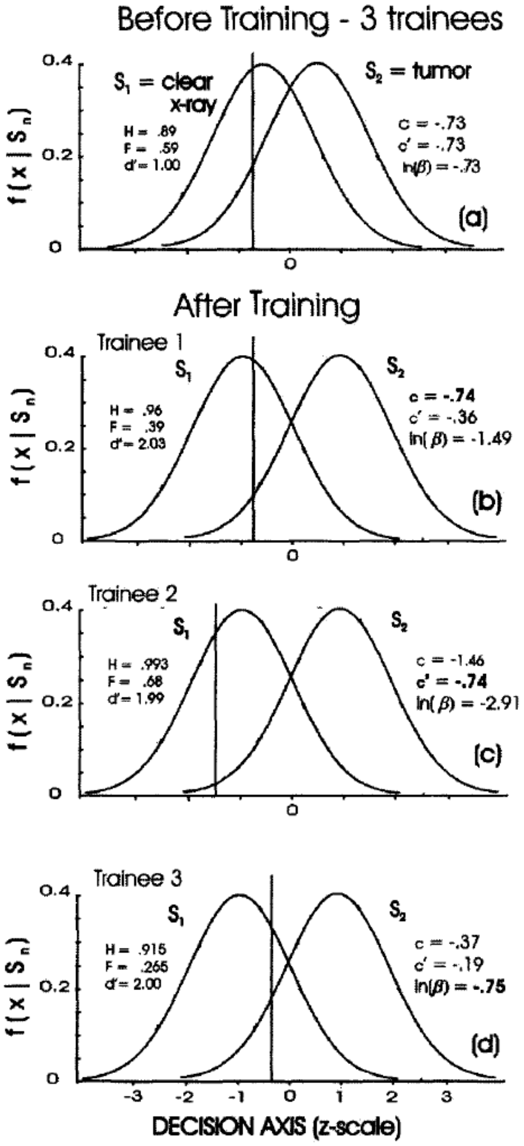


FIG. 2.3. Decision spaces for the three radiology trainees of Example 2b. In each case the hit rate, false-alarm rate, sensitivity, and three alternative criterion measures are shown. (a) Before training, $d' = 1.0$. The criterion c, the relative criterion c', and log likelihood ratio equal –0.73 for all trainees. (b) Trainee 1, after training; increased sensitivity and approximately the same criterion location c as before training. (c) Trainee 2, after training; increased sensitivity and approximately the same relative criterion location c' as before training. (d) Trainee 3, after training; increased sensitivity and approximately the same value of log likelihood ratio [ln(β)] as before training.

Figure 3.5: (MacMillan & Creelman 2005: 32).

each word independently' (2005: 168). In order to 'take account of the difference in difficulty between 2AFC and yes-no' (Macmillan and Creelman, 2005: 168) d' must be adjusted downwards by a factor of $\sqrt{2}$. This would therefore be expressed as:

(3.4) $$d' = \frac{1}{\sqrt{2}}[z(H) - z(F)]$$

This will give much lower values for d' for 2AFC tasks than for tasks with a 'Yes-No' design. Indeed, Macmillan and Creelman write that the 2AFC paradigm is popular because not only does it discourage bias, but performance levels tend to be high, which allows for the investigation of sensitivity to small differences between stimuli (2005: 179).

3.2.3 Response bias

Macmillan and Creelman write that response bias statistics can reflect either 'the degree to which yes responses dominate or the degree to which no responses are preferred' (2005: 29). They also state that a positive bias in the data is 'a tendency to say no, whereas a negative bias is a tendency to say yes' (2005: 29). This may initially seem counter-intuitive, but it can be seen in Figure 3.5 – in all four of the decision space graphs, the vertical criterion line *k* is on the left of the point where the two distributions cross, meaning a negative bias, and the trainees are more inclined to respond 'yes'. The three main measures of response bias will now be discussed.

Criterion location: *c*

The Criterion Location is one measure of response bias, represented simply by lowercase *c*. The formula for *c* is:

(3.5)
$$c = -\frac{1}{2}[z(H) + z(F)]$$

This is a fairly simple method of measuring response bias, as it does not take account of sensitivity. Indeed, mathematical proofs show that, out of all types of response bias measures, 'only *c* is [statistically] independent of *d*'' (Macmillan and Creelman, 2005: 41).

Relative criterion location: *c'*

Another method of showing response bias is Relative Criterion Location (*c'*), which is simply the Criterion Location (*c*) scaled relative to discrimination performance (*d'*). The formula is simply:

(3.6)
$$c' = \frac{c}{d'} = -\frac{1}{2}\frac{[z(H) + z(F)]}{[z(H) - z(F)]}$$

The main reason for using *c'* over *c* as a measure of bias is that it takes account of the person's sensitivity. For example, there is virtually no difference in the criterion location *c* between (a) and (b) for Trainee 1 in Figure 3.5 (-0.73 and -0.74), even though sensitivity improves after training: *d'* is higher in (b). Since the criterion is now on the other side of the mean of the leftmost distribution, due to the higher *d'*, the trainee's bias can be said to have changed: this possibility is reflected by the *c'* value changing from -0.73 to -0.36.

In relation to Relative Criterion Location *c'*, Macmillan and Creelman note that ‘when *d′* varies, one must decide whether in discussing bias one wishes to take account of sensitivity’ (2005: 33). A problem with the use of this calculation may arise in some circumstances. If sensitivity is very poor, so that *d′* is a very small number, then once *c* is divided by *d′* to find *c′* (as in equation 3.5), *c′* will become a very large number. This large number gives an artificially large value for response bias, even though the original criterion line *k* may be very close to 0, meaning very little bias. Worse, if sensitivity is so poor that *d′* is negative (i.e. when $F > H$: False Alarm rate is even marginally greater than Hit rate), *c′* will be an artificially large number of the opposite sign to the original *c*, which may heavily skew future calculations. This data point (or points) may then have to be excluded as an outlier. When dealing with ‘extreme’ data, where sensitivity is likely to be marginal, it may be best to avoid *c′* for these reasons.

Likelihood ratio β

Another measure of bias is the Likelihood Ratio (β). On each of the decision spaces in Figure 3.5 there is a value for the height of each of the distribution curves S_1 and S_2. At the point where the two distributions cross in the middle of the decision space the ratio between these values is 1 (they are the same height), but at any other point where the criterion *k* crosses the curves, there will be a number which represents the difference in the heights of the two curves. The ratio can be calculated by the formula:

$$\beta = e^{cd'} \tag{3.7}$$

By taking logarithms, an equivalent form is:

$$ln(\beta) = cd' = -\frac{1}{2}[z(H)^2 - z(F)^2] \tag{3.8}$$

Often, the Log Likelihood Ratio $ln(\beta)$ is reported, presumably because calculation is easier than the formula to find β alone. As can be seen in Figure 3.5, if the criterion line *k* is to the left of centre, the value of $ln(\beta)$ is negative, but if the line were to the right of the centre it would have a positive value.

3.2.4 Summary

In summary, the fundamental properties of a person's perceptual decisions, sensitivity (*d′*) and response bias (*c*, *c′*, or β), can be easily measured using SDT. An ideal observer, or rather, a perfectly efficient, unbiased observer, will have a sensitivity index *d′* which tends towards positive infinity (*d′*$\rightarrow \infty$), and will have no bias, so that the criterion *k* is exactly in the middle of the two distributions ($c = 0$, $\beta = 1$, *c'*$\rightarrow 0$ [tends towards 0, as *d'*$\rightarrow \infty$]). However, in perceptual experiments this ideal observer manifests as a ceiling effect, which is less than useful for most research questions.

In most experiments this does not happen, but sometimes extreme data is investigated using SDT. For example, an experiment may include stimuli which are very difficult for some participants to distinguish, producing very small – sometimes negative – *d'*, and this is interesting in itself. In this case, the researcher may wish to use a measure of bias which does not fall foul of the problems that can be caused by a small, negative *d'*, in other words, avoiding Relative Criterion Location *c'* (discussed above).

Likelihood Ratio β may be a useful alternative, as it is also scaled to *d'*. However, with a very small/negative *d'* it may suffer from the same problem as *c'*, albeit in a different way. As the neutral bias ratio is 1, values for bias may be artificially close to 1, which would not be helpful for analysis.

Criterion Location *c* therefore seems to be the better candidate for use as a measure of bias in these types of task, as it does not take any account of sensitivity and cannot be affected by very large or small data. Although this may not be ideal, it allows for the comparison of unaffected bias values, without the possibility of skewed data. Furthermore, as mentioned above, MacMillan and Creelman write that mathematical proofs show that ‘only *c* is [statistically] independent of *d'*, [out of all response bias measures]’ (Macmillan and Creelman, 2005: 41), so it also seems to be the logical candidate in terms of statistical analysis.

At this point it should be said that the term ‘Signal Detection’ is potentially misleading in this thesis. The reason is that the task described here is slightly different to some two-alternative-forced-choice tasks, in that the listeners do not perceive two stimuli at once, or one immediately following the other, and then report, which was (for example) strongest out of A or B. Because the listeners in this experiment hear *one signal*, then are presented with *two options*, and finally being asked to choose which option best described the signal, this could be described more as a ‘signal assignment’, or an ‘option detection’ task. In fact, Signal Detection Analysis could be thought of as nothing more than a tool, which is very flexible in terms of describing human performance of different types, in many different settings.

The predictions for the signal detection analysis for Experiment 1 are that the easiest minimal pairs to choose between will be those with the /i/ vowel, for example *bead/beard*, and those with /ʌ/, for example *bud/bird*, will be the hardest, reflected in reduced sensitivity and a greater response bias. Between the listener groups, it is predicted that the Glasgow group will show the greatest sensitivity to difference between stimuli and the least response bias, and the Cambridge listeners will show much less sensitivity to difference, with much more response bias than the Glasgow group. The Intermediate listener group is predicted to show a pattern of results between the other two groups, for both sensitivity and response bias.

3.3 Results

The previous section dealt with the theory behind signal detection and the reasons for using it in specific research contexts. The present section builds upon this knowledge, and will detail the procedure that was followed when the signal detection analysis was completed for Experiment 1. It is important at this stage to repeat the research question for this analysis: 'What is the role of experience in the perception of fine phonetic detail for a contrast?' Signal detection analysis can help us answer this question by showing whether groups of listeners with different degrees of experience differ in terms of bias, or sensitivity, or both.

Classification of responses

Responses to a two alternative forced choice (2AFC) task are usually 'Yes'/'No', 'Present'/'Absent', or some other binary choice regarding the presence or prominence of an object or state. However in this research, the binary nature of the choice between words like e.g. *hut/hurt* is more like 'Left'/'Right', meaning that for the listener, neither choice acts as the 'target'. This means that when applying the four terms 'Hit', 'Miss', 'Correct Rejection' and 'False Alarm' to the responses, one type of response must be chosen as the 'target', in order to calculate the formulae. It was decided – because the /r/ was closest to being the 'target' in this study – that the member of each minimal pair which orthographically had an <r> present, e.g. *hurt*, *beard*, etc., would be treated as the object that participants were 'aiming for' in the 2AFC. In other words, if a listener heard '*hurt*' and responded by selecting HURT (the *correct* response), this would be classified as a 'Hit', and if, for the same stimulus, the listener selected HUT (the *incorrect* response), this would be treated as a 'Miss'. To complete the 4-way matrix, if a listener heard the /r/-less member of the minimal pair, e.g. *hut*, and responded by selecting HUT (the *correct* response), this was classified as a 'Correct Rejection', and if for the same stimulus they selected HURT (the *incorrect* response), this was classified as a 'False Alarm'. The classification of these responses is summarised in Table 3.1.

	Response	
Stimulus	‘HURT’	‘HUT’
hurt	Hit	Miss
hut	False Alarm	Correct Rejection

Table 3.1: Classification of 2AFC responses for signal detection analysis. Stimulus *hurt* represents all stimuli with a postvocalic /r/ (*hurt*, *bird*, *third*), and Stimulus *hut* represents all stimuli without a postvocalic /r/ (*hut*, *bud*, *thud*). Similarly, Response HURT represents all response options with a postvocalic /r/ and Response HUT represents all response options without a postvocalic /r/.

The assignment of these classifications to the 2AFC responses was conceptually and practically a little cumbersome, because treating the /r/ words as the ‘target’ of each minimal pair seemed to be artificially imposing a hierarchy on the choice, when in fact the task was simply to discriminate between two options. Indeed, the way signal detection analysis is used here, it can be thought of as signal ‘discrimination’ analysis. However, it provided a useful structure of terminology in which to conduct the signal detection analysis, which was unrelated to the presence or absence of /r/ in the stimuli.

The hit rate *H* and false alarm rate *F* were calculated separately for each listener. Within each listener, the rates were calculated for each speaker class, and for each vowel environment. To demonstrate this, Table 3.3 shows the classification for a single subject's responses to only the WC /ʌ/ stimuli – that is, out of all the words in Table 3.2, only the ones in the right-hand column are classified here.

/i/	/ʌ/
bead/beard	*bud/bird*
feed/feared	*hut/hurt*
weed/weird	*thud/third*

Table 3.2: Minimal pairs used in the experiment

It is important to note that *d′* and *c* scores were calculated by subject, not by word pair. A ‘by word pair’ approach would have produced a (very slightly) different set of values for *d′* and *c*, but as the difference between this approach and the ‘by subject’ approach was negligible, and it was easier to implement from the structure of the data, the ‘by subject’ analysis was therefore chosen.

	Response		
Stimulus	‘HURT’	‘HUT’	(total)
hurt	Hit = 7	Miss = 11	(18)
hut	F.A. = 1	C.R. = 17	(18)

Table 3.3: Classification of 2AFC responses for subject ‘ec05’ (Cambridge group), responding to WC /ʌ/ stimuli. ‘*hurt*’ = all *hurt*, *bird* & *third* stimuli; ‘*hut*’ = all *hut*, *bud* & *thud* stimuli. ‘HURT’ & ‘HUT’ represent the responses in a similar fashion.

3.3.1 Statistical analysis

For each of the analyses in this chapter, the statistical program R was used to run linear mixed effects models (LMERs), in order to determine which of the experimental factors were important for any variation in the results. When building each of the models, a fully saturated model was constructed, including all of the experimental factors and all theoretically relevant interactions among them, and non-significant effects and interactions were then eliminated. Another possible approach taken by researchers using LMER models is to construct a very basic model, including only the effect that is hypothesised to be the most important or interesting to the research question. The model is then systematically built up to include more and more effects, and eventually interesting interactions may emerge as significant.

However, the saturated-model approach (i.e. ‘backwards’ stepwise regression) was taken in this project because of the relatively large number of factors (i.e. it would have been very time-consuming to manually build up the models; i.e. ‘forwards’ stepwise regression), and in order to account for any potentially complex and unforeseen interactions, it was decided that a more data-led approach to the modelling was the wisest course of action.

Using the lme4 package in R, saturated linear mixed effects models were created, in order to uncover which of the experimental factors most affected the dependent variable. In each of the analyses, the initial model included the following factors of interest, beginning with the fixed effects, followed by the random effects:

Fixed effects:

Group: Which listener group the participant was in, i.e. ‘Glasgow’, ‘Intermediate’, or ‘Cambridge’.

Class: Whether the stimulus was produced by one of the two middle class speakers or one of the two working class speakers, i.e. ‘MC’ or ‘WC’.

Vowel: Whether the vowel in the stimulus was phonemically /i/ or/ʌ/.

Random effect:

Subject: The effect of participant was included as a random intercept in the model to allow for likely, and potentially large, variation between participants' response behaviour.

Trial was not included in the models for this analysis. This was because it was necessary to average over items in order to conduct the signal detection analysis. In Experiments 2 and 3, described in later chapters, trials were randomised by participant.

An alpha level of .05 was used for all models. For each analysis below, the following saturated model was run. It included all three fixed effects, as well as all interactions including the 3-way interaction:

$lmer([dependent\ variable] \sim (group + class + vowel)\hat{}3 + (1|subject))$

In order to remove non-significant effects, the R package lmerTest's step() function was then applied to the model. step() tests the significance of each of the random effects, each interaction, and finally the main effects, removing from the model any non-significant effects, resulting in the best-fitting model. See the Results section of Chapter 2 for a more complete explanation of how the function works.

3.3.2 Sensitivity

Using the *H* and *F* values for each participant, it was possible to apply formula 3.9, obtaining the value for *d'*.

$$d' = \frac{1}{\sqrt{2}}[z(H) - z(F)] \quad (3.9)$$

These values for each participant were then averaged within listener groups in order to create the graphs below, resulting in mean values for sensitivity *d'*, for all listener groups, and for all stimulus types (WC /ʌ/, MC /ʌ/, WC/i/, & MC /i/). Therefore, the values for *d'* represent the listeners' ability to distinguish between the two members of each minimal pair they choose between (with or without an /r/), for each vowel environment (/i/ or /ʌ/), and for each speaker class (WC or MC).

After step() was run on the fully saturated model, the best-fitting model for *d'* (summary in Table 3.4) was:

$lmer(d' \sim (group + class + vowel)\hat{}3 + (1|subject))$

Table 3.4: Model summary for *d'* (Experiment 1)

	d'
group_Glasgow	0.053
	(0.150)
group_Intermediate	0.046
	(0.150)
class_WC	−0.120
	(0.139)
vowel_ʌ	-0.416^{***}
	(0.139)
group_Glasgow X class_WC	0.149
	(0.194)
group_Intermediate X class_WC	0.082
	(0.194)
group_Glasgow X vowel_ʌ	0.122
	(0.194)
group_Intermediate X vowel_ʌ	−0.252
	(0.194)
class_WC X vowel_ʌ	-2.344^{***}
	(0.196)
group_Glasgow X class_WC X vowel_ʌ	1.608^{***}
	(0.274)
group_Intermediate X class_WC X vowel_ʌ	0.955^{***}
	(0.274)
Constant	3.697^{***}
	(0.107)
Observations	248
Log Likelihood	−175.462
Akaike Inf. Crit.	378.924
Bayesian Inf. Crit.	428.112
Note:	*p<.1; **p<.05; ***p<.01

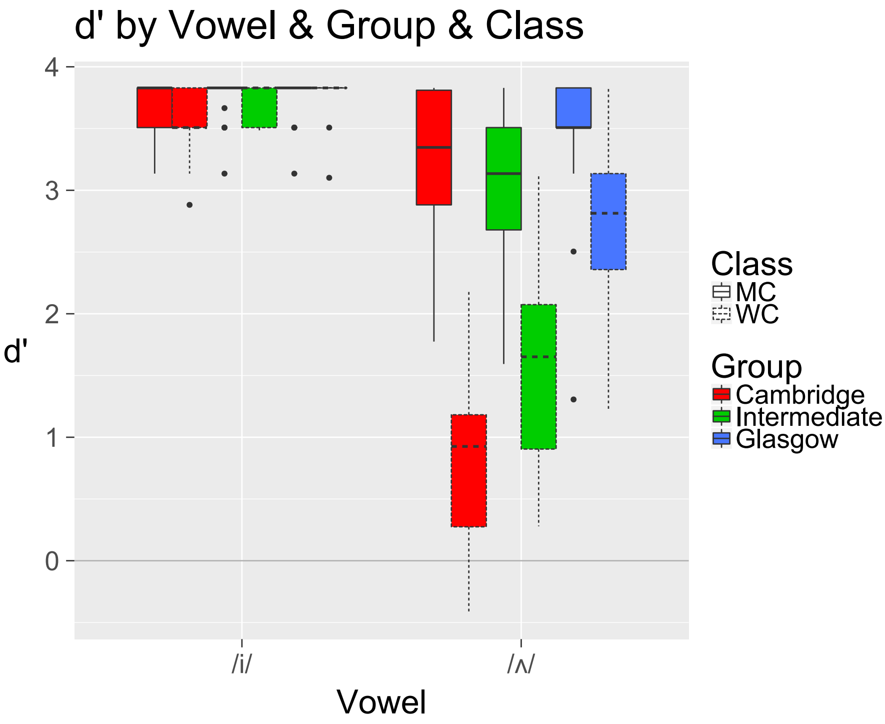


Figure 3.6: Experiment 1 *d′* by Vowel, Group, & Class

Figure 3.6 shows *d′* by Vowel, i.e. whether the stimuli had an /i/ or an /ʌ/; by group, i.e. whether the listener was in the Cambridge (red), Intermediate (green), or Glasgow (blue) listener group; then by Class, i.e. whether the speaker was middle class (solid line) or working class (dotted line).

This was a significant 3-way interaction (Pr(>F)<.001, F=17.3548). The *d′* differences for this interaction were achieved using manual factor comparisons with Bonferroni correction, using the pairwise.t.test() function in R. This was done because lmerTest's step() function only shows factor comparisons for effects up to and including two way interactions. Because there were three comparisons in each of the vowel conditions, the significance level (alpha = .05) was adjusted by dividing by three, resulting in a new alpha of .01667. Consequently, there were no differences in *d′* for any comparisons for /i/ stimuli, but significant differences in sensitivity for /ʌ/ stimuli, specifically between middle class and working class /ʌ/ stimuli for Cambridge (p<.001), Intermediate (p<.001), and Glasgow listeners (p<.001).

3.3.3 Response bias

As discussed in the section above which introduced signal detection theory, there are different measures of bias which can be used: *c*, *c'*, and β. For the reasons stated in that section (e.g. statistical independence from *d'*, among other reasons), it was decided that 'response bias', *c*, would be used.

Again using the *H* and *F* values for WC /ʌ/ stimuli, formula 3.10 was applied, resulting in the value for response bias, *c*, for each individual participant.

$$c = -\frac{1}{2}[z(H) + z(F)] \tag{3.10}$$

The *c* values for each participant were then averaged within listener group, resulting in mean values for *c*, for all listener groups, and for all stimulus types (WC /ʌ/, MC /ʌ/, WC/i/, & MC /i/).

Because of the way in which the *H* and *F* values were calculated, a **positive** response bias *c* indicates that, when the listener responds to a minimal pair, they are biased towards reporting that they heard a word *without* an /r/; that is, given multiple choices of 'HUT' or 'HURT' throughout the task, they are more likely to respond 'HUT'. Conversely, if there is a **negative** value for response bias *c*, this indicates that the listener is more likely to report hearing a word *with* an /r/.

After step() was run on the fully saturated model, the best-fitting model for *c* (summary in Table 3.5) was:

$lmer(c \sim (group + class + vowel)\hat{}3 + (1|subject))$

Table 3.5: Model summary for *c* (Experiment 1)

	c
group_Intermediate	-0.136 (0.094)
group_Glasgow	0.120 (0.094)
class_WC	0.853^{***} (0.087)
vowel_i	0.161^{*} (0.087)
group_Intermediate X class_WC	-0.785^{***} (0.122)
group_Glasgow X class_WC	-0.687^{***} (0.122)
group_Intermediate X vowel_i	0.144 (0.122)
group_Glasgow X vowel_i	-0.080 (0.122)
class_WC X vowel_i	-0.758^{***} (0.123)
group_Intermediate X class_WC X vowel_i	0.716^{***} (0.173)
group_Glasgow X class_WC X vowel_i	0.610^{***} (0.173)
Constant	-0.210^{***} (0.067)
Observations	248
Log Likelihood	-65.660
Akaike Inf. Crit.	159.321
Bayesian Inf. Crit.	208.509
Note:	*p<.1; **p<.05; ***p<.01

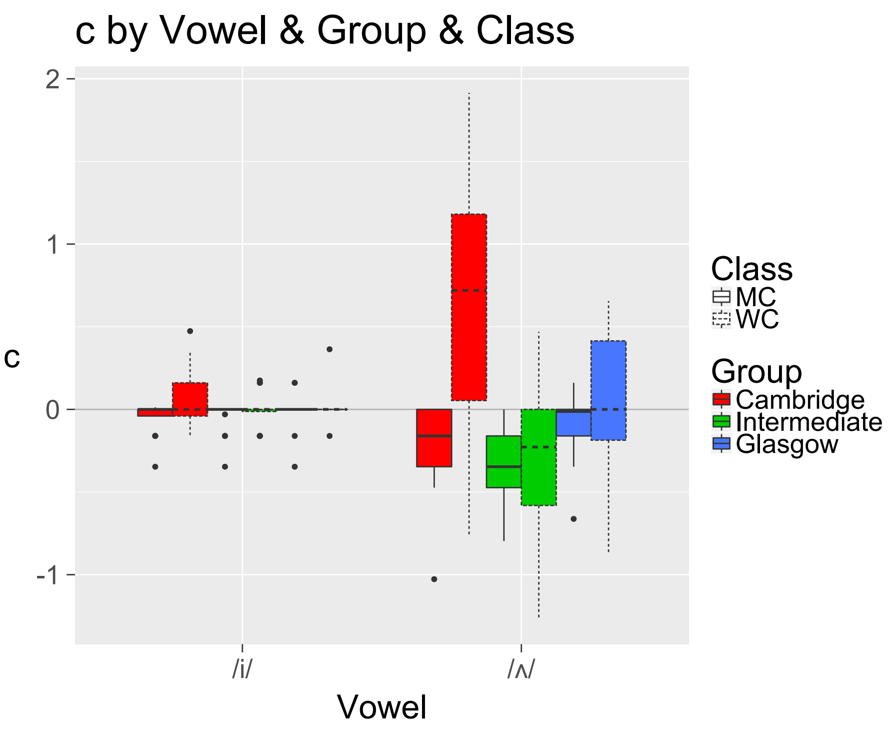


Figure 3.7: Experiment 1 *c* by Vowel, Group, & Class. Positive values of *c* indicate a bias towards responding HUT

Figure 3.7 shows *c* by Vowel, i.e. whether the stimuli had an /i/ or an /ʌ/; by group, i.e. whether the listener was in the Cambridge (red), Intermediate (green), or Glasgow (blue) listener group; then by Class, i.e. whether the speaker was middle class (solid line) or working class (dotted line).

This was a significant 3-way interaction (Pr(>F)<.001, F=9.9558), showing no differences in *c* for any comparisons for /i/ stimuli, using the Bonferroni-corrected significance level of .01667 (see above). There was no difference between classes for /ʌ/ stimuli for the Intermediate listener group (p=.52), or the Glasgow group (p=.08). However, there was a significant difference in bias between middle class and working class speakers for the Cambridge listeners hearing the /ʌ/ stimulus pairs (p<.001). The pattern was the same as for the Glasgow listeners (who were tending towards hearing middle class /ʌ/ stimuli as more /r/-ful, but this was not significant) but was much more extreme. When middle class speakers were heard, Cambridge listeners were biased towards reporting *hurt*-like words (c=-0.2100), but when hearing working class speakers they were biased towards reporting *hut*-like words (c=0.6428).

3.4 Discussion

The focus of this chapter has been a more detailed analysis of the results of Experiment 1, which was originally conducted for Lennon (2013). In that project, the primary finding was that in the 2AFC task, the most accurate listeners were the group who were the most familiar with the Glaswegian accent, namely those who grew up in and around Glasgow, and the least accurate were the listeners who grew up in the South East of England and lived in Cambridge at the time of testing – these listeners had the lowest amount of long-term experience with Glaswegian speakers. The secondary finding was that the Glasgow group also had the lowest reaction times when responding, which was hypothesised to be because of a lower cognitive load on the listeners.

A ‘crossover’ effect was found in both the accuracy and response time analyses, such that the Intermediate listeners appeared to be better at processing the more ‘difficult’ working class *hurt* words than *hut* words. While an explanation for this pattern was suggested in Lennon (2013), it was only possible to hypothesise about the explanation for these unexpected results (i.e. perceptual hypercorrection), given the fairly basic nature of the analysis methodology. A further analysis was needed, making use of a more sophisticated method which would be much more sensitive to the responses made by the participants. Signal detection analysis was decided upon as it allowed for the high level of scrutiny of the listeners' responses that was required. This methodology was applied to the research question for this chapter, which was:
‘What is the role of experience in the perception of fine phonetic detail for a contrast?’

This new analysis showed a clear effect of accent experience on a listener's ability to distinguish between the perceptually ambiguous words, which are known from Chapter 2 to be acoustically very similar. It also uncovered an effect of listener group on the bias that they display – in other words, how predisposed they may be to responding one way or another. While the listeners in Glasgow showed very little bias in their responses, the Cambridge listeners (who had the least experience with working class Glaswegian) were very biased to responding *hut*, whether the stimulus was *hut* or *hurt*. What follows is a discussion of these results, interpreting them in terms of their likely causes and implications.

3.4.1 Sensitivity

The main finding for listener sensitivity, as shown in the *d'* values, was that the more experience listeners had with the Glaswegian linguistic environment, the more sensitive they were to differences between the stimuli. The main effect of Group followed predictions for familiarity, such that sensitivity was highest for Glaswegian listeners, less for Intermediates, and worst for Cambridge. In essence, this was the main reported finding of the original analysis in Lennon (2013), and confirms the finding of that paper that long-term experience is very important for a listener to be able to successfully discriminate between difficult word pairs.

Another prediction was prompted by Lennon's (2013) results, and also by the results of the acoustic analysis detailed in Chapter 2 of this thesis, namely that the middle class word pairs would be easier to discriminate than the working class pairs. This was supported by the main effect of Class, which showed that there was a very large difference in sensitivity to MC pairs than to WC pairs. This was unsurprising, and once again supports the previous work.

The final main effect of Vowel was also highly significant, with word pairs like *bead/beard* being much easier to discriminate than *hut/hurt* pairs. There was no expectation that listeners would have difficulty in discriminating between these /i/ word pairs, whether they were produced by a middle class speaker or a working class speaker. While the formant structure of the /r/ is very different between the Glaswegian middle class and working class sociolects, it is also very different to the formant structure of the preceding /i/ in each case. Thus, listeners would have no difficulty in perceiving the obvious ‘glide’ from /i/ to /r/ in either the middle class or working class speakers' productions.

Importantly, all possible interactions were significant, showing that no effect can be interpreted on its own. This too, is anticipated by the acoustic analyses. All possible factor interactions were significant for *d'* in this experiment: GroupXClass, GroupXVowel, ClassXVowel, and the three-way interaction of GroupXClassXVowel. Taken together, the interactions all contribute to the story, in that the easiest pairs to discriminate were generally the middle class and /i/ stimuli, by Glaswegian listeners. At the other end of the scale, the high acoustic similarity between working class *hut* and *hurt* words is a likely explanation for the low *d'* for the Cambridge listeners.

This all appears to support the patterns in the main effects fairly neatly. However, a closer look at Figure 3.6 shows that middle class *hut/hurt* words (the boxes on the right with the solid outlines) are more difficult to discriminate than other ‘easy’ contrasts, such as the working class and middle class *bead/beard* pairs (the boxes on the left of the graph). This may be because of the phonetic similarity of the middle class *hut* words to both *hut* and *hurt* words produced by the working class speakers; i.e. all three word types have relatively steady formant frequencies with low F2 and high F3 throughout the vocalic rime section, which is contrasted with the highly rhotic middle class *hurt* words, which have a lowered F3, making them acoustically distinct from the other three word types. Perception of the middle class *hut* words may therefore have suffered because they were heard randomised in the same block as the working class *hut/hurt* words. The listeners might have been experiencing extra cognitive loading for this reason, compared to a hypothetical situation in which they only heard one speaker, or one class.

However, it is also the case that the listeners' processing of the middle class *hut* words does not suffer as much as their processing of the working class *hut* and *hurt* words, as seen when comparing the solid lined boxes to the dotted lined boxes in Figure 3.6. This may be even further evidence for talker/phoneme integration in perception. The fact that the MC *hut* words are being produced by the middle class Speakers ‘A’ and ‘B’, whose voices are now known to the listeners, means that the listeners may be using some of their knowledge about those speakers in an indexical fashion, in order to help their perception of the identity of the word or phoneme. This is addressed in an aspect of the design of Experiment 3, which is described in Chapter 5.

Overall, sensitivity to difference reveals some of the challenges experienced by the listeners in this experiment, but the clearest pattern is that listener experience greatly affects sensitivity to difference in fine phonetic detail for phonemic contrasts. A closer look at differences in response bias provides another source of information.

3.4.2 Response bias

Like the *d'* results, there was a large effect of Class, with listeners being biased towards reporting hearing /r/-ful words for the middle class speakers, and /r/-less words for the working class speakers. A closer inspection of the data helps to clarify this: the three-way interaction of VowelXGroupXClass in Figure 3.7 shows that the working class stimuli evoked much more variable responses than the middle class stimuli.

At first glance it might seem odd that there was no main effect of vowel for response bias, given that there was such a large difference for sensitivity, as discussed above. However, closer inspection reveals why there may be no difference. Figure 3.7 shows that there is a great deal of both positive and negative bias in the /ʌ/ stimuli, across different factors. Looking at this data only from the perspective of the factor of Vowel is uninformative here, as it effectively washes out the differences, keeping the average *c* for /ʌ/ relatively close to zero, which is similar to /i/. This explains the apparent lack of difference between the vowels, and underlines the importance of closer and more sophisticated inspection of this type of data.

There was however a big effect of Group, in that the Cambridge listeners were biased to reporting /r/-less words, the Intermediate group were even more biased to reporting /r/-ful words, and the Glasgow listeners were, in general, not biased either way. The fact that there are such large differences in main effects shows that long term listener experience has a big effect on the perception of fine phonetic detail for contrasts between words. However, to interpret these results more effectively we must look more closely at the interactions, all of which were significant.

The Intermediate listeners appear to be following a different pattern than the Glasgow and Cambridge groups. The boxes for Cambridge and Glasgow in Figure 3.7 both show a bias towards reporting *hurt* for middle class words and *hut* for working class words. This seems logical, as the more strongly-rhotic middle class tokens are, in general, more likely to evoke more /r/ responses than the weaker, more vowel-like derhoticised working class tokens. The pattern is in the same direction for both groups, but it is simply a stronger effect for Cambridge than Glasgow.

However the Intermediate listeners show a bias towards reporting hearing /r/-ful words in both the strongly rhotic middle class tokens *and* the weakly rhotic working class tokens. These new results from response bias indicate that these Intermediate listeners are indeed ‘perceptually hypercorrecting’. The Intermediate listener group is made up of people whose linguistic experience when growing up was similar to that of the Cambridge listener group, as they all progressed through childhood in England perceiving and producing non-rhotic accents. The difference between the groups is that the Intermediate listeners had an average of 3.6 years of living in the Glasgow area, hearing Glaswegian accents for that period of time. In comparison, the Cambridge listeners had almost no exposure to the Glaswegian linguistic environment, so they were relatively naive to the existence of derhoticised /r/, unlike the Intermediate group. Their inexperience with a feature that is very close to being a plain vowel, is probably (and perhaps understandably) responsible for their tendency to mistakenly classify the /r/ in working class *hurt* words as a plain vowel. This explanation addresses the ‘crossover’ effect reported in Lennon (2013).

When hearing middle class /ʌ/ stimuli, the response bias for all groups was biased towards responding *hurt*. As well as the result that responses to the working class stimuli were biased towards *hut*, this result was unexpected. However, it is perhaps even more surprising, given the relative ease with which listeners processed the 'easier' middle class *hut/hurt* pairs, compared to the working class *hut/hurt* pairs. A possible explanation for this pattern of results is that listeners may have been processing the stimuli in such a way that they were 'building in' the identity of the speaker to the stimulus they were hearing, and therefore allowing for how likely the speaker was to produce a certain variant. Thus, because they knew that Speakers 'A' and 'B' (the middle class speakers; as distinct from 'C' and 'D', the working class speakers) were likely to produce strongly rhotic *hurt* words, based upon the evidence that Speakers 'A' and 'B' had actually produced *hurt*, or *bird* earlier in the task, they could then use this knowledge when they next heard Speaker 'A' or Speaker 'B', and respond accordingly. When they next heard either of these two speakers, they may have identified them by their voice quality as being likely to produce a strongly rhotic /r/, in order to set up the expectation that they were likely to be 'r-ful'.

In this way, it may be that the listeners were 'packaging' the phonetic detail along with the identity of the speakers to aid their perception. This suggests that the listeners are building up an inventory of individual exemplars for each speaker/word 'instance'. This appears to be consistent with exemplar theory, as previous research has suggested that the talker is processed along with the phoneme or the word, when hearing a stimulus (e.g. Mullennix & Pisoni 1990; Cole et al. 1974; etc.).

However, a Bayesian approach to speech perception might equally be used to explain these results. As Smith writes, listeners may make decisions based on a combination of knowledge or expectation, and evidence (2013). Furthermore, Kleinschmidt, Weatherholtz & Jaeger write that the 'ideal adapter' *probabilistically infers* the likelihood of each possible linguistic unit, given their prior knowledge of the distribution in question (Kleinschmidt, & Jaeger 2015; Kleinschmidt, Weatherholtz & Jaeger 2018).

3.4.3 Summary

The results presented in this chapter have shown that listener experience matters, but in concert with other aspects – the phonetic context (i.e. the preceding vowel) and the social class of the speaker. This is most extreme for the least experienced, and least for the most experienced, but even they have problems with some contrasts, showing that the acoustic patterns of these stimuli are indeed ambiguous.

If the response bias and sensitivity results are taken together, it may be deduced that the Intermediate listeners *have* improved their perception of derhoticised /r/, in that they are much better at discerning differences between the words than the

Cambridge group. However, their strong bias towards hearing /r/ in all the /ʌ/ tokens – compared with the Glaswegian listeners – indicates that they are still a long way from being native-like in their perception of the feature, suggesting that much more than three years' casual (and possibly infrequent) exposure to an unfamiliar phonetic feature is needed for improvement. It therefore could be said that the Intermediate listeners were almost as inefficient in discriminating the working class *hut/hurt* pairs as the Cambridge listeners, just in a different way.

This also raises the question of what happens when inexperienced listeners (like the Cambridge listener group) *start* to learn fine phonetic detail. How long does it take to reach the stage of proficiency that the Intermediate listeners in Experiment 1 have achieved? This question will be addressed in the next chapter.

Chapter 4

Experiment 2: Short-term adaptation

4.1 Introduction

The results from Experiment 1 showed the influence of listener experience on perceiving phonetic detail for both ends of the Scottish rhotic sociolinguistic continuum. Experiment 2 will add another dimension to this research, by examining the role of short-term learning in a listener's perception of unfamiliar phonetic detail. Specifically, we now focus on the perception of the acoustically ambiguous derhoticised /r/, which was clearly the most challenging variant for all listeners. Where Experiment 1 tested listeners with different levels of experience on their ability to distinguish similar words, Experiment 2 will examine what happens when listeners *begin* to learn the distinction between these words. In other words, they will be tested on their ability to learn fine phonetic details which present themselves as subtle acoustic differences.

In Experiment 1, the Intermediate listeners – who had some experience with Glaswegian – displayed the most interesting pattern of responses. They displayed a bias towards the presence of /r/, whether the word they heard was produced as *hut* or *hurt*, meaning that they were hypercorrecting their perception of derhoticised /r/. These results clearly show that increased familiarity with the unfamiliar derhoticised /r/ variant alters its perception. However, this effect appears to vary depending on the amount of experience the listener has. Because the Glaswegian listeners displayed by far the best performance, their familiarity with the variant clearly aids their discrimination between *hut* and *hurt* words. However, it is less clear what the effect is for the Intermediate listeners: their perception is altered, but it certainly cannot be said to be as ‘correct’ as that of the Glaswegian listeners. It is also interesting to note that the Intermediate listeners do not show a strictly intermediate degree of performance (i.e. between the Cambridge and Glasgow groups), as if they were ‘half way there’ in terms of learning the *hut/hurt* distinction. If they did not show any hypercorrection, as indicated by their bias in responding *hut*, it could be said that they were learning in a rather more linear fashion.

Experiment 2 uses the same three listener groups, with varying experience of Glaswegian as before (Group), in line with the design of the previous experiment. Each group's perceptual performance is tested before and after hearing some Glaswegian speech. Furthermore, in order to shed light on the importance of some key acoustic differences which *are* present between e.g. *hut* and *hurt*, these acoustic differences are removed for a control group of listeners. Accordingly, the experiment has two conditions. The experimental group – the ‘Natural’ condition – hear the unchanged speech in the exposure phase, and their level of improvement (if any) is measured in the difference in performance between their Pretest and Posttest. The control group – the ‘Altered’ condition – hear the same read passage, but the target words will have been acoustically manipulated to remove difference along some key acoustic dimensions. The acoustic manipulation is directly based on the new findings given in the acoustic analysis presented in Chapter 2. If the experimental group improve more than the control group, this may be interpreted as evidence for listeners perceptually learning from the acoustic differences that are present in the unmodified words.

The initial prediction is that in the initial task (before being exposed to the Glaswegian speech), all listener groups will show similar results to the long term familiarity found in Experiment 1 for both sensitivity to stimulus difference and response bias, such that Glasgow listeners will be the most sensitive to difference with very little bias, Cambridge listeners will be the least sensitive and will show a large amount of bias towards hearing no /r/, and Intermediate listeners will have an intermediate sensitivity but will show signs of perceptual hypercorrection in their response bias, reporting more /r/ words (i.e. the opposite pattern to the Cambridge listeners). Response times will also be analysed, and following the response time results in Lennon (2013) (see Figure 3.2), the prediction is that the Glasgow listeners will be the fastest, Cambridge the slowest, and the Intermediate listeners between the other two groups. Responses to words canonically without /r/ are predicted to be slightly faster than responses to words with /r/, but because of the acoustic similarity between the minimal pairs this effect is not expected to be strong.

Following the listeners' exposure to the Glaswegian speech, it is predicted that there will be very little improvement between pretest and posttest for the groups in the Altered condition, in which minimal pairs have been made more similar along certain acoustic parameters. However there is predicted to be a degree of improvement for listeners in the Natural condition, as they will be exposed to the words with the original features, and therefore have the opportunity to learn differences between the minimal pairs. Response times are predicted to improve slightly between pretest and posttest for these listeners, with sensitivity increasing for all listener groups. Responses for the Cambridge listeners are predicted to become less biased towards reporting no /r/, and for the Intermediate listeners they are predicted to become less biased towards over-reporting the presence of /r/. For the Glasgow listeners, much less change is predicted in the response bias, as they are already predicted to display little bias.

4.2 Experiment 2

4.2.1 Design

The two-alternative-forced-choice (2AFC) experimental design was again chosen, primarily because a number of analysis methods are possible from the experiment's output. 2AFC tasks, when implemented using the perceptual testing software DMDX, provide both response accuracy and reaction time data for each participant. For this experiment it is therefore possible to investigate the descriptive statistics of both accuracy and reaction time, and subject the results to more sophisticated statistical analyses such as linear mixed effects modelling and signal detection analysis, as implemented in the previous chapter.

The experiment has three sections:

1. Pretest: a two-alternative-forced-choice task (2AFC)
2. Exposure: a read passage, produced by the same native working class Glaswegian speaker as in the Pretest
3. Posttest: a second 2AFC task

Listener experience (Group):

For this experiment, it was judged to be important to replicate the same participant groups as Experiment 1. This was because Experiment 2 was seen partly as a development of Experiment 1, as long term learning could be inferred from any differences in accuracy and reaction times between listener groups. Therefore, in line with the design of the previous experiment there are three groups of listeners with different levels of experience with Glaswegian.

Condition:

Participants were divided into two Experimental Conditions, Natural and Altered, depending on the phonetic detail contained in the exposure passage. One group – the Natural condition – hears the speech in the exposure phase, and their level of improvement (if any) is measured in the difference in performance between their pretest and posttest tasks. The other group – the Altered condition – act as a control group. They will hear the same read passage, but all the target tokens have been acoustically manipulated so they are identical in three key acoustic dimensions: F2 trajectory, F3 trajectory, and duration of the vocalic portion. In other words, the existence of Altered signals effectively leads to the presentation of homophones.

In order to change the vocalic portion's duration, a steady section of the voicing is removed, or duplicated and spliced in, to shorten or lengthen the vocalic portion respectively. To avoid the possibility of lexical learning, that is, listeners simply remembering the pronunciation of one word and applying that knowledge across experimental tasks, different target words appear in the exposure phase (e.g. Barden & Hawkins 2013). For example, if *cut/curt* and *fussed/first* appear in the pre- and post-test phases, they will not appear in the exposure phase: minimal pairs such as *bud/bird* and *thud/third*, etc. will appear instead. This will mean that listeners learn (or not) from the acoustics, not from the individual words themselves.

4.2.2 Participants

Six groups of listeners were recruited from both the University of Cambridge ('Cambridge' groups), and the University of Glasgow ('Intermediate' and 'Glasgow').

- Natural Condition (passage with unaltered parameters):
  - Cambridge: raised in S.E. England n = 21
  - Intermediate: raised in England, living in Glasgow for more than 1 year n = 22
  - Glasgow: raised in Greater Glasgow area, living in Glasgow n = 21
- Altered Condition (passage with altered parameters):
  - Cambridge: raised in S.E. England n = 21
  - Intermediate: raised in England, living in Glasgow for more than 1 year n = 22
  - Glasgow: raised in Greater Glasgow area, living in Glasgow n = 21

In both locations, recruitment was achieved through posters in faculties and colleges, use of previous researchers' participant lists, online notice boards, and experimental recruitment databases:

4.2.3 Creation of acoustic stimuli

The stimuli for this experiment were tokens of working class Glaswegian, produced by one of the speakers recorded previously for Experiment 1. Perceptual improvement is measured from Pretest to Posttest (with Exposure appearing between the two), so the stimuli were the same in both Pretest and Posttest. In order to test whether listeners could learn from the relationship between F2 and F3 (which were shown in Chapter 2 to be important for distinguishing between minimal pairs), differences were artificially removed using Praat's source-filter resynthesis, and these stimuli would be used for one of the two listening conditions.

Recordings

For the two-alternative-forced-choice sections of the experiment, stimuli were segmented from high-quality recordings of a read word list, produced by one native speaker of working class Glaswegian. The stimuli used in the experiment were sets of minimal pairs with or without postvocalic /r/, as well as distractor minimal pairs. These are listed in Table 4.1. Two tokens of each word were used, so the total number of words in each of the Pretest and the Posttest is 96. The stimuli in Pretest and Posttest were not altered or resynthesised, whereas both versions of the story were. The details of this are in the next section.

Stimuli were produced by one of the working class speakers from the MSc experiment (male, 28 years old, from Maryhill in the North West of Glasgow). This was partly because he was personally known to the researcher, but also because of the need for a reliable speaker for this experiment – because he had been recorded for this project before, he was familiar with what was required of him. More importantly, this speaker is a very typical, natural user of derhoticised /r/, so his pronunciation did not have to be coached in any way. The use of the same speaker also means that there can be a degree of comparability between the experiments.

The recordings were made using a Beyerdynamic TG H74 high-quality headset microphone, through a Rolls LiveMix MX34 2-channel mixer, onto a Dell desktop computer, with Audacity recording software. They took place in the sound-attenuated recording booth at Glasgow University Laboratory of Phonetics. Once the sound level checks were completed, the speaker was asked to read from a short story (Appendix 2 – to be found at the end of the thesis).

There is a large number of pre-existing passages used in the fields of phonetics, sociolinguistics, and psychology, designed to elicit particular words or phonemes in certain environments. However, no existing passage was found to have the correct words or environments to act as the Exposure passage for this experiment, nor could any existing passage be adequately adapted for the purpose. A new passage was therefore written, by deciding on the target words that would need to appear, then constructing a story around them. All the target words appeared in phrase-final position, primarily to avoid effects of coarticulation from following segments. It was also deemed necessary to ensure that postvocalic /r/ appeared nowhere but the target words, so no other words in the story contained an /r/. This meant that the only postvocalic /r/ exemplars that would be heard by the participants would be in tightly controlled environments. The story was around six minutes in length, and was just under 1000 words. The word pairs used in the Exposure story are shown in Table 4.1 (there were two tokens of each word in the passage), but see Appendix 2 for their surrounding context.

Test		Exposure	
Target pairs	Distractor pairs	Target pairs	Distractor pairs
bust/burst	*bad/pad*	*bud/bird*	*Ben/pen*
cud/curd	*bait/beat*	*bun/burn*	*bet/pet*
cuss/curse	*bake/beak*	*hut/hurt*	*big/pig*
cut/curt	*ban/pan*	*shut/shirt*	*bin/pin*
fussed/first	*baste/beast*	*thud/third*	*bit/pit*
spun/spurn	*bat/pat*	*tonne/turn*	*bunch/punch*
	beg/peg		*coast/cost*
	bunk/punk		*code/cod*
	butt/putt		*fade/feed*
	coat/cot		*fate/feet*
	cone/con		*goat/got*
	cop/cope		*hate/heat*
	dot/dote		*hope/hop*
	make/meek		*mane/mean*
	mop/mope		*mate/meat*
	not/note		*shape/sheep*
	same/seem		*shown/shone*
	snake/sneak		*soak/sock*

Table 4.1: Minimal pairs used in both Pretest and Posttest (left), and embedded in Exposure story (right). Two tokens of each word in each target pair appeared in the Pretest/Posttest, or in the Exposure story; total = 24 (each Test) & 24 (Exposure). Two tokens of each word in each distractor pair appeared in the Pretest/Posttest, or in the Exposure story; total = 72 (each Test) & 72 (Exposure). Total words = 96 (each Test) & 96 (Exposure).

Before the recording session, the speaker was informed not to worry if he made any mistakes, as sections could be spliced together later – he should just begin again at the start of the paragraph in which the mistake was made. Each time this happened a new recording would be started, so the breaks could easily be found at the editing stage later. The headset microphone itself was fairly lightweight and unobtrusive, and the speaker worked in a music studio, so was very familiar with recording equipment. He was aware that the researcher would be listening on headphones just outside the booth (with the door closed for sound insulation), so there was no issue whenever he needed to be asked to repeat a section. This was all done with the aim of creating a relaxed and friendly atmosphere, in order to elicit as naturalistic a speech style as possible. The recordings were later spliced together, at zero crossings where necessary, in order to create a seamless sound file. Since Praat has the function of automatically moving the cursor to zero-crossings, enabling precise placement of spliced material, the margin of error for this process was zero.

Once all sections of the story were recorded, the speaker was asked to read from a word list (Appendix 3), which included all the target and distractor tokens for the 2AFC tasks (Table 4.1).

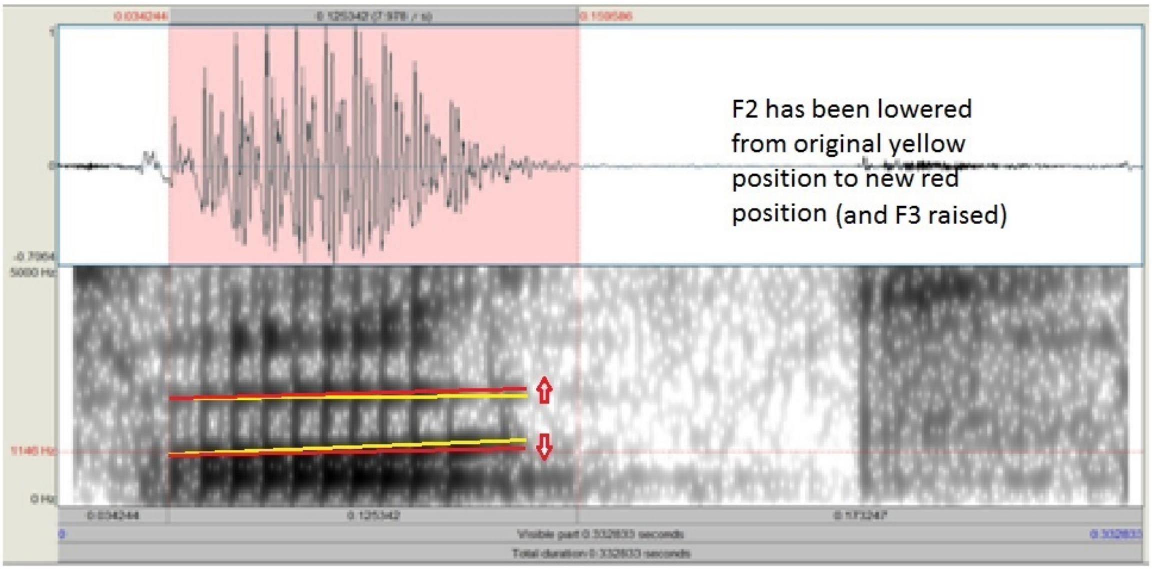
Creation of Altered stimuli

The acoustically manipulated tokens for the Altered Condition were created by altering the three acoustic dimensions of F2, F3, and vocalic duration, using Praat's source-filter synthesis (Boersma & Weenink 2006, Boersma 2006). The following protocol was used in the process, and is adapted from Praat's online tutorial, accessible at: http://www.fon.hum.uva.nl/praat/manual/Source-filter_synthesis_4_Using_existing_sounds.html. Bullet points denote instructions on which buttons to press in Praat's object menu:

1. Resample the sound file to 12kHz (keep the original, as it is required later):
   - *Convert > Resample... > 12000, 50*
   - (12000 = new sampling freq., 50 = precision/samples)
2. To extract the source, create an LPC object from the resampled sound:
   - *Analyse spectrum > To LPC (burg)... > 12, 0.025, 0.02, 50.0*
   - (12 = prediction order, 0.025 = window length, 0.02 = time step, 50.0 = pre-emphasis frequency)
   - Keep time step at 0.02s so number of formant points is manageable
3. Select this LPC object and the resampled sound from step 1. Press:
   - *Filter (inverse)*

4. Rename it to something like ‘source_[orig._wav_name]’:
    - *Rename...*
5. To extract the filter, select the original un-resampled sound and make a Formant object:
    - *Analyse spectrum > To Formant (burg)... > 0.01, 6, 6000, 0.025, 50*
    - (Keep time steps at 0.02s again)
6. From this make a FormantGrid object:
    - *Down to FormantGrid*
7. Change F2 track by opening the FormantGrid object and dragging the points, once the 2nd formant row is selected:
    - *View & Edit > [Ctrl+2] > [drag the points]*
    - (It may be useful to group this window with the original sound file, to monitor temporal position.)
8. Once all the points have been moved, select the source and filter together (source_... & FormantGrid objects), and recombine them:
    - *Filter*
9. Open the resulting new object for inspection, and save as a WAV file.

The formants in the resulting stimuli followed the red lines in Figures 4.1 & 4.2, having been altered from their original position, along the yellow lines.



F2 has been lowered from original yellow position to new red position (and F3 raised)

Figure 4.1: Resynthesised formant structure, *hut* words

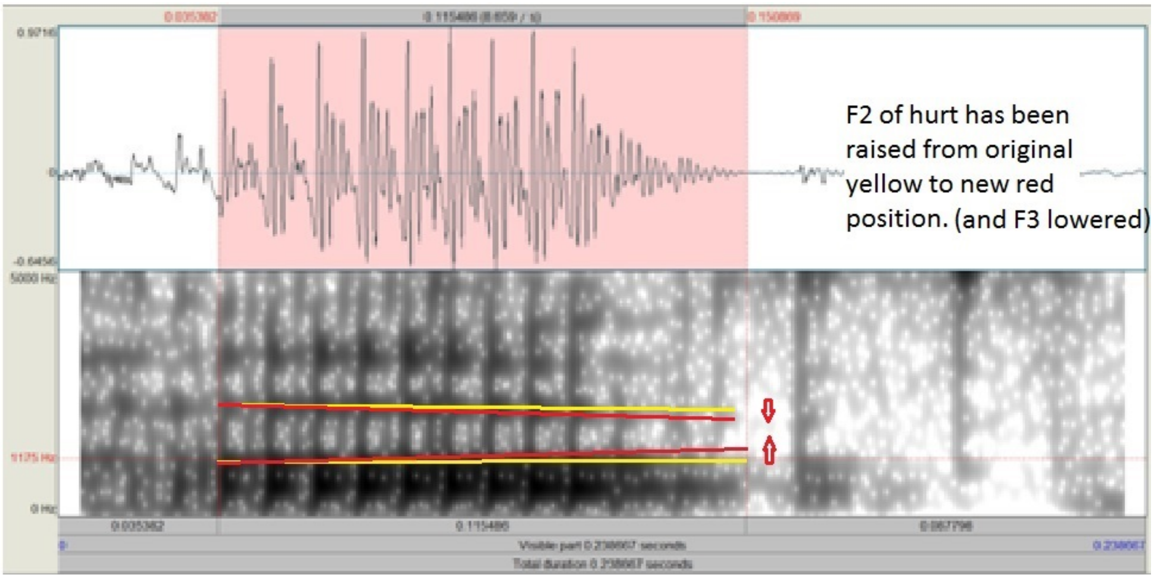

0.035382
0.115486 (8.659 / s)
0.150869
0.9716
0
-0.6456
5000 Hz
1175 Hz
0 Hz
F2 of hurt has been raised from original yellow to new red position. (and F3 lowered)
0.035382
0.115486
0.087798
0
Visible part 0.238667 seconds
0.238667
Total duration 0.238667 seconds

Figure 4.2: Resynthesised formant structure, *hurt* words

Duration changes were done for the Altered condition in order to further ‘neutralise’ the differences between e.g. *hut* and *hurt* words. In Stuart-Smith (2007), the difference between words with and without /r/ (e.g. *hat* and *heart* in that study) was around 17%, and Table 2.3 in Chapter 2 shows similar duration differences between e.g. *hut* and *hurt* words in the stimuli for Experiment 1 in this thesis. The midpoint for neutralising the difference between *hut* and *hurt* words for Experiment 2 was therefore chosen by calculating the difference between the minimal pair counterparts, and making the manipulation accordingly.

For example, the duration of the vocalic portion of the *hurt* stimuli differed from the duration of the vocalic portion of the *hut* stimuli by 12%. The duration of the vocalic portions of the *hut* stimuli was always shorter than the durations of the vocalic portions of the *hurt* stimuli. The ‘midpoint’ of the two stimulus types was thus 6% longer for *hut* and 6% shorter for *hurt*. The durations would therefore be adjusted accordingly, to create the stimuli for the Altered condition.

For changes in duration, Praat can be used to run a script which implements the PSOLA procedure (Valbret, Moulines & Tubach 1992), which allows for automated duration changes for a large number of stimuli. It was decided that PSOLA would not be used, due to the relatively low number of stimuli in this experiment. The duration changes were done by calculating the percentage difference to the midpoint and either copying and pasting one, two, or three full periods on the waveform (whatever was closest to the percentage difference) for lengthening duration, or cutting periods, for shortening duration.

Subsequent signal manipulation for Altered and Natural stimuli

The method of doing Praat's source-filter resynthesis as described above, results in a sound file which is of a lower quality than the original file, such that it acoustically resembles the audio of a compressed internet video call. The quality of the processed sounds in this experiment could not be described as poor – that is, acoustic features of the speech were still very much audible – but there would have been a noticeable, and likely distracting, difference in quality between the processed words and the rest of the passage. Therefore, the whole of the rest of the passage was processed in the same way as the target words, so that the sound quality remained uniform throughout. Furthermore, to avoid an experimental confound, the passage for the Natural condition was also processed in this way (i.e. without making any formant or duration manipulations, but ensuring the sound quality was the same for both listener conditions).

After testing for the experiment had been completed, it was found that by subjecting the sound file to Praat's source-filter resynthesis in a slightly different way, a much better quality file could have been produced. This approach involves separating the sound file into two parts: by low-pass filtering to produce the file's 0-6kHz frequency band, and high-pass filtering to produce the file's 6-12kHz frequency band. These bands are then individually processed as above, then recombined to create the desired stimulus. This method results in much less distortion than the approach used in the present experiment, where the file was processed as a single frequency band. Because this alternative method was discovered after testing, it was too late to change the stimuli for this experiment. However, it will be used in future for any similar manipulations.

Once the target word files were fully manipulated in both formant and duration dimensions, and the remainder of the passage had been subjected to the same filtering, the target words were spliced back into the appropriate places. Transitions always occurred at zero-crossings on the waveform, in order to avoid acoustic clicking artefacts. The resulting sound file was used as the passage for the listeners in the Altered condition.

4.2.4 Experimental procedure

The perceptual experiment was presented on a Lenovo laptop computer, using the perceptual testing software DMDX (Forster & Forster 2003), playing stimuli over Sennheiser HD800 headphones. Participants were tested in sound-attenuated recording booths in Glasgow University Laboratory of Phonetics, and the Phonetics Lab in the Department of Theoretical and Applied Linguistics at the University of

Cambridge. Each participant was welcomed, then asked to sign the consent form and to read the information and instruction sheets. The researcher stayed with each participant for the short practice section at the start of the first task, then they were left alone to complete the rest of the task. This was the same for each of the two 2AFC tasks, Pretest and Posttest.

Task 1: Pretest

DMDX had been prepared before the participant's arrival, and each participant had a unique filename for the output of their results. Furthermore, each participant had a uniquely randomised order of presentation of the trials. Each participant carried out a two-alternative-forced-choice (2AFC) word identification experiment. They heard a stimulus, then were asked to choose (using two labelled keys) which word they thought they had just heard, out of two options that appeared on the computer screen immediately after the stimulus had finished. For example, if they were played a stimulus '*bust*', the two options 'BUST' and 'BURST' would then appear on the left and the right of the screen, and they would choose one by pressing the corresponding key beneath it, labelled 'L' or 'R'. For each repetition, the position of the /VrC/ word on the screen varied pseudo-randomly, 50% on each side, to prevent dominant hand bias: this order was counterbalanced across tokens. Response time was measured from the point at which the visual response options appeared on screen. There was a time-out at 2500ms. After the time-out, the next stimulus was played. In Task 1, 96 randomised stimuli were played to each participant, with breaks after every 20 stimuli: Task 1 lasted approximately 10 minutes. DMDX coded the responses as either 'correct' or 'incorrect', and response times were recorded. Once participants had completed the task they alerted the researcher, who was outside the booth.

Task 2: Exposure passage

Before they were left alone in the booth for Task 2, participants were asked to read the instructions, which described a task in which they had to count how many times an animal is mentioned in the story, and tally them up in the space on the instruction sheet. This was a distractor task, in order to hold their attention so they would be more likely to attend to the phonetic detail of the speech.

Participants listened to one of the two resynthesized versions of the short story, played using VLC player (VideoLAN 2013), as read by the WC Glaswegian speaker. At the end of the passage there was a short beep, indicating to the participants that the story was finished and that they should call the experimenter

Task 3: Posttest

After a short break, the participant then completed another version of the 2AFC, which was of the same design as Task 1 (above), with the same stimuli (different, randomised running order).

Debrief/end of experiment

Participants were asked to fill in a question sheet (Appendix 6), including main places they have lived, how long they spent there, etc. There was also a question about how easy or hard they found the speaker and story to understand. They were debriefed, and at this stage the researcher was happy to answer any questions the participant had. They were paid, and thanked for their participation.

4.3 Results

4.3.1 Statistical analysis

For each of the analyses in this chapter, linear mixed effects models were run, as explained in Experiment 1.

Using the lme4 package in R, saturated linear mixed effects models were created, in order to uncover which of the experimental factors most affected the dependent variable. In each of the analyses, the initial model included the following factors of interest, beginning with the fixed effects, followed by the random effects:

Fixed effects:

Group: Which listener group the participant was in, i.e. ‘Glasgow’, ‘Intermediate’, or ‘Cambridge’.

Condition: Which experimental condition the participant was in, i.e. ‘Altered’ or ‘Natural’.

Coda: Whether the stimulus canonically had an /r/, e.g. whether the word in the minimal pair was *cut* or *curt*.

Test: Whether the stimulus was heard in the Pretest or in the Posttest.

Random effects:

Subject: The effect of participant was included as a random intercept in the model to allow for likely, and potentially large, variation between participants' response behaviour.

Trial: The stimuli were randomised within each of the blocks, and the pattern of randomisation was unique to each participant. Trial was therefore included in the model, again as a random effect. However, it was later determined that trial should not have been included as a random effect in this manner, because the meaning of the term ‘random’ in statistical modelling does not refer to the nature of the stimulus presentation, but to the ability to generalise the results patterns to new instances. Nevertheless the term is still presented here (and again in Chapter 5), to reflect the models which were actually run.

An alpha level of .05 was used for all models. For each analysis below, the following saturated model was run. It included all four fixed effects, as well as all interactions including the 4-way interaction:

$$lmer([dependent\ variable] \sim (group + test + coda + condition)\hat{}\,4 + (1|subject) + (1|trial))$$

In order to remove non-significant effects, lmerTest's step() function was again applied to each model.

4.3.2 Response time

After step() was run on the fully saturated model, the best-fitting model for log(rt) (summary in Table 4.2) was:
$lmer(logrt \sim group + condition + test + coda + groupXcondition + groupXtest + groupXcoda + testXcoda + groupXtestXcoda + (1|subject) + (1|stimulus))$

Table 4.2: Model summary for log(rt) (Experiment 2)

	log(rt)
group_Intermediate	-0.097 (0.064)
group_Glasgow	-0.151^{**} (0.064)
condition_NatExp	-0.157^{***} (0.061)
test_Posttest	0.061^{**} (0.028)
coda_*curt*	0.105^{**} (0.046)
group_Intermediate X condition_NatExp	0.251^{***} (0.085)
group_Glasgow X condition_NatExp	0.196^{**} (0.085)
group_Intermediate X test_Posttest	-0.088^{**} (0.039)
group_Glasgow X test_Posttest	-0.066^{*} (0.036)
group_Intermediate X coda_*curt*	-0.135^{***} (0.038)
group_Glasgow X coda_*curt*	-0.139^{***} (0.036)
test_Posttest X coda_*curt*	-0.184^{***} (0.039)
group_Intermediate X test_Posttest X coda_*curt*	0.123^{**} (0.053)
group_Glasgow X test_Posttest X coda_*curt*	0.132^{***} (0.050)
Constant	6.749^{***} (0.053)
Observations	4,144
Log Likelihood	−1,314.525
Akaike Inf. Crit.	2,665.051
Bayesian Inf. Crit.	2,778.980
Note:	*p<.1; **p<.05; ***p<.01

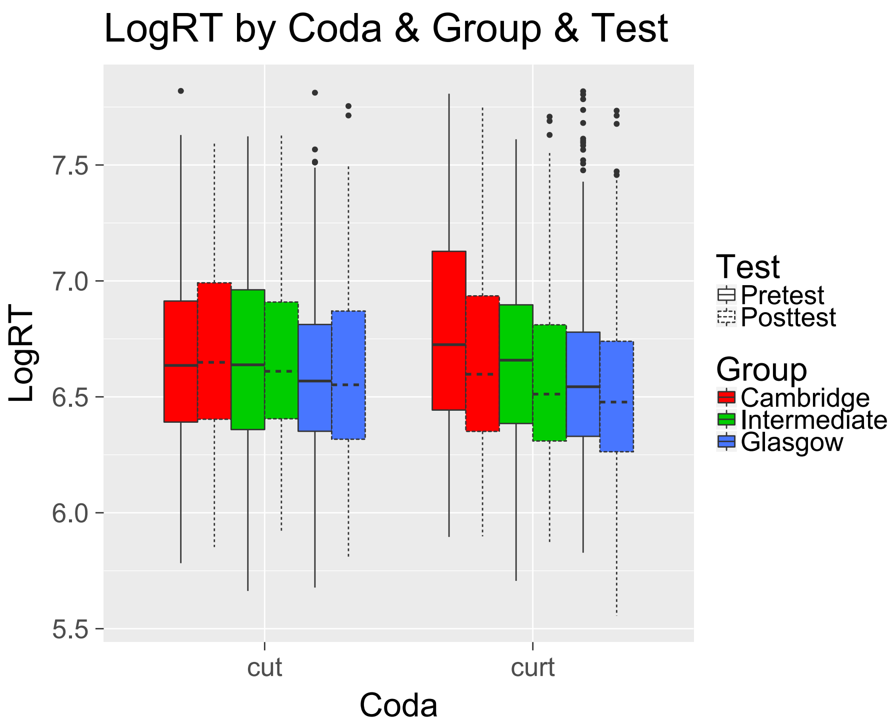


Figure 4.3: Experiment 2 log(rt) for responses to correct stimuli, by Coda, Group, & Test

The log response time (log(rt)) data presented here only plots the 4144 correct responses out of a possible 6131 (68%) across all listeners. Figure 4.3 shows log(rt) by coda, i.e. whether or not the stimulus canonically contained an /r/, then by Group (blue = Glasgow; green = Intermediate; red = Cambridge), then by Test (Pretest = solid line, Posttest = dotted line). This was a significant 3-way interaction (Pr(> F) = .019, F = 3.9744).

There were no significant differences between Pretest & Posttest for any group's *cut* words, but everyone gets significantly faster (Cambridge: p = .001; Intermediate: p = .001; Glasgow: p = .01) for *curt* words. The 'Group' part of the interaction is that Cambridge listeners behave slightly differently from the other two groups for *cut*, in that they get slightly *slower* from Pretest to Posttest (note: this is only a trend, with p = .2), whereas Glasgow and Intermediate listeners seem to get slightly faster (although neither of those differences was significant).

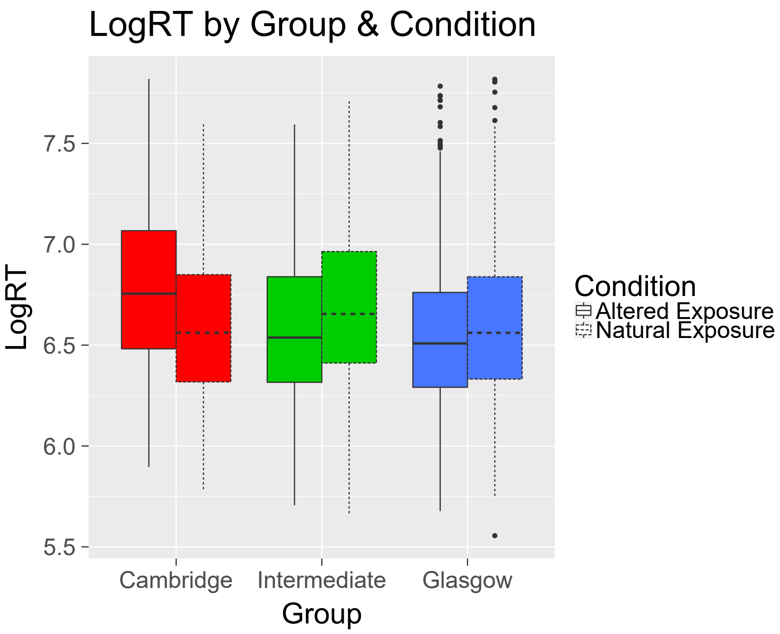


Figure 4.4: Experiment 2 log(rt) for responses to correct stimuli, by Condition & Group

Figure 4.4 shows log(rt) by condition, i.e. whether or not the stimulus was in the Natural or the Altered Exposure condition, then by group (blue = Glasgow; green = Intermediate; red = Cambridge). This was a significant interaction (Pr(> F) = .01, F = 4.7815), showing that for the Altered Exposure condition, Cambridge listeners were slower than both Intermediate (p = .004) and Glasgow listeners (p < .001), but in the Natural Exposure condition there were no group differences. Cambridge listeners also differed between conditions, with those in the Altered Exposure condition responding slower than those in the Natural Exposure condition (p = .011).

4.3.3 Sensitivity

After step() was run on the fully saturated model, the best-fitting model for *d'* (summary in Table 4.3) was:

$lmer(d'group + test + (1|subject))$

Table 4.3: Model summary for *d'* (Experiment 2)

	d'
group_Intermediate	0.536^{***}
	(0.124)
group_Glasgow	1.959^{***}
	(0.125)
test_Posttest	0.123^{**}
	(0.053)
Constant	0.242^{***}
	(0.092)
Observations	256
Log Likelihood	−231.889
Akaike Inf. Crit.	475.778
Bayesian Inf. Crit.	497.049
Note:	$^{*}p<.1$; $^{**}p<.05$; $^{***}p<.01$

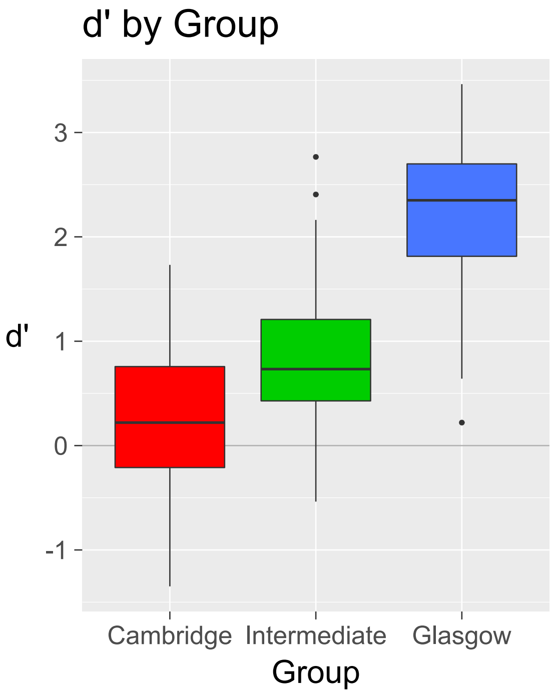


Figure 4.5: Experiment 2 d': Group

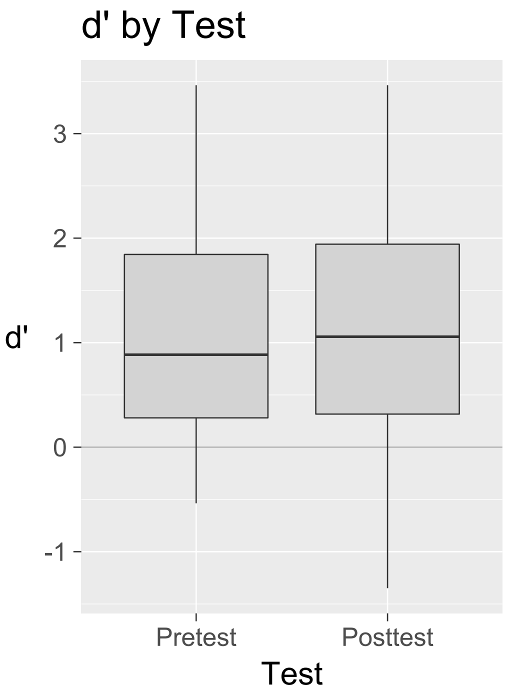


Figure 4.6: Experiment 2 d': Test

Figure 4.5 shows the participants' d' depending on whether they were in Glasgow, Intermediate, or Cambridge listener groups. This was a significant main effect (Pr(>F)<.001, F=131.3917), such that Glasgow listeners were the most sensitive to difference ($d'=2.2620$), Cambridge were the least ($d'=0.3034$), and Intermediate were between the other groups ($d'=0.8397$). All group differences were highly significant ($p<.001$).

Figure 4.6 shows the participants' d' depending on whether the stimulus was heard in Pretest or in Posttest. This was a significant main effect (Pr(>F)=.023, F=5.3012), such that, overall, participants' d' was higher in Posttest ($d'=1.1965$) than in Pretest ($d'=1.0736$).

4.3.4 Response bias

After step() was run on the fully saturated model, the best-fitting model for *c* (summary in Table 4.4) was:

$lmer(c \sim group + test + (1|subject))$

Table 4.4: Model summary for *c* (Experiment 2)

	c
group_Intermediate	-0.333^{***}
	(0.105)
group_Glasgow	-0.267^{**}
	(0.106)
test_Posttest	-0.252^{***}
	(0.043)
Constant	0.117
	(0.078)
Observations	256
Log Likelihood	−182.830
Akaike Inf. Crit.	377.660
Bayesian Inf. Crit.	398.932
Note:	*p<.1; **p<.05; ***p<.01

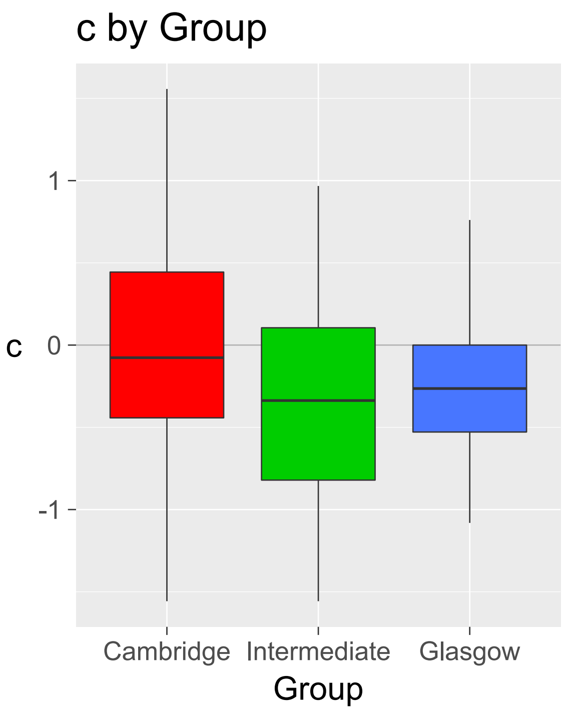


Figure 4.7: Experiment 2 *c*: Group. Positive *c* value indicates bias towards CUT

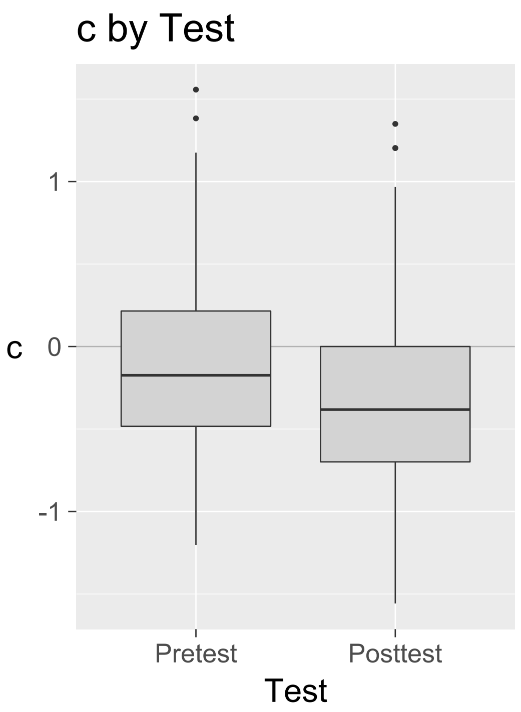


Figure 4.8: Experiment 2 *c*: Test. Positive *c* value indicates bias towards CUT

Figure 4.7 shows the participants' *c* depending on whether they were in Glasgow, Intermediate, or Cambridge listener groups. This was a significant main effect (Pr(> F) = .005, F = 5.6361), such that the Cambridge listeners had an overall lack of bias across conditions and factors, but their *c* was different to both Glasgow (*c* = -0.2757, p = .013), and Intermediate (*c* = -0.3418, p = .002) listener groups, who both showed bias towards reporting e.g. *curt*. There was no difference in bias between Glasgow and Intermediate listeners (p = .528).

Figure 4.8 shows the participants' *c* depending on whether the stimulus was heard in Pretest or in Posttest. This was a significant main effect (Pr(> F) < .001, F = 34.7264), such that overall, participants were more biased towards responding e.g. *curt* in Posttest (*c* = -0.3347) than in Pretest (*c* = -0.0828).

4.4 Discussion

This chapter presented Experiment 2, which followed a perceptual learning paradigm in order to reveal what happens when listeners are exposed to ambiguous phonetic variants over a short period of time. This discussion will assess the extent to which these results address the research question:
'How does experience relate to the learning of ambiguous fine phonetic detail for a phonemic contrast?'

4.4.1 Response time

From the analysis of log response time, there were two significant main effects: Group, and Test. The existence of a main effect of Group is unsurprising, and appears to support the long term familiarity results found in Experiment 1. In short, this main effect indicates that the listener group who were the most familiar with the accent variety represented by the stimuli in this experiment – Glasgow – were the quickest listeners to respond to stimuli across the whole experiment, with the least familiar Cambridge listeners significantly slower than the Glasgow group ($p = .005$). The Intermediate listeners were not *significantly* faster or slower than the other two groups, but the group's average response time was between those of Glasgow and Cambridge, as expected.

This effect does not seem to be as strong as the much clearer pattern of listener experience on response time in Experiment 1. This was not described in the previous chapter (which only described new analyses), but Lennon (2013) did analyse response time for Experiment 1. Although it was not subjected to the same rigorous statistical analysis as in this thesis, the pattern of the effect of listener experience on response time in Experiment 1, namely that the Glaswegians responded to the stimuli with the fastest response times (2013: 20), mirrors the response time result in Experiment 2, described here. The key difference is that the result for Experiment 2 is weaker.

However, direct comparisons between the two sets of results should be made with caution, as there are a number of differences between the two experiments. They were structurally different, with Experiment 1 consisting of two tasks (2-alternative-forced-choice task, /r/ strength rating task), and Experiment 2 consisting of 3 tasks (Pretest, Exposure, Posttest).

There was an even greater amount of difference within each task, as Experiment 2 presented listeners with one working class speaker, whereas Experiment 1 presented listeners with four speakers, across two sociolects: middle class and working class.

This final concern raises the important issue of what happens when a listener hears multiple talkers or accents in a single listening environment, compared to hearing just one. The existence of more than one talker in listening experiments has been attributed to processing costs for the listener in a number of influential studies (e.g. Cole et al. 1974; Mullennix & Pisoni 1990; Goldinger 1996, 1998). The results of these studies all lend support to exemplar theories of speech perception, as they appear to show that the listener processes the phonetic content *and* indexical information about the speaker in an integrated fashion. This issue will be addressed much more directly in the design and discussion of Experiment 3, described in the next chapter.

The other main effect which was significant was Test, and as with the Group effect, the fact that Test is significant is also unsurprising, but for a very different reason. It is likely to be due to improvement because of the order of tasks in the experiment. In other words, listeners hear the stimuli in the Pretest, learning the speaker's voice and gaining proficiency in the procedure as the task progresses, then they hear the same speaker for six minutes in the Exposure task, learning still, then they complete the Posttest, in which they hear the same speaker once again.

There were no significant main effects relating to the factors of Condition or Coda. This may be due to the relative difficulty of the experiment, compared to Experiment 1, which had many more effects due to large differences between the factors. For example, in Experiment 1 there were large acoustic differences between middle class and working class stimuli in the Class factor. In contrast, in Experiment 2 there were very small acoustic differences between stimuli in the Altered and Natural conditions, meaning that much smaller or absent main effects are unsurprising. The interactions between factors were more interesting to examine.

Close examination of the significant factor differences in the interaction of Group by Condition reveals an interesting pattern: the three listener groups were no different to each other in the Natural condition, but the Cambridge listeners in the Altered condition were significantly slower than the other two groups. There may be an explanation for this pattern, if the nature of the speech in the Altered condition is taken into account. First, the listeners hear the unaltered /r/ productions in the Pretest, then they hear /r/ productions with altered F2, F3 and duration (neutralising the differences along these dimensions between the /r/-ful and /r/-less words), and finally they hear the unaltered /r/ productions again in the Posttest. This suggests that the least familiar listeners in Cambridge may have been the most susceptible to phonetic irregularities, and that they might be incurring a processing cost due to a potentially idiosyncratic speaker.

The fact that they are unfamiliar with both the speaker *and* the accent (unlike the other two listener groups, who have at least a few years of experience with the Glaswegian accent in general, though not the speaker), could mean that their perceptual 'model' for this speaker/accent is relatively unstable at this very early stage of their learning of this speaker/accent combination. This interpretation is compatible with exemplar models which suggest that sparse exemplar clouds can give rise to weak representations, because of the comparatively strong influence of new exemplars. It is also consistent with the Bayesian model of speech perception as put forward by Kleinschmidt and colleagues (e.g. Kleinschmidt, Weatherholtz & Jaeger 2018), who assert that listeners build upon their prior beliefs about a speaker or an accent, updating and adapting as they hear more. The fact that the Cambridge listeners have relatively few prior beliefs about Glaswegian /r/ production might mean that they have to incur extra processing, compared to the other listener groups. The unusual acoustic neutralisation of the Altered condition may act as an inhibiting factor which slows their processing speed.

The other significant 2-way interaction was Coda by Group (Figure 4.3), where Cambridge listeners were significantly slower than the other two listener groups when responding to all *curt* stimuli (boxes on the right of the graph), but there was no difference between the groups when responding to *cut* stimuli (boxes on the left). This may just be an artefact of the Cambridge listeners' lack of familiarity with derhoticised /r/, resulting in a longer processing time for the more 'difficult' /r/-ful words. Although it is very vowel-like in its formant structure, derhoticised /r/ has a degree of pharyngealisation or uvularisation, which probably makes it sound more 'unusual' than the /r/-less words. If this were the case, this would very likely be responsible for increasing the perceptual load on the Cambridge listeners when they hear derhoticised /r/ variants.

The three way interaction of Group by Test by Coda was significant ($Pr(>F) = .019$, $F = 3.9744$), and it shows that every listener group gets much faster from Pretest to Posttest for *curt* words (on the right of Figure 4.3). This interaction because the Cambridge listeners seem to behave differently from the other groups. When responding to /r/-less words (the left of Figure 4.3), the Cambridge listeners are slightly faster in Pretest than Posttest, though this difference is only a trend, at $p = .2$. The other two groups both seem to get slightly faster from Pretest to Posttest.

One possibility is that the Cambridge listeners are beginning to mistakenly think that the speaker's *hut* words are in fact *hurt* words. This may be because the context in the Exposure story provided them with information that *it is possible* *that – at least for this speaker – words with high F3 and low F2 can contain an /r/*, and that these 'words' are very similar to the actual *hut* words. In essence, they are learning that '*hut* can be similar to *hurt*', so it seems that they are thinking about their decision for a little longer in the Posttest task. This effect may therefore be responsible for overriding the otherwise universal reduction in response times from Pretest to Posttest.

This could be seen as the Cambridge listeners beginning to treat both the *cut* and *curt* words in a similar fashion, following the Exposure section. Consequently, they might be shifting their perception to be more in line with the longer term experience of the Intermediate group. This could be evidence for the 'seeds of change', as discussed by Ohala (e.g. 1993), in that the Cambridge listeners might be starting to 'perceptually hypercorrect' even after gaining such a little amount of exposure. A closer look at the results for both sensitivity and response bias may shed further light on this.

4.4.2 Sensitivity

The main effect of Group was expected, and the pattern of the Glaswegian listeners being by far the most sensitive to difference between *cut* and *curt* words unsurprisingly replicates the long-term experience pattern seen in Experiment 1, as described in Chapter 3. This is likely to be due to many of the same issues as described for the main effect of Group in the response time section.

The other significant main effect, Test, was also expected, as it is not surprising for listeners to improve their sensitivity to differences between stimuli from Pretest to Posttest, if the effect of 'task learning' is taken into account. However, a closer look at the two marginal interactions may shed further light onto this pattern.

The two-way interaction of Condition by Test is marginally significant at $p = .0512$, showing that there was in fact almost no Pretest to Posttest improvement in sensitivity for all listeners in the Altered exposure condition, but there was a relatively large Pretest to Posttest improvement in sensitivity for listeners in the Natural exposure condition. This suggests that the Natural exposure condition (which retains the F2, F3, & duration differences between *hut* and *hurt* stimuli), aids the improvement in sensitivity, whereas the Altered condition (in which the F2, F3, & duration differences were neutralised) does not.

In order to investigate this further, we can inspect the differences between the groups, as shown in Figure 4.9. There was no significant 3 way interaction, but the trends represented in the graph are interesting to examine.

The near-significance of the Group by Test two-way interaction may come from the fact that the Glasgow listeners in the Natural exposure condition (blue boxes

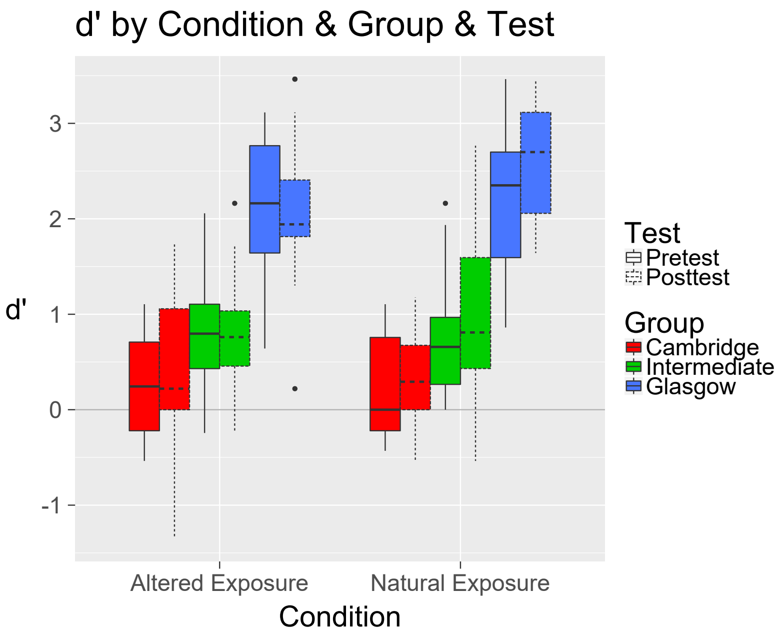


Figure 4.9: Sensitivity *d'* to stimulus pairs by condition, group and test

on the right) seem to be the only group to improve their sensitivity, however the Intermediate listeners in the Natural exposure condition also seem to notably improve. The Glasgow and Intermediate groups in the Altered exposure condition do not appear to improve at all.

A plausible explanation for this pattern may be that the more experience that listeners have of the Glaswegian accent in general, the more they benefit from increased exposure to one particular speaker. If we look again at Figure 4.9, we can see that the Intermediate listeners in the Natural exposure condition (green boxes on the right) improve from Pretest to Posttest, but the Glasgow listeners (blue boxes) look as if they improve even more – even though their sensitivity was already very high. In contrast, the Cambridge listeners (red boxes) hardly seem to improve at all, in either Altered or Natural exposure conditions, *even though their initially low sensitivity had a lot of room to improve*. This seems like a clear benefit for increased exposure to an accent when learning about a speaker.

When placing these results in the context of the wider literature, these findings, for short-term learning of a subtle fine grained phonetic detail, look consistent with exemplar-based learning. Overall, sensitivity to stimulus difference does not improve a great deal in this relatively short experiment, but the small differences that do appear are interesting. Listeners improve their sensitivity to differences between words, but only when the speaker is internally-consistent. That is, in the Altered exposure condition, the speaker had a different pattern of formant structures in his *hut* and *hurt* words in the Exposure story than in the Pretest and

Posttest. In the Natural exposure condition, the speaker had very similar formant structures in all three sections of the experiment.

Nevertheless, perhaps the pattern described above indicates a more nuanced form of learning than simply a broad exemplar account of fully integrated speaker-and-accent learning. If it were the case that listeners were learning the speaker and accent *together*, one might expect there to be improvement in sensitivity in both the Natural exposure condition *and* the Altered exposure condition, because even in the Altered condition, it is likely that listeners would still have the opportunity to learn something about the speaker. This was not the case, so it appears that the imbalance in improvement between conditions may show that there is indeed a benefit for long-term exposure to the accent, when processing a new speaker. That is, Glasgow listeners have a large amount of experience of Glaswegian speech, and this seems to be a good foundation on which to build their learning of the speaker. The Intermediate listeners have much less experience (around three years), but still enough to help them learn the speaker. In contrast, the Cambridge listeners have almost no experience of working class Glaswegian in general, so they are likely having to work hard to learn both the speaker *and* the accent at the same time.

4.4.3 Response bias

As with sensitivity, response bias had significant main effects of Group and Test, though differences in bias are a little harder to interpret than differences in sensitivity. Overall, there was a slight bias towards responding *curt*, at $c = \text{-}0.209$, but this was due to a number of factors which will now be looked at.

All listener groups were biased overall towards responding *curt*, that is, towards reporting that they had heard an /r/-ful word when choosing between pairs. The Intermediate and Glasgow listeners showed significantly more bias towards responding *curt* than the Cambridge listeners, which caused the significant main effect of Group. This is difficult to interpret, but the large spread of responses in the Cambridge listeners (red box in Figure 4.7) can be more closely examined in Figure 4.10. There was no significant three-way interaction in this graph, which shows the response bias data broken down by exposure condition, listener group, and test, so the patterns cannot be interpreted as particularly meaningful, but the trends are interesting.

The Cambridge listeners initially showed a bias towards reporting that they heard /r/-less words, as their bias was positive. This was predicted to be the case, as this was the pattern in Experiment 1: those listeners with little experience of the Glaswegian accent would unsurprisingly report hearing most words with

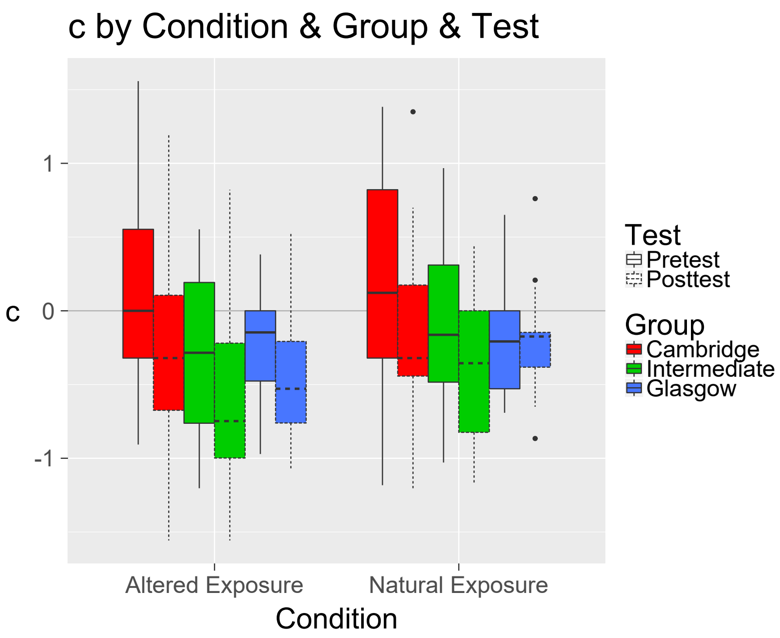


Figure 4.10: Response bias *c* by condition, group and test. Positive values of *c* indicate a bias towards responding CUT.

flat, vowel-like formant structures as words with vowels. Interestingly though, the Cambridge listeners change their bias towards reporting /r/-ful words in the Posttest task, which is in line with the long-term pattern for Intermediate listeners hearing /ʌ/ words, as seen in Figure 3.7 in the previous chapter (green boxes with dotted lines, on the right of the graph).

Although this data is a trend, the change is evidence that once the Cambridge listeners have heard the Exposure story, they then begin to ‘perceptually hypercorrect’, therefore showing the same long-term pattern of perception as the Intermediate listeners, but only after a very small amount of exposure. This could be taken as evidence for rapid adaptation to an unfamiliar dialect, as has been found in other studies.

The other significant main effect was Test (Figure 4.8), which showed an overall inclination for listeners to report hearing a *curt* stimulus in Posttest, *after* they had heard the exposure story. An explanation for this could be that once the listeners have heard the speaker in the Exposure story, they have learned that he uses a highly vocalic variant (with F2 & F3 far apart), when the context strongly indicates that he intended to produce a canonical /r/. For example, when he says ‘...he could feel his skin beginning to burn...’, or ‘...he didn't want to cause it any hurt’, this teaches the listeners that, when they hear a word which has a low F2 and high F3, it is very likely that the speaker is intending to produce a word with an /r/. This knowledge simply does not exist in the Pretest, due to lack of experience of the context.

A trend seen in Figure 4.10 is that in both conditions there appears to be a swing towards reporting *curt* in Posttest, which could indicate that the listeners are learning from the context about the potential for the speaker to produce canonical /r/. However, the most surprising pattern is that of the Glasgow listeners in the Altered exposure condition (blue boxes on the left), which do seem to show a swing in bias towards responding *curt*, in a similar fashion to the Cambridge and Intermediate listener groups. A possible explanation may be that the Glasgow listeners are being confused by the difference between the fact that the speaker's *hut/hurt, cut/curt* productions change from Pretest to Exposure, then again from Exposure to Posttest. They may have therefore had to “unlearn” what they previously knew about derhoticised /r/ for this speaker, treating him as an idiosyncratic speaker, meaning they have a comparable amount of experience with the speaker as the other listener groups. A Bayesian interpretation for this may be that these Glasgow listeners are being forced to update their prior beliefs with new information.

4.4.4 Summary

This chapter described Experiment 2, which was a perceptual learning experiment with a Pretest, Exposure and Posttest design. It tested listeners on their ability to learn the fine phonetic detail of Glaswegian derhoticised /r/, making use of two listening conditions to do so. The level of experience that a listener had with working class Glaswegian was controlled by having three listener groups, in the same way as in Experiment 1.

The research question for this experiment was: ‘How does experience relate to the learning of ambiguous fine phonetic detail for a contrast?’. The clearest result from Experiment 2 is that the listeners from Glasgow were overall the best-performing group, which supports the overall pattern of long term learning seen in Experiment 1. The Intermediate and Cambridge listeners, who had much less experience with Glaswegian, showed mostly predictable patterns in response time, sensitivity, and response bias, with some intriguing patterns emerging.

The Cambridge listeners appeared to be showing the beginnings of perceptual change in their rapid swing in response bias, to match the long-term experience pattern of the Intermediate listeners.

Chapter 5

Experiment 3: The dynamics of discrimination

5.1 Introduction

From the acoustic analysis in Chapter 2, it is known that there is a statistically significant difference between the formants in working class *hut*-type words and *hurt*-type words from very early in the vocalic portion, but it is not yet known how the differences between these formant structures translate to what is heard by the listener, or indeed whether this comparison can be made directly; i.e. whether the acoustic signal maps directly onto perception. It is also clear from the results of the previous experiments in this thesis that there is a degree of competition between alternatives, in other words, a degree of attraction to the incorrect competitors. However, the results of the previous experiments cannot give information about the fine detail of the strength of the competition effect while listeners are hearing the stimuli, then make their choice.

Experiments 1 and 2 showed that Glaswegian listeners are the most efficient of all three of the listener groups at making the perceptual distinction between working class *hut/hurt* minimal pairs, responding in the two experiments to the working class *hut/hurt* stimuli with an overall accuracy of 89% and 90% respectively. Therefore, the influence of competitor attraction on /r/ perception by Glaswegian listeners was investigated in this experiment. Another reason for testing only Glaswegian listeners in this experiment was that it was deemed impractical to include too many factors in what was a new methodology to the experimenter, and the relative lack of time and project funding compared to previous experiments was an issue which helped to cement this decision.

The next experiment in the thesis follows on from the previous two, both of which investigated familiarity and learning effects (both long-term and short-term) in the perception of derhoticised /r/ in working class Glaswegian. Each of them used natural speech stimuli, presenting them to listeners who were asked to complete a 2-alternative-forced-choice task, either with or without being exposed to a read passage. In short, they both investigated the amount of exposure to the Glaswegian accent a person needs in order to accurately perceive the distinction between e.g. *hut* and *hurt*. In both experiments, a positive long-term learning effect was found, however in the short-term experiment the learning effect was quite small.

While these first two experiments show that perception of derhoticised /r/ varies both between listener groups and with the amount of exposure, this experiment investigates the degree of similarity between the minimal pairs, which are already known to be confusable to all but the most familiar listeners. The experiment described in this chapter provides a deeper understanding of how difficult the discrimination is, by analysing mouse tracking data from a 2-alternative-forced-choice design to quantify this perceptual similarity. This experiment presents listeners with the challenging working class words, as well as the 'easier' middle class words, as a comparison.

The predictions for this experiment are that the listeners would find middle class stimuli the easiest to perceive, especially the strongly-rhotic *hurt*-type words. The most challenging words to perceive are predicted to be the working class *hurt* words, as they are the most similar to the *hut* words.

In order to address the role of challenging listening conditions – the final research question of this thesis – it was decided that this experiment would present listeners with both working class and middle class stimuli randomised together in the same listening task. It was predicted that this would increase the difficulty of perceiving each stimulus, because of the fact that the accent would be unknown before the start of each upcoming word. This motivated a blocked design, with middle class and working class single-talker blocks presented first, and the mixed-talker block presented last. The single-talker blocks were included so that the relative difficulty of perceiving each stimulus in the mixed-talker block (i.e. 'challenging listening conditions') could be compared against a baseline.

5.1.1 Mouse tracking

Because of the wish to analyse the timecourse of listener perception, mouse tracking was chosen. Eye tracking had been considered, as it is an established methodology in this field (e.g. Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy 1995; McQueen & Viebahn 2007; Huettig & McQueen 2007; Huettig, Rommers & Meyer 2011; Koops, Gentry & Pantos 2008; Dahan, Drucker & Scarborough 2008; Salverda

& Tanenhaus 2010; McMurray, Clayards, Tanenhaus & Aslin 2008; Creel, Aslin & Tanenhaus 2008; Creel & Tumlin 2011; Beddor, McGowan, Boland, Coetzee & Brasher, 2013; Robertson 2015). However, eye tracking was quickly deemed impractical for this thesis due to cost and lack of expertise.

The established analysis methods of analysing response time and percentage of correct responses (and the somewhat related Signal Detection Analysis, in Chapters 3 and 4) can reveal much about a participant's behaviour when hearing audio stimuli in an experimental setting, including the perceptual load and the degree to which stimuli are confusable. Such analyses are a good way to quantify the difficulty experienced when processing a particular stimulus, and this data is collected by analysing participants' eventual responses, usually after they hear the stimulus.

However, it is more difficult to use these analysis techniques when the researcher wishes to uncover more detail about ‘online’ perception, that is, how a listener processes a stimulus as it is being heard. A methodology such as mouse tracking allows this detail to be analysed, because – in a similar way to eye tracking – it records the precise movements of the device (the mouse cursor) which is used to make the final response, *before* that response is made. The participant's ‘journey’ towards making their decision can therefore be analysed.

After considering the use of software such as R to run the experiment, the software MouseTracker (Freeman and Ambady, 2010) was decided upon. This was because of the relative ease of setting up experiments using MouseTracker, along with its flexibility in terms of data output. Despite the fact that the use of mouse tracking is relatively new to speech perception studies, there have been a few high-quality studies using it. Some of these studies have used the MouseTracker program which is used for this chapter, including Barca & Pezzulo (2012), who examined Italian listeners' perception of Italian words and non-words, and Dimopoulou (2014), who studied the role of reinforcement when training native Greek listeners to perceive the dental-retroflex phonetic contrast of Hindi.

One of the first studies to use mouse tracking for phonetic research was Spivey, Grosjean & Knoblich (2005), who used the methodology to measure the effect on spoken-word recognition of parallel activation of lexical alternatives. In comparing the method with eye-tracking, the authors write that a disadvantage of eye-movement evidence for parallel activation of alternatives is that it relies on measuring differences between averaged ‘categorical’ data – i.e. steady eye-gaze fixations to an object over time – to produce ‘continuous’ functions. In contrast, mouse tracking takes advantage of the fact that many arm movements are nonballistic (unlike most eye movements), and therefore can be recorded as continuous measures in order to ‘observe graded effects of a competing object pulling the movement in its direction' (2005: 10393). They note that recorded cursor trajectories serve as a 'record of the mental trajectory traversed as a result of the continuously updated interpretation of the linguistic input' (2005: 10398).

Spivey et al.'s paper clearly sets out the procedure involved in mouse tracking, for a 2AFC task between words with similar onsets. Thus, it has been cited many times in the methodologies of subsequent mouse tracking literature. In their analysis, they used a method which measured the area under the cursor trajectories in order to quantify attraction to linguistic competitors (2005: 10395), however they did not specifically use the terminology 'Area under the Curve', as many later studies (including the present research) have done.

One such study is Sulpizio, Fasoli, Maass, Paladino, Vespignani, Eyssel & Bentler (2015), who looked at the relationship between the characterisation of listeners' judgements of a speaker's sexual orientation in one language, and those of another language (Italian and German). Area Under Curve (AUC) measurements showed that there was a general bias to reporting that speakers were heterosexual, despite actual sexual orientation, in both Italian and German. They noted that, from a methodological standpoint, similar results were obtained whether the experiment was mouse tracking (analysing 2AFC bias using trajectories), or a rating task (a Likert scale was used to obtain degrees of sexual orientation judgments), 'suggesting that the type of judgement does not modify the perception of speakers' [sexual orientation]' (2015: 22). The implication here is that, for this study at least, the use of mouse tracking is warranted as a research method for listener judgements of phonetic features in speech.

Farmer, Anderson & Spivey (2007) analysed the effects of attractor items on the perception of 'garden-path' sentences. Mouse tracking results converged with previous eye tracking results, finding that visual context constrains the resolution of syntactic ambiguity in the visual-world paradigm. Their results 'tie in nicely with converging evidence for a close-knit relationship between language processing, visual perception, and motor action' (2007: 592).

In a study which further argues the case for employing the methodology to answer a range of research questions, Farmer, Liu, Mehta & Zevin (2009) investigated the manner in which Italian natives – who were late learners of English – perceived English vowels, in e.g. *pin/pen/pan* words. Farmer et al. discuss the merits of mouse-tracking when investigating a participant's mouse curvature when moving towards their chosen response, which allows the observation of 'confusability on the dynamics of response execution itself' (2009: 2589). In their justification for using the mouse-tracking methodology instead of eye-tracking, the authors write that while curvature can occasionally be seen in individual saccadic eye movements (Doyle & Walker 2001), individual arm and hand movements can show much more dramatic curvature (Tipper, Howard, & Jackson, 1997) which can be interpreted as the ‘dynamic blending of two mutually exclusive motor commands (Cisek & Kalaska 2005)’ (2009: 2589). They go further, stating that mouse-tracking can yield many more data points per second (30-60) than eye-tracking (2-3 saccades), and therefore mouse-tracking data can reveal spatiotemporal dynamics of the listener's categorisation process itself, not only the final, ‘offline’, result of such a process (Farmer et al. 2009: 2589).

However, Franco and Johnson (2011) compared eye-tracking and mouse-tracking methods, finding that for their particular study using the *decision moving window* paradigm, involving movement of either the mouse or the eyes to uncover certain hidden data on a screen, eye-tracking may be ‘a more natural interface’, allowing ‘freedom from the psychological tether of the mouse’ (2011: 747). Nevertheless they did comment that the reason behind tracking methods is to allow researchers to explain the ‘how’ (the ‘process’) of decisions, not just the ‘what’ (the ‘outcome’). Previously researchers had to use the outcome to infer the ‘how’ (2011: 740).

A much earlier paper, Lohse & Johnson (1996), claimed that methods like mouse tracking ‘increase the amount of effort needed to acquire information’, however subjects manage this extra load by adopting strategies to acquire information in a systematic way, and they write that ‘these strategies tend to be more rigorous and systematic than those observed with eye-tracking equipment’ (1996: 96).

Mouse tracking, then, seems to offer a novel, accessible alternative to eye tracking, and was adopted here. No comparison with eye tracking has yet been made for the perception of contrast between Glaswegian minimal pairs, and would be most interesting.

At the analysis stage, in order to quantify this ‘mental trajectory’, a measure of spatial attraction – Area under the Curve (AUC) – was employed for this experiment (following Freeman and Ambady, 2010). Using this approach, if an effect of the ‘competing object’ is observed, i.e. the mouse moves in the direction of the incorrect competitor, this can be taken to mean a degree of perceptual similarity exists between words in a minimal pair.

However, while it is informative, AUC cannot shed light on the dynamic properties of the trajectories, although it can be seen as a step in the right direction when using the experimental method of mouse tracking. In order to more fully exploit the capabilities of mouse tracking, an additional analysis technique was used: Discrete Cosine Transformation (DCT). DCT describes a trajectory in terms of a set of coefficients, each relating to a different property of the trajectory, and (depending upon the level of complexity of the analysis) these properties can inform the researcher about, for example, the perceived similarity of a stimulus to another, or the level of indecision a listener has when making a particular choice. All of this allows for a deeper understanding of the perceptual journey taken by a listener when processing the stimuli, which is why mouse tracking was chosen as the method for this experiment.

5.2 Experiment 3

5.2.1 Design

Like Experiments 1 & 2, this experiment used a two-alternative-forced-choice (2AFC) paradigm. However, this time the stimuli were presented in three blocks, comprising three separate experimental tasks: the motivation for this was described in the introduction to this chapter. To ensure balance, the presentation order of the Single talker blocks alternated by participant, meaning participant ‘01’ heard MC, then WC, then Mixed, and participant ‘02’ heard WC, then MC, then Mixed.

5.2.2 Participants

51 participants were tested in this experiment (29 female; age 18-34, mean 22), recruited using the same subject pool of Glaswegian listeners as for Experiments 1 and 2, as well as posters around the University of Glasgow. There were two separate presentation orders in the experiment, with 26 participants hearing MC, then WC, then Mixed blocks, and the remaining 25 hearing WC, then MC, then Mixed blocks.

5.2.3 Materials

Materials for Experiment 3 are listed in Table 5.1. The experimental target words were the same set of 24 /CV(r)C/ words used in Experiment 2. The 24 distractors were a subset of the distractors used in Experiment 2 (those with the /ɔ/-/o/ and /i/-/e/ vowel contrasts).

The working class recordings from Experiment 2 were re-used (recall that they were from a 28 year old male speaker from Maryhill, a working class area of Glasgow). New recordings of middle class speech were required. As the two speakers used in Experiment 1 had moved away, another male speaker was recruited (a 22 year old male speaker from Bearsden, the same suburb of Glasgow where the middle class speakers from Experiment 1 were raised).

Target words	Distractor words
bud bird	*beak bake*
bun burn	*beat bait*
bust burst	*beast baste*
cud curd	*con cone*
cuss curse	*cop cope*
cut curt	*cot coat*
fussed first	*dot dote*
hut hurt	*meek make*
shut shirt	*mop mope*
spun spurn	*not note*
thud third	*seem same*
tonne turn	*sneak snake*

Table 5.1: Minimal pairs used in the experiment

The new middle class speaker was recorded under the same conditions as for the speakers in the previous experiments. He was recorded in the sound-attenuated booth at Glasgow University's English Language & Linguistics department, using a lightweight Beyerdynamic TG H74c Condenser headset microphone, at a sampling rate of 44.1kHz.

After recording and excising the words from the wordlist, stimuli were edited so that each word was preceded by a 500ms silence.

5.2.4 Procedure

Instructions and equipment setup

Each participant was tested individually in the sound-attenuated booth (the same booth used to record the speakers) at the English Language & Linguistics department in the University of Glasgow. They were asked to sign a consent form, and to read an instruction sheet.

The instruction sheet stated that there would be three blocks, that recordings of words would be heard over the headphones, and that their task each time would be to choose which of the two displayed words they thought they heard, using the mouse.

It then detailed the following steps, which were the same for each task:

- There will be a short practise session before the start of each task.
- On the top-left and top-right of the screen will be 2 words. Take a second to familiarise yourself with the location of each word!
- After a second, a START button will appear at the bottom of the screen.

- When you click on the START button, one of the words will start to play over your headphones. At the same time, you should immediately start to move the mouse, and click on the word you heard.
- Make your choice as quickly as you can! After 2 seconds the program will move to the next word. (Don't worry if you miss an item: the program will continue)
- After you have heard 25 words there will be a break: when you're ready, press Enter to continue to the next 25.
- Please ask the researcher if there is anything you would like to have explained further.
- When you have finished, please call the researcher.

The screen was positioned level with the participants' eyes, and 1 centimetre on the screen corresponded to approximately 1° of visual arc (cf. McQueen & Viebahn 2007). The experimenter then verbally reiterated that on each trial participants should slowly but immediately begin to move upwards after clicking the START button. Their attention was also drawn to a small black arrow pointing upwards, placed directly above the START button (in piloting it was found that a small visual prompt helped the participants to remember to move upwards), but they were told not to worry about directly following it, and simply to treat it as a prompt or reminder to begin moving upwards. They were also informed that the experimenter would return to the booth between each task to start the next one, and that they would stay with them during each practice session, which was done before each of the three tasks. They would then be left alone for each of the tasks.

Once the participant was happy that they understood the procedure of the experiment, the experiment was started.

Experimental blocks

Each participant completed the experiment in the following order:

1. Practice session
2. First Single talker block (counterbalanced MC or WC): 16 practice trials then 48 trials, freely randomised per participant
3. Second Single talker block: 16 practice trials then 48 trials, freely randomised per participant
4. Mixed talker block: 16 practice trials (half MC half WC) then 96 trials, freely randomised per participant

In all blocks, there was a break after every 24 trials.

Task procedure

The timings and other details of stimulus presentation and response elicitation were determined by informal piloting with eight participants.

At the start of each trial, a white screen was displayed with the two response options (e.g. *hut* and *hurt*) at the top left and top right, in black boxes (see Figure 5.1). After a two-second delay, which allowed the participant to become familiar with the response options' relative positions, a grey 'START' button appeared at the bottom centre of the screen. As soon as the participant clicked that button, the soundfile was triggered to begin playing. After clicking, the participant would then begin to move the mouse upwards *before* they began to hear the word. If they did not start moving within 500ms (i.e. during the silence at the start of the soundfile), a dialogue box would appear with the message 'Please start moving immediately, even if you are unsure of a response yet!', and that trial would be discounted. For trials that were not discounted, the participant would then have a total of 2500ms to move their cursor to their chosen response button and click on it.

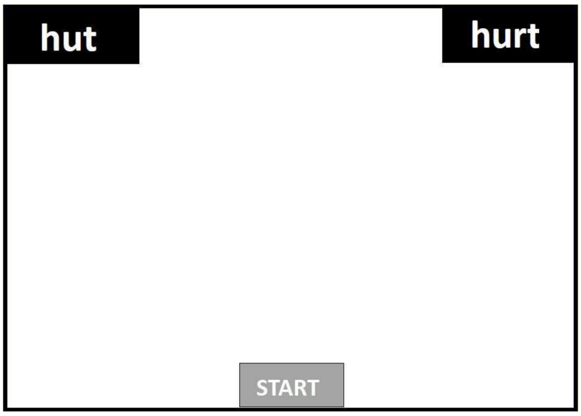


Figure 5.1: Example of decision space for each trial in MouseTracker

The practice sessions before each block were quite long, to allow the participant to practise following the task's instructions, as piloting indicated that the task was fairly difficult. For example, on a couple of occasions during the practice sessions participants clicked the start button then immediately moved the cursor straight upwards, hitting the top edge of the screen, then waited for the word to finish playing fully before initiating horizontal movement towards their chosen response. This type of movement is difficult to analyse, so when it happened, the experimenter could advise the participant to move more gradually and smoothly, in order that they generally followed an arc (approximately) across the screen. This task difficulty was the reason for all practice sessions being monitored by the experimenter.

The practice session before the start of each Single talker block consisted of 16 stimuli from the upcoming block (middle class or working class), randomised by participant, including 10 distractors and 6 target words. The first 4 of these words were always the same distractor words (*beak*, *bake*, *con*, and *cone*), so that the participant always started by hearing some unambiguous words, then heard examples of some potentially ambiguous words. This ensured that as the initial task learning was taking place during the very first few trials, there was no interference as yet from potentially difficult phonetic information. The case could be made that no target words should have appeared in the practice sessions, but it was decided that a more representative sample of the upcoming block may reduce the possibility of novel and potentially surprising information (namely, the more challenging working class *hut/hurt* tokens) causing extra perceptual difficulty.

The practise session before the Mixed talker block had the same 16 words, but half of them were middle class stimuli and half were working class. Furthermore, all 16 were randomised from the beginning, as there was no need to allow for the participant to learn the trial procedure by this stage of the experiment.

Once each practice session was complete, the experimenter started the experimental task then left the booth. Each of the two Single talker blocks had 48 automatically randomised trials, each of which began automatically after a two-second gap following the previous response. As in the practice sessions, correct responses were left/right counterbalanced. Within each Single talker block there was a break after the first 24 trials, with an onscreen display informing the participants that they should press Enter when ready to restart. Once Enter was pressed, the task continued with the final 24 trials. The final block, Mixed talker, consisted of all the previous stimuli from both Single talker blocks, fully randomised. This meant that the participants would hear both the middle class and working class talkers' stimuli presented together in the same task, and there was a total of 96 trials in this final block. Again, there was a break after every 24 stimuli, meaning each participant had three breaks in the Mixed talker block.

At the end of the experiment the participants were asked to fill in a questionnaire, which asked them their place of birth, where they spent the majority of time living, in which areas of the Glasgow conurbation they have lived, and where their parents/guardians grew up. This was to gain a more detailed picture of their accent experience, to ensure that they had not lived too far away from areas where they would be likely to hear Glaswegian accents. The experiment lasted no longer than 30 minutes per participant.

5.3 Results

While the results section for Experiment 3 will deal with the analysis of mouse cursor trajectories, a response time analysis will be presented first, afforded by the fact that, despite its differences to a button-pressing task, the methodology of this mouse tracking study is in essence a 2AFC task – the same as the previous experiments. These simple analyses will allow for a degree of comparison between the Glaswegian listener groups in Experiments 1, 2, & 3. Of course, they are not directly comparable: although each experiment had very similar WC stimuli, they were presented in a slightly different manner in all three experiments. Nonetheless, any comparison will still be worthwhile, after such caveats are applied.

In order to improve the overall picture of the listeners' responses to this experiment even further, signal detection analysis could have been run here, as the responses are again given in a 2AFC paradigm, similar to Experiments 1 & 2. However this analysis was not done, as it was felt that the new methodology of Mouse Tracking would benefit from different analyses being run this time, and that there was much less information that could be gained from running another Signal Detection Analysis.

5.3.1 Statistical models

As in the previous chapter, all the analyses in this experiment were subjected to linear mixed effects models using the lme4 package in R. When building each of the models, a fully saturated model was constructed, including all of the experimental factors. Non-significant effects and interactions were then eliminated. Once again, as in the previous experiments, this approach was taken because of the relatively large number of factors, and in order to account for any potentially complex and unforeseen interactions, it was decided that a more data-driven approach to the modelling was the wisest course of action.

The initial model included the following factors of interest, beginning with the fixed effects, followed by the random effects:

Fixed effects:

Coda: Whether the stimulus canonically had an /r/, e.g. whether the word in the minimal pair was *hut* or *hurt*. This factor contained the levels ‘u’ and ‘r’.

Class: Whether the stimulus was produced by the middle class speaker or the working class speaker. The levels were ‘mc’ and ‘wc’.

Blocktype: Whether the stimulus appeared in one of the Single talker blocks, or the Mixed talker block.

Presentation: Half of the participants heard the MC Single talker block first, followed by the WC Single talker block, then finished with the Mixed block. The other half heard the WC Single talker block first, followed by the MC Single talker block, then the Mixed talker block. This was designed into the experiment purely to counterbalance the participants' experience of MC and WC stimuli, but since the order of presentation could conceivably have an effect on the participants' performance, it was initially included in the models. The levels were ‘mw’ and ‘wm’, representing the different presentation order of the blocks.

Random effects:

Subject: The effect of participant was included as a random factor in the model to allow for likely, and potentially large, variation between participants' response behaviour.

Trial: The stimuli were randomised within each of the blocks, and the pattern of randomisation was unique to each participant. Trial was therefore included in the model, again as a random effect. However, as with Experiment 2 (see the equivalent section of Chapter 4), trial should not have been included as a random effect, as this does not reflect the correct usage of the term ‘random’ in statistical modelling.

Target word: Finally, the target word was included, because each of the natural stimuli will have had many acoustic and durational differences. For example, tokens of *bud* generally have much shorter durations than tokens of *first*.

An alpha level of .05 was used for all models. For each analysis below, the following saturated model was run. It included all four fixed effects, as well as all interactions including the 4-way interaction:

$$lmer([dependent\ variable] \sim (coda + class + blocktype + presentation)\hat{}4+ \\ (1|subject) + (1|target\_word) + (1|trial))$$

In order to remove non-significant effects, lmerTest's step() function was applied again.

The fixed effect of **Presentation** will not be discussed further, as it failed to approach significance as a main effect in every one of the models, nor did it contribute to any interactions. The random effect of **Trial** will also be discussed no further, because it was also eliminated from every model by step().

5.3.2 Response time

After step() was run on the fully saturated model, the best-fitting model for log(RT) (summary in Table 5.2) was:

$lmer(logrt \sim coda + class + blocktype + coda{:}class + class{:}blocktype + (1|subject) + (1|target\_word))$

Table 5.2: Model summary for log(RT) (Experiment 3)

	log(RT)
coda_*hurt*	-0.019
	(0.015)
blocktype_Mixed	0.029^{***}
	(0.005)
class_WC	0.082^{***}
	(0.006)
coda_*hurt* X class_WC	0.039^{***}
	(0.007)
blocktype_Mixed X class_WC	-0.030^{***}
	(0.007)
Constant	7.346^{***}
	(0.017)
Observations	4,486
Log Likelihood	3,377.012
Akaike Inf. Crit.	$-6,736.023$
Bayesian Inf. Crit.	$-6,678.345$
Note:	*p<.1; **p<.05; ***p<.01

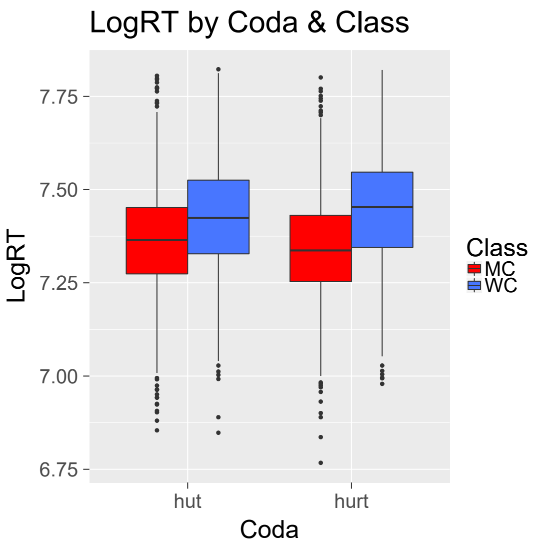


Figure 5.2: Experiment 3 log(rt) for responses to correct stimuli, by Coda & Class

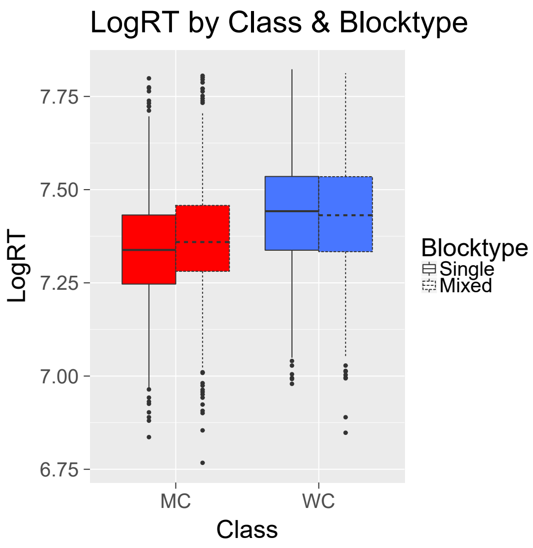


Figure 5.3: Experiment 3 log(rt) for responses to correct stimuli, by Class & Blocktype

Like the previous experiments, the log-transformed response time (log(rt)) data presented here only plots the correct responses. Figure 5.2 shows log(rt) by coda, i.e. whether or not the stimulus had an /r/, then by class (red = MC; blue = WC). This was a significant interaction (Pr(>F) < .001, F = 35.0251), showing that responses to MC *hut* stimuli were faster than to WC *hut* stimuli (p < .001), and that responses to MC *hurt* stimuli were faster than to WC *hurt* stimuli (p < .001). In this experiment, faster responses mean that the participant took less time between the start and end clicks, meaning an easier decision.

Figure 5.3 shows mean log(rt) by class, then by blocktype, i.e. whether the stimulus appeared in one of the Single talker blocks, or in the Mixed talker block (solid lines = Single; dotted lines = Mixed). This was a significant interaction (Pr(>F) = 0, F = 20.7568), showing that participant responses were significantly slower when responding to MC stimuli in the Mixed talker block than when responding to MC stimuli in the Single talker block (p < .001), but there was no difference between the blocks for the WC stimuli (p = .8).

5.3.3 Area Under the Curve

The output of MouseTracker can be presented in two main ways: normalised time, or raw time. These different presentations allow for different analysis methods to be used, which in turn means that the data can be interpreted in fundamentally different ways. The main difference between the two approaches is that in raw time, the duration between the start click (trial initiation) and end click (response selection) is preserved, so that the path of the trajectories can be analysed in real time. However, because the end points of such trajectories will always be separated in time, an overall spatial attraction measure cannot be taken.

Conversely, in a normalised time analysis, the duration of each trajectory is slightly adjusted so that the end points are brought together. This clearly means that duration cannot be analysed, either between factors or between trials. This is because Trajectory A, lasting (for example) 1000ms from start to finish, will have the same number of normalised points as Trajectory B, lasting 2000ms, so comparing a potentially interesting feature of the trajectories at, say, normalised timepoint no.50, is meaningless: this is actually a comparison of an event at 500ms of real time for Trajectory A, and 1000ms of real time for Trajectory B. All the analyses in this section of Chapter 4 therefore do not involve relating trajectories to time-dependent features of either other trajectories, or of the experimental stimuli. A raw time analysis is in progress, and will be completed in future work.

The advantage of using normalised time is that because each trajectory has the same start and end points, these trajectories can be mapped directly on top of each other, allowing for a direct comparison of their overall shapes. However the ‘direct’ comparison of normalised time trajectories must still be treated with a little caution, as it does not represent the reality of the participant's hand movements, but – due to the time normalisation – can be thought of as a ‘description’ of them, and each description is ‘good enough’ to perform sophisticated analyses not possible with the raw trajectories. One final caveat when directly comparing trajectories relates to potential shortcomings of the computer hardware, although this may be more of a concern for raw, time-sensitive analyses. Latency in the computer hardware, such as the computer display or the USB mouse, may introduce inaccuracies when comparing, for example, response times between trajectories. More sophisticated hardware (and software) such as eye-tracking technology can control for the issue of latency, but as stated before, the eye-tracking methodology was deemed impractical for this project. Nevertheless, this was not seen as a major issue in the mouse tracking methodology: response time analysis was completed, but a number of other analyses which were not time-dependent were also completed. Lastly, the same hardware and software was used throughout the experiment, to remove the possibility of large differences in system latency due to differences in hardware.

Mouse tracking experiments make use of the virtual space on the display screen (see Figure 5.1) to measure participants' arm movements, as they hear the stimuli. As such, many studies emphasise the importance of spatial attraction as a measure of the strength of the competitor. In order to achieve this, an analysis method known as Area Under the Curve can be applied, allowing for consistent, comparable measures to be obtained from the trajectory data.

For each experimental trial, the MouseTracker software records the cursor's trajectory between the start and end mouse clicks. Figures 5.4 and 5.5 are cropped screenshots of the Graphical User Interface (GUI) of MouseTracker's ‘MT analyzer’ program, when displaying selected subsets of cursor trajectories. Figure 5.4 shows the cursor trajectories of two individual trials completed by the experiment's first participant, MW01, as they responded to stimuli in the middle class Single talker block. The participant responded correctly to both of these trials.

The purple trajectory (which finished at the top right of the screen) was their response to the MC_hut.wav stimulus, and the blue trajectory (which finished at the top left) was their response to the MC_hurt.wav stimulus. While the analysis of individual trajectories is statistically uninformative, it is interesting to inspect them to understand the pattern of the participants' hand movements, and how these are recorded by the program. In the present analysis, the trajectories' time duration has been converted to 101 normalised timepoints. Figures 5.4 and 5.5 show that these points are represented along each of the trajectories. Close inspection of these trajectories reveals that on both trials the participant clicked the start button at the bottom of the screen, started to move slowly upwards, as represented by the closeness of the timepoints, then appeared to speed towards their chosen response, as can be seen by the relatively widely-spaced points. They then slowed down as they reached the button and clicked to record their response, finishing the trial.

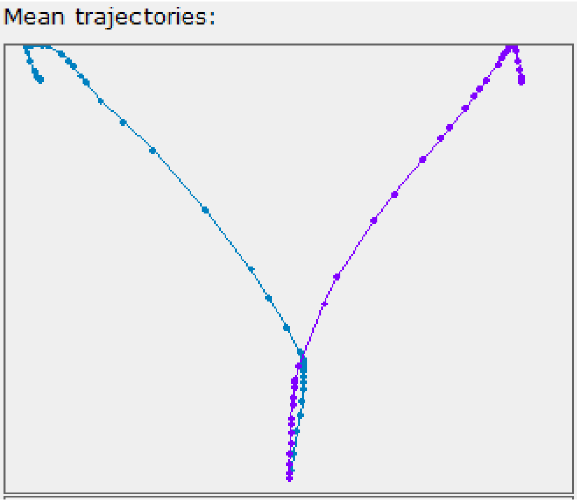


Figure 5.4: responses by MW01

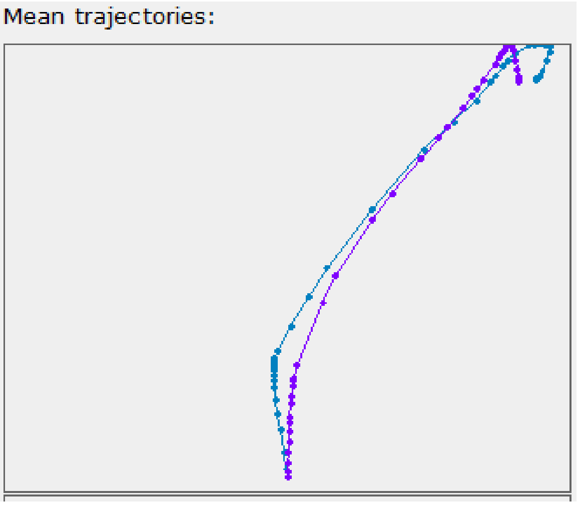


Figure 5.5: left trajectory flipped

The program shows trajectories by condition, and displays them in the same space, as in Figure 5.4. The program then ‘flips’ the left trajectories, so they can be easily compared with the trajectories on the right, as seen in Figure 5.5. This visual inspection does not form part of any of the analyses, but it does give a sense of the way that AUC is calculated, i.e. by displaying multiple trajectories on top of one another.

In order to calculate the AUC for each individual trajectory, MouseTracker calculates an idealised straight line between the start and end clicks, as in Figure 5.6. It then calculates the area between this line and the trajectory. This area is in MouseTracker's coordinate space, where x ranges from -1 to 1, and y from 0 to 1. The resulting number (which has no units, other than units of area in MouseTracker's coordinate space) can be used as a measure of overall spatial attraction towards the incorrect competitor, as the participant hears the stimulus. Each trajectory's AUC, having been calculated individually, is averaged for analysis, depending on the desired factor comparisons.

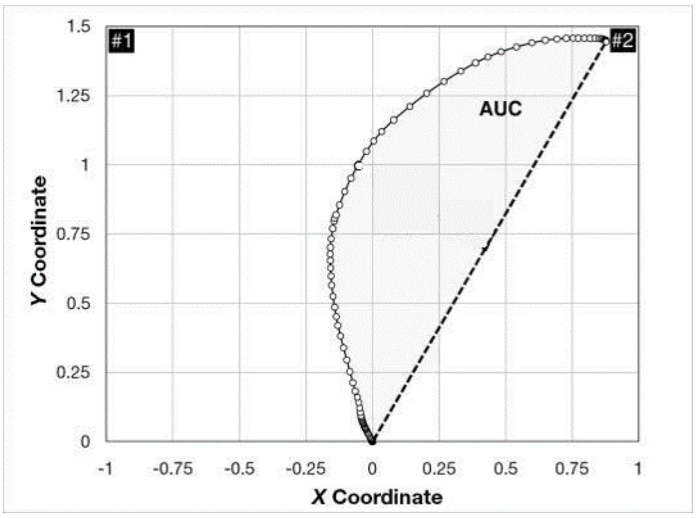


Figure 5.6: Freeman & Ambady 2010: 229

After step() was run on the fully saturated model, the best-fitting model for AUC (summary in Table 5.3) was:

$lmer(AUC \sim (coda + class + blocktype)\hat{}3 + (1|subject) + (1|target\_word))$

Table 5.3: Model summary for AUC (Experiment 3)

	AUC
coda_*hurt*	-0.114^{*}
	(0.063)
blocktype_Mixed	0.015
	(0.049)
class_WC	0.091^{*}
	(0.050)
coda_*hurt* X blocktype_Mixed	0.007
	(0.068)
coda_*hurt* X class_WC	0.138^{*}
	(0.071)
blocktype_Mixed X class_WC	-0.089
	(0.071)
coda_*hurt* X blocktype_Mixed X class_WC	0.247^{**}
	(0.100)
Constant	0.733^{***}
	(0.064)
Observations	4,486
Log Likelihood	$-5,641.532$
Akaike Inf. Crit.	11,305.060
Bayesian Inf. Crit.	11,375.560
Note:	*p<.1; **p<.05; ***p<.01

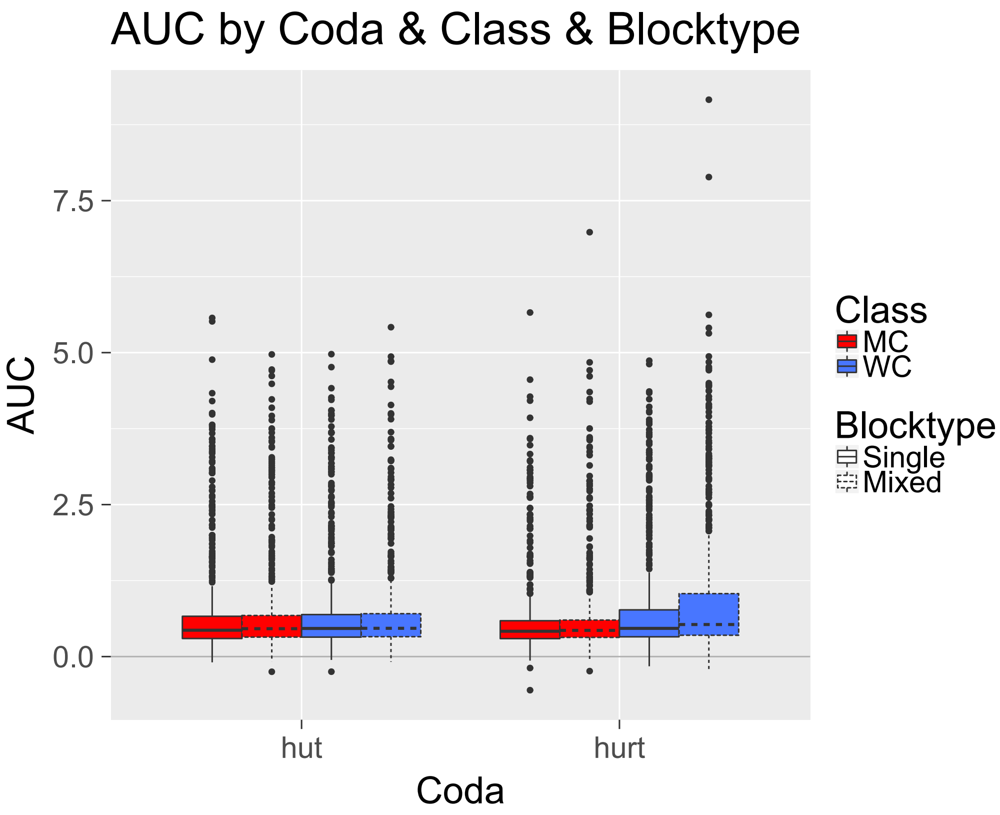


Figure 5.7: Experiment 3 AUC for responses to correct stimuli, by Coda, Class, & Blocktype

Figure 5.7 shows AUC by coda (*hut* = left, *hurt* = right), then by class (MC = red, WC = blue), then within class is blocktype (Single = solid lines, Mixed = dotted lines). This graph represents the significant 3-way interaction of Coda X Class X Blocktype (Pr(>F) = .013, F = 6.1374), such that there are no significant differences between AUC between blocktypes Single and Mixed, except for working class /r/ stimuli (blue boxes on the right; p = .006).

In general, trajectories had a lower AUC when listeners heard middle class stimuli than when they heard working class stimuli, meaning that there was an overall greater difficulty for listeners choosing between working class *hut* and *hurt*, than when they were choosing between middle class *hut* and *hurt*. This result was unsurprising, as the middle class stimuli are known to be easier to distinguish, as found in the previous experiments.

Furthermore, there was no difference between responses to middle class and working class *hut* (red vs. blue boxes on the left), but a large difference between middle class and working class *hurt* (red vs. blue boxes on the right), with the AUC for the middle class stimuli (red) much lower. This means that listeners found it relatively easy to respond to middle class *hurt* words, and much harder to respond to working class *hurt* words. This was in line with expectations.

Overall, the most perceptually challenging word types – *hurt* words produced with a derhoticised /r/ – result in the greatest amount of curvature when participants were making their decision, and even more so when the listening conditions are more difficult; that is, when the speaker is heard alongside another speaker in the Mixed block.

5.3.4 Discrete Cosine Transformation

While the AUC analysis described above is a very useful measure of general spatial attraction to incorrect competitors, a further measure was required, in order to delve into the dynamics of the trajectories. A curve analysis such as the kind described in the acoustic analysis for comparison of the formant trajectories in Chapter 2 (i.e. Figure 2.7) would have provided a measure of whether or not the trajectories were different to each other, but it lacked the overall statistical power for describing the more complex properties of the dynamic changes to the cursor trajectories along their path. Discrete Cosine Transformations (DCT) (Harrington, 2010: 304) were chosen for this purpose. A DCT is a mathematical operation that, when applied to a signal, decomposes it into a set of sinusoids. When these are summed, the original signal is reconstructed. This is a similar process which enables compression of wav audio signal into an mp3 file, or an image into a jpeg.

When a DCT is applied to a curve, the result is a set of coefficients, e.g. k_0, k_1, k_2, k_3, etc. The more coefficients that are used, the more detailed the description of the curve will be, therefore the closer to the original signal a reconstruction would be if the coefficients were to be summed. However, only four coefficients are used in this type of analysis, as the use of many more would be to needlessly over-describe the curve (Harrington, pers. comm.). Figure 5.8 (left) shows that these first four coefficients are proportional to the mean (k_0), the slope (k_1), the curvature (k_2) (terms adapted from Harrington, 2010: 311-12), and the 'noisiness' (k_3) of the original signal. For the purposes of this analysis, the term 'noisiness' is analogous to the overall complexity of the cursor trajectory's shape, which may reflect changes in the direction of the cursor, and therefore a degree of indecision by the participant.

The DCT coefficients can in fact be used to redraw the original curve. However this reconstructed curve is very close to, but not exactly the same as, the original curve. An example is Figure 5.8 (right), which is from Harrington (2010: 309). This is analogous to mp3 and jpeg formats (and other lossy compression methods) being a good approximation of the original sound or picture, but with some of the detail missing.

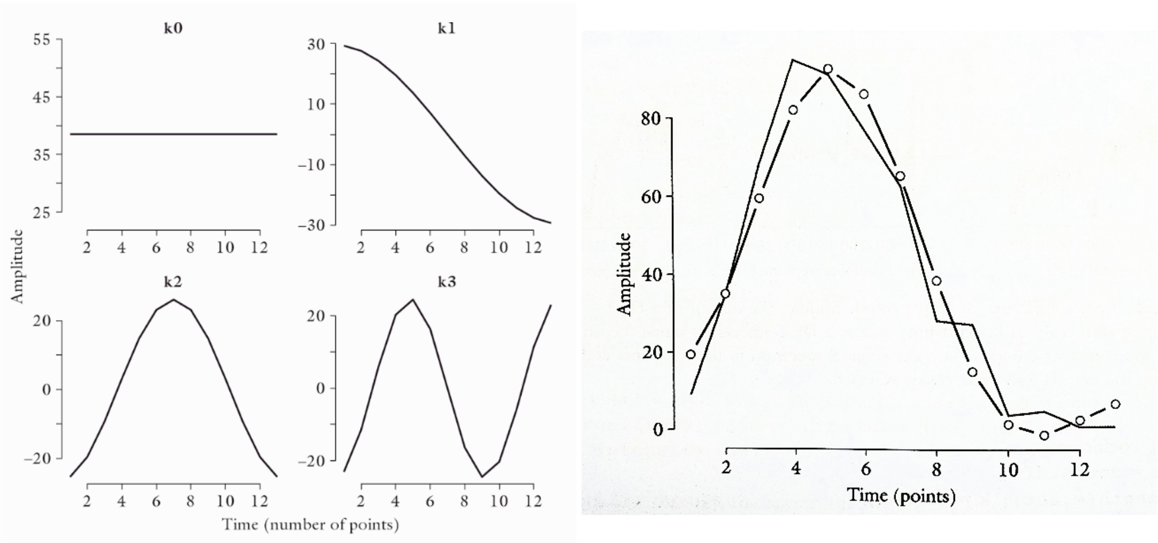


Figure 5.8: From Harrington (2010: 307 & 309): ‘The first four half-cycle cosine waves that are the result of applying a DCT to [a raw signal]’ (left) and ‘The raw signal (solid) and a superimposed DCT-smoothed signal (dotted) obtained by summing k_0, k_1, k_2, k_3’ (right)

Other phonetic studies have employed this technique for the analysis of formant trajectories (e.g. Watson and Harrington, 1999; Rathcke, Stuart-Smith, Timmins, and José 2012), but here it is applied to describe the properties of the cursor trajectories.

In order to understand how DCT is applied to the cursor trajectories, we must look at these trajectories from a different, more abstract perspective than we have done so far with the AUC analysis. Since cursor trajectories in the mouse tracking experiment move across 2-dimensional space (x- and y-coordinates on the screen) as well as through time (between start and end clicks), there are three dimensions in play. DCT can only be applied to curves with two dimensions, so we must re-visualise the trajectories in two ways: 1. **‘x-coordinates’** on the y-axis and normalised time along the x-axis; and 2. **‘y-coordinates’** on the y-axis and normalised time along the x-axis. These two graphs are shown in the rightmost panels of Figure 5.9, which shows participant MW_01's right-remapped responses to MC *hut* (purple) and *hurt* (blue), and in Figure 5.10, showing the same participant's responses to WC *hut* (purple) and *hurt* (blue).

DCT analyses are performed separately for x-by-time and y-by-time. In each analysis, the coefficient k_0 refers to the mean y-axis value of each cursor trajectory on the graph. For example, both of the lines in the x-by-time graph in Fig.5.9 (starting at $y=0$ and ending at $y=1$) appear to have the mean value of roughly 0.4. Therefore, we could say that the ‘xk_0’ of those trials is around 0.4. If the decomposed X and Y graphs are compared with the cursor trajectory graph to

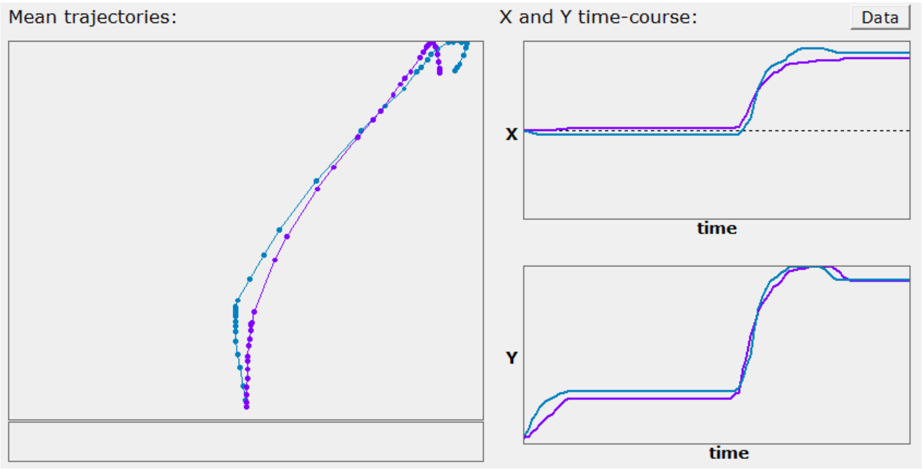


Figure 5.9: MW01's MC *hut/hurt* trajectories decomposed into x- & y-coordinates by time

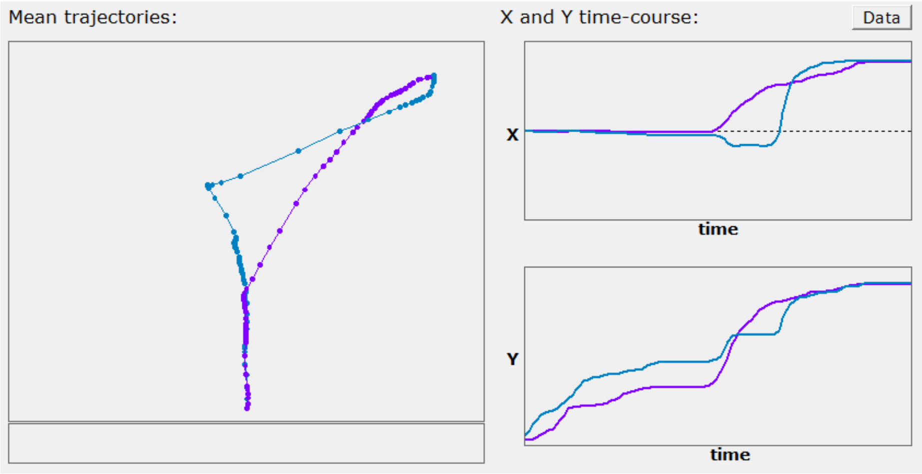


Figure 5.10: MW01's WC *hut/hurt* trajectories decomposed into x- & y-coordinates by time

their left, the participant appears to have spent just over half of the time in those two particular trials moving almost directly upwards, then stopping, explaining the relative lack of x- and y-coordinate change.

For this analysis, the R package EMU-R was used in order to obtain the four DCT coefficients of k_0, k_1, k_2, and k_3. Each of the following sections takes one of these coefficients at a time, starting with xk_0, xk_1, xk_2, and xk_3, then finishing with yk_0. The coefficients yk_1, yk_2, and yk_3 are very hard to interpret individually in the context of this work, so they will not be shown. Coefficients yk_1 and yk_3 each had only two significant main effects – Class and Blocktype – and yk_2 had only one – Class. These effects mirror the effects found in the results of the other y- and x-coefficients, and simply back up their results, albeit to a much lesser degree.

xk_0: Mean x-coordinate

After step() was run on the fully saturated model, the best-fitting model for xk_0 (summary in Table 5.4) was:

$$lmer(xk_0 \sim (coda + class + blocktype)\hat{}3 + (1|subject) + (1|target\_word))$$

Table 5.4: Model summary for xk_0 (Experiment 3)

	xk_0
coda_*hurt*	0.006
	(0.010)
blocktype_Mixed	−0.035***
	(0.008)
class_WC	−0.036***
	(0.008)
coda_*hurt* X blocktype_Mixed	0.019*
	(0.011)
coda_*hurt* X class_WC	−0.006
	(0.011)
blocktype_Mixed X class_WC	0.035***
	(0.011)
coda_*hurt* X blocktype_Mixed X class_WC	−0.045***
	(0.016)
Constant	0.334***
	(0.009)
Observations	4,486
Log Likelihood	2,686.120
Akaike Inf. Crit.	−5,350.239
Bayesian Inf. Crit.	−5,279.743
Note:	*p<.1; **p<.05; ***p<.01

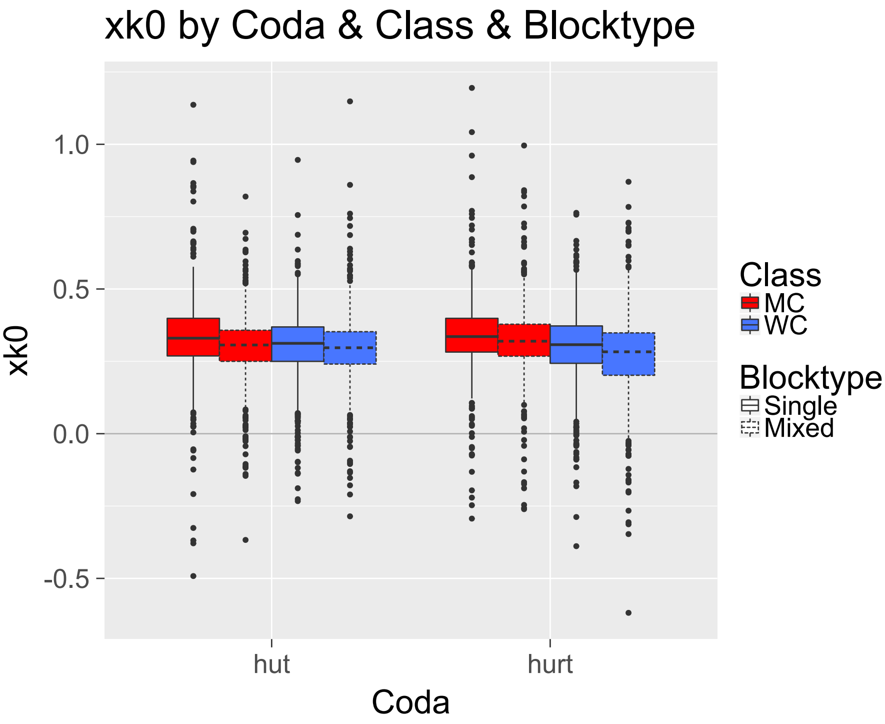


Figure 5.11: Experiment 3 xk_0 for responses to correct stimuli, by Coda, Class, & Blocktype

Figure 5.11 shows xk_0 by coda (*hut* = left, *hurt* = right), then by class (MC = red, WC = blue), then within class is blocktype (Single = solid lines, Mixed = dotted lines). This graph represents the significant 3-way interaction of Coda X Class X Blocktype (Pr(> F) = .004, F = 8.3207), such that there are significant differences in xk_0 between blocktypes Single and Mixed, for middle class /ʌ/ stimuli (red boxes on the left; p < .001), for middle class /r/ stimuli (red boxes on the right; p = .026), and for working class /r/ stimuli (blue boxes on the right; p = .003). The lowest mean x-coordinate of all is when listeners heard working class *hurt* stimuli in the Mixed talker block. This means that the Mixed block makes the most difficult word types even harder to process than they already are.

There was, overall, more difference between the classes in *hurt* stimuli than in *hut* stimuli. The lower xk_0, corresponding to trajectories with overall less deviation from the centre of the screen, means that listeners found it harder to decide that working class *hurt* words were correct, than to decide that middle class *hurt* words were correct. This was not surprising, given what is known from other results in all three experiments in this thesis.

xk_1: **Slope**

After step() was run on the fully saturated model, the best-fitting model for xk_1 (summary in Table 5.5) was:

$lmer(xk_1 \sim coda + class + blocktype + coda{:}class + (1|subject) + (1|target\_word))$

Table 5.5: Model summary for xk_1 (Experiment 3)

	xk_1
coda_*hurt*	-0.016^{***}
	(0.006)
blocktype_Mixed	0.014^{***}
	(0.002)
class_WC	0.013^{***}
	(0.003)
coda_*hurt* X class_WC	0.025^{***}
	(0.005)
Constant	-0.387^{***}
	(0.006)
Observations	4,486
Log Likelihood	5,123.020
Akaike Inf. Crit.	−10,230.040
Bayesian Inf. Crit.	−10,178.770
Note:	*p<.1; **p<.05; ***p<.01

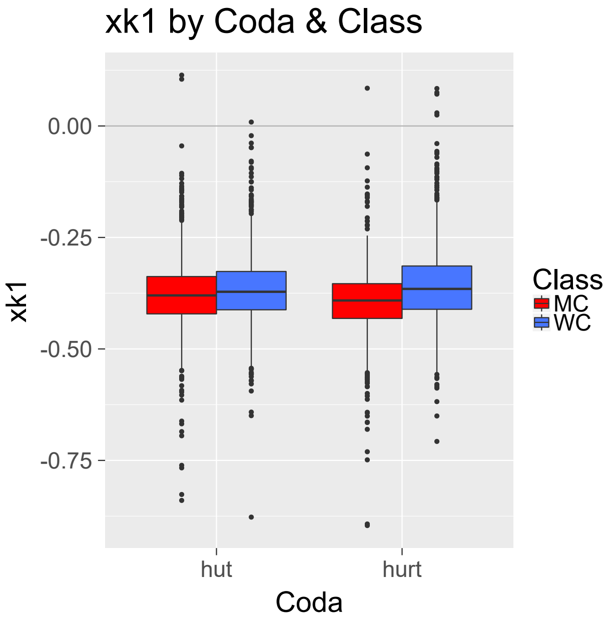


Figure 5.12: Experiment 3 xk_1, responses to correct stimuli: Coda & Class

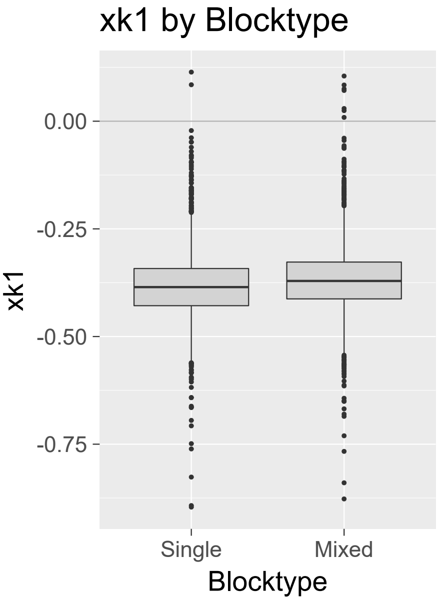


Figure 5.13: Experiment 3 xk_1, responses to correct stimuli: Blocktype

Figure 5.12 shows mean xk_1 by coda, i.e. whether or not the stimulus had an /r/, then by class (red = MC; blue = WC). This was a significant interaction (Pr(>F)<.001, F = 31.7813), such that the difference (p<.001) in slope between middle class *hurt* stimuli (-0.3956) and working class *hurt* (-0.3575) is much more than the difference (p<.001) in slope between middle class *hut* (-0.3794) and working class *hut* (-0.3668). Also, the difference between middle class *hut* and *hurt* is significant (p = .008), but the difference between working class *hut* and *hurt* is non-significant. A greater slope means an overall more direct route from the start button to the correct response, therefore a higher numerical value of xk_1 means that listeners found it easier to decide upon the correct response. In this interaction, this means that the easiest stimuli were the middle class *hurt* stimuli.

Figure 5.13 shows the xk_1 depending on whether the stimulus appeared in one of the Single talker blocks (either MC or WC), or in the Mixed talker block. This was a significant main effect of Blocktype (Pr(>F)<.001, F = 40.3883), such that the slope for stimuli in the Single blocks was greater (p<.001) (-0.3820) than the slope for stimuli in the Mixed block (-0.3677). This can be interpreted as the listener being more certain of the correct response by the time they have reached the end of their trajectory.

xk_2: Curvature

After step() was run on the fully saturated model, the best-fitting model for xk_2 (summary in Table 5.6) was:

$$lmer(xk_2 \sim (coda + class + blocktype)\hat{}3 + (1|subject) + (1|target\_word))$$

Table 5.6: Model summary for xk_2 (Experiment 3)

	xk_2
coda_*hurt*	−0.003
	(0.008)
blocktype_Mixed	0.011*
	(0.006)
class_WC	0.017***
	(0.006)
coda_*hurt* X blocktype_Mixed	−0.012
	(0.009)
coda_*hurt* X class_WC	0.003
	(0.009)
blocktype_Mixed X class_WC	−0.020**
	(0.009)
coda_*hurt* X blocktype_Mixed X class_WC	0.030**
	(0.012)
Constant	0.215***
	(0.006)
Observations	4,486
Log Likelihood	3,704.635
Akaike Inf. Crit.	−7,387.270
Bayesian Inf. Crit.	−7,316.774
Note:	*p<.1; **p<.05; ***p<.01

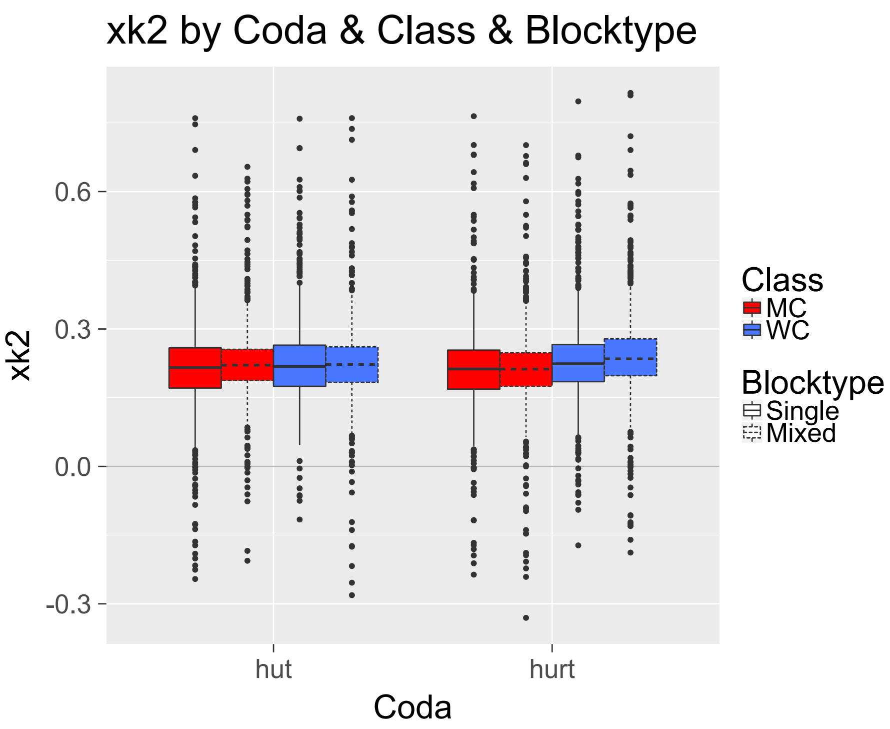


Figure 5.14: Experiment 3 xk_2 for responses to correct stimuli, by Coda, Class, & Blocktype

Figure 5.14 shows xk_2 by coda, then by class (red or blue), then within class is blocktype. This three-way interaction was significant (Pr(>F) = .018, F = 5.6386), such that there is a significant difference in xk_2 between middle class and working class *hurt* stimuli in the Mixed blocktype ($p < .001$), but no such difference between classes for *hut* stimuli in the Mixed blocktype ($p = .71$). A lower degree of curvature in the trajectory means that the participant followed a more direct route to the correct response.

xk_3: Noisiness

After step() was run on the fully saturated model, the best-fitting model for xk_3 (summary in Table 5.7) was:

$$lmer(xk_3 \sim (coda + class + blocktype)\hat{}3 + (1|subject) + (1|target\_word))$$

Table 5.7: Model summary for xk_3 (Experiment 3)

	xk_3
coda_*hurt*	0.021**
	(0.009)
blocktype_Mixed	-0.018^{***}
	(0.006)
class_WC	-0.021^{***}
	(0.006)
coda_*hurt* X blocktype_Mixed	0.001
	(0.008)
coda_*hurt* X class_WC	-0.024^{***}
	(0.008)
blocktype_Mixed X class_WC	0.015*
	(0.008)
coda_*hurt* X blocktype_Mixed X class_WC	-0.028^{**}
	(0.011)
Constant	-0.045^{***}
	(0.009)
Observations	4,486
Log Likelihood	4,024.601
Akaike Inf. Crit.	$-8{,}027.201$
Bayesian Inf. Crit.	$-7{,}956.705$
Note:	*p<.1; **p<.05; ***p<.01

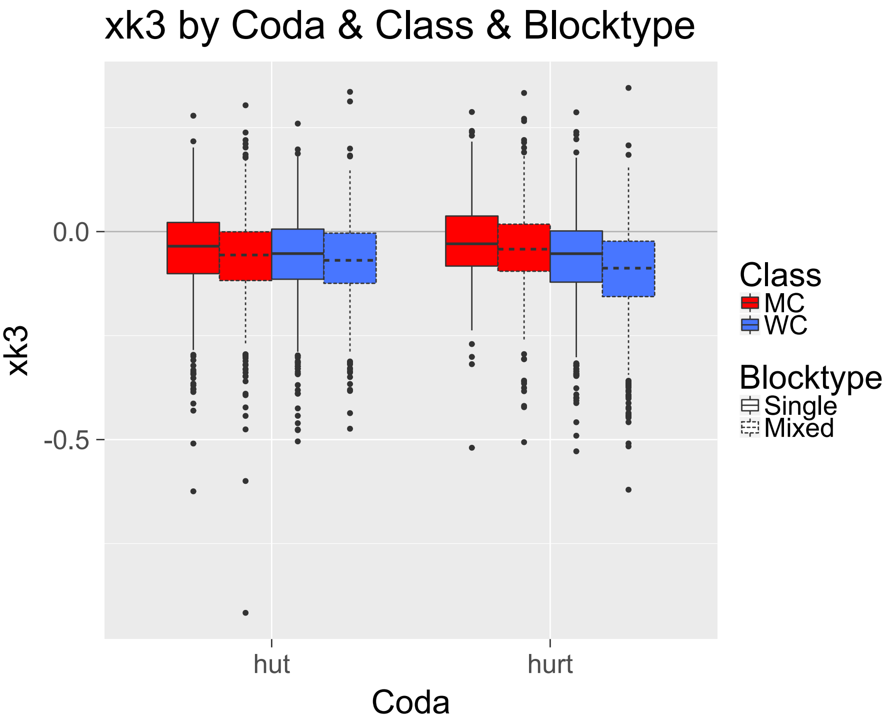


Figure 5.15: Experiment 3 xk_3 for responses to correct stimuli, by Coda, Class, & Blocktype

Figure 5.15 shows xk_3 by coda, then by class (red or blue), then by blocktype. This three-way interaction was significant (Pr(>F)=.014; F=6.0040), such that xk_3 was greater in the Mixed blocktype than in the Single blocktype for all stimulus types, except for working class *hut* stimuli (blue boxes on the left of the graph). More noisiness means that the trajectory more closely follows a curve with two changes in direction, so a greater (negative) numerical value of noisiness relates to more indecision in the participants' responses.

yk_0: Mean y-coordinate

After step() was run on the fully saturated model, the best-fitting model for yk_0 (summary in Table 5.8) was:

$lmer(yk_0 \sim class + blocktype + (1|subject) + (1|target\_word))$

Table 5.8: Model summary for yk_0 (Experiment 3)

	yk_0
blocktype_Mixed	0.011**
	(0.005)
class_WC	0.016***
	(0.005)
Constant	1.168***
	(0.020)
Observations	4,486
Log Likelihood	1,530.616
Akaike Inf. Crit.	−3,049.233
Bayesian Inf. Crit.	−3,010.780
Note:	*p<.1; **p<.05; ***p<.01

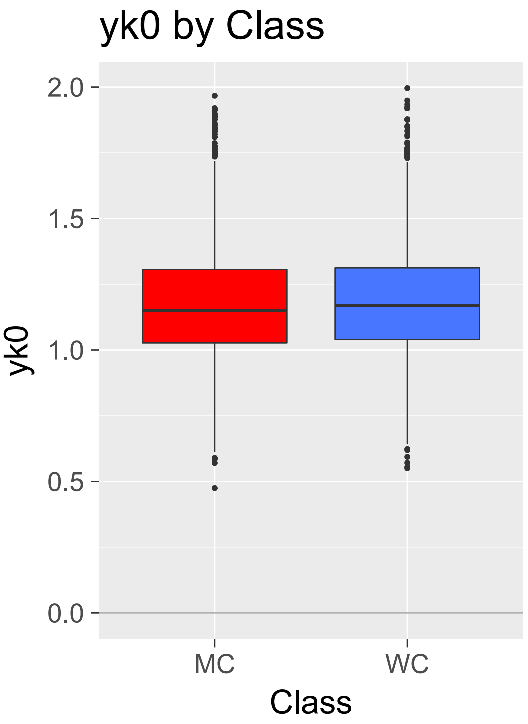


Figure 5.16: Experiment 3 yk_0 for responses to correct stimuli, by Class

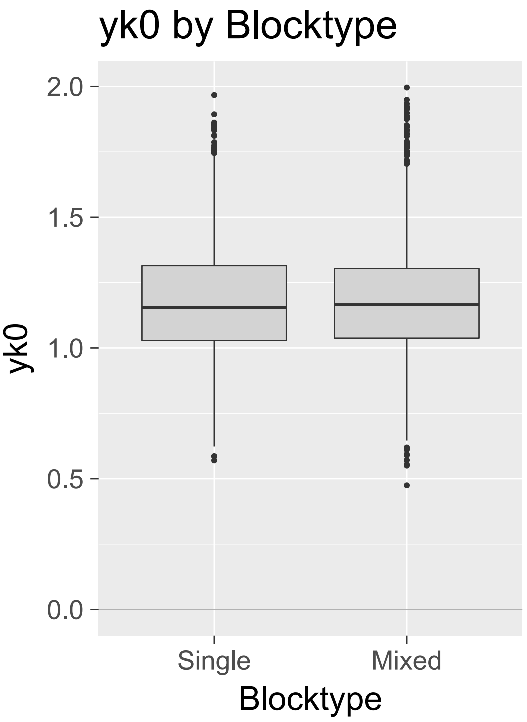


Figure 5.17: Experiment 3 yk_0 for responses to correct stimuli, by Blocktype

Figure 5.16 shows how the participants' mean yk_0 differed depending on whether the stimulus was middle class (MC) or working class (WC). This was a significant main effect (Pr(> F) = .002, F = 9.6725), such that the mean y-coordinate of trajectories in response to hearing working class stimuli was higher (p = .002) than trajectories in response to hearing middle class stimuli. A greater mean y-coordinate of the trajectory may be interpreted as the participant spending more time in the vicinity of the response buttons, taking more time to choose between them, meaning greater indecision.

Figure 5.17 shows the yk_0 depending on whether the stimulus appeared in one of the Single talker blocks (s), or in the Mixed talker block (r). This was a significant main effect (Pr(> F) = .029, F = 4.7955), such that the mean y-coordinate of trajectories in response to hearing stimuli which appeared in the Mixed talker block, was higher (p = .029) than the mean y-coordinate of trajectories in response to hearing stimuli which appeared in the Single talker blocks.

5.4 Discussion

In previous chapters, listeners were tested on their ability to discriminate perceptually similar Glaswegian words. They were observed to process these words in different ways, depending on their exposure to them over both long-term and short-term time spans. However, it was not possible to know in any great detail how Glaswegian listeners – seemingly the most ‘fluent’ listeners – discriminated these words until a more targeted examination was done, focusing on their ‘online’ perception of derhoticised /r/. As discussed in the introduction to this chapter, the mouse tracking methodology enables a much more detailed analysis than is afforded by more traditional methods such as response time, and the following discussion will interpret these analyses together.

With one motivation for this experiment being an investigation of the online processing of derhoticised /r/, the research question was formulated:
‘How do experienced listeners process ambiguous fine phonetic detail for a contrast?’

A second motivation for this experiment arose from an earlier experimental finding, discussed at the end of Chapter 3. This finding was that, when presented with more than one talker, listeners may have been experiencing difficulty in identifying variants in a two alternative forced choice task. This prompted the second research question:
‘Do harder listening conditions affect the online perception of ambiguous fine phonetic detail for a contrast?’

This section will discuss possible interpretations of the results for the analyses of response time and area under the curve, and finally the discrete cosine transformation's curve coefficients.

The results presented in this chapter show an overall pattern that is supported by taking all of the results together. First, middle class *hurt* words are the easiest for listeners to accurately and quickly identify, and this pattern was maintained across all presentation conditions. Figure 5.2 shows that responses to middle class *hurt* words have the lowest response time, Figure 5.7 shows that the area under the curve is the lowest for these stimuli, and a low value for spatial attraction indicates a more distinctive stimulus. In terms of the DCT coefficients, the significant interaction effects described above show that response trajectories to middle class *hurt* words have the greatest mean x-coordinate (Figure 5.11), the greatest slope (Figure 5.12), and the lowest amount of curvature (Figure 5.14), which relate to more overall movement towards the correct response, an earlier tendency to move towards it, and a more direct route taken. Trajectories in response to these stimuli also show the least noisiness (Figure 5.15), meaning that there is less indecision in the path taken by the cursor. These overall findings will now be investigated in more depth, by investigating the individual factors.

As seen in Figure 5.2, responses to middle class stimuli were much faster than to working class stimuli. This was unsurprising, as it replicated the finding reported in Experiment 1 that discrimination between middle class *hut* and *hurt* words is easier than discrimination between working class *hut* and *hurt* words.

Interestingly, when stimuli appeared in one of the Single blocks, either they were significantly faster than responses to stimuli in the Mixed block (for middle class stimuli), or they showed no difference (for working class stimuli) – see Figure 5.3 for this pattern. This is interesting because stimuli in the Mixed blocks were heard *after* those in the Single blocks, and since the same actual stimuli were repeated, one might expect the listeners to be faster upon hearing them again, since they had heard them only a few minutes previously. Combined with the task learning effect which was theorised to be in effect in Experiment 2, one might expect an overall decrease in the response time from the Single talker blocks to the Mixed talker blocks.

A simple interpretation is that this is the result of a fatigue effect, as the Mixed block was always the last block to be done by the listeners. Another interpretation is that listeners were struggling in the Mixed block, compared to the Single blocks, because their perception has somehow been made more difficult by the listening conditions. This appears to be a direct answer to the secondary research question of this chapter, which concerned the difficulty of listening conditions and their effect on online perception.

The randomisation of the two speakers seems to have caused difficulty for the listeners, but the design of this experiment means that it is not easy to pinpoint the exact source of this difficulty, as there may actually be more than one factor at play. Firstly, the fact that there were two different talkers, each with their own voice qualities, may have placed an extra perceptual load on the listeners. Secondly, there was the additional element of the existence of two accents in the Mixed block, which is also likely to account for part of the processing cost that resulted in increased response times. Because of these complications, it could be said that listeners effectively have multiple options when asked to choose which word they heard in the Mixed block, rather than a binary choice between two words, produced by one speaker – with one accent – in the Single blocks. In summary, the factors of talker and accent are confounded, so it is not possible to explicitly unpack why the difficulty arose for the listeners. Future work could shed more light on these effects. Clopper (2017) found complex interactions between dialect familiarity, speaker identity, and lexical information.

If we look at the significant two-way interaction of Class by Blocktype for response time (Figure 5.3) a very interesting pattern emerges, which clarifies the effect of the processing cost reported above. Responses to the middle class stimuli were indeed significantly slower in the Mixed talker block than in the Single talker block. But in contrast, responses to the working class stimuli were in fact no different in their response time from the Single block to the Mixed talker block. This is rather surprising, because discrimination between middle class *hut* and *hurt* is much easier than the same discrimination in working class stimuli, so why should their response times suffer when the working class response times do not? This will be explored in more depth in the sections below.

For the Area Under the Curve result, the most perceptually challenging word types – *hurt* words produced with a derhoticised /r/ – result in the greatest amount of curvature when participants were making their decision, and even more so when the listening conditions are more difficult; that is, when the speaker is heard alongside another speaker in the Mixed block. This validates the use of AUC as a measure of spatial attraction for this study. However, if we now look more closely at each of the discrete cosine transformation coefficients we will be able to explore features of the participants' responses which dynamically vary over the course of the trajectories.

The first coefficient, xk_0, directly relates to the mean x-coordinate, and when listeners were responding to middle class stimuli it was much greater than when they were responding to working class stimuli. This means that, when hearing middle class stimuli, either one of two general patterns were followed:

1. Listeners moved towards the correct response earlier in the trajectory than for working class stimuli, or;

2. Listeners spent more time near the correct response, than for working class stimuli.

Of course, it could have been a combination of both of these patterns, and it likely was for most trajectories. However whether it was an *earlier movement to the correct response* or a *general proximity to the correct response*, cannot be determined from the xk_0 coefficient alone, as it only represents an average figure. Nevertheless it is possible to say that listeners are indeed more ‘swayed’ towards the correct response for middle class words than for working class words, highlighting the general pattern seen in many previous sections in this thesis that the middle class *hut/hurt* pairs are relatively easy to distinguish.

Middle class *hut* and working class *hut* words also had significantly different mean x-coordinates, although less so than the difference between classes for *hurt* words. This *hut* difference shows that listeners found it easier to process the middle class speaker's stimuli than the working class speaker's stimuli. The perception of working class *hut* words probably suffered because of their similarity to working class *hurt* words, meaning that even in the Single working class block, listeners still had trouble in deciding that the word was definitely /r/-less, whereas this difficulty was not present when hearing the middle class *hut* words, as they were very different to the middle class *hurt* words. Interestingly, middle class *hurt* was easier than middle class *hut* (which only achieved $p = .087$), possibly because listener's processing of middle class *hut* words suffered due to their acoustic similarity to the working class *hut* and *hurt* words, which are harder to distinguish. Perception of the otherwise unambiguous middle class *hut* words might be getting 'caught in the crossfire', when heard alongside the ambiguous working class words.

The next coefficient, xk_1, should be interpreted such that the greater the negative slope there is in the x-coordinate, the more time the participant spent in the region of the correct response *towards the end of the trajectory's path*, instead of hovering in the middle. As Harrington writes, the slopes have a negative value because the k_1 is the inverse of the slope of the curve which it describes, so that 'there is almost complete (negative) correlation between [a spectral slope and k_1], i.e., greater positive slopes correspond to greater negative k_1 values' (2010: 312).

The significant interaction for xk_1 (slope) was Coda by Class (Figure 5.12), with the Slope difference between middle class *hurt* and working class *hurt* stimuli being much greater than the difference between middle class *hut* and working class *hut* stimuli. This means that the middle class *hurt* words appear to be much easier than the working class *hurt* words, because the trajectories spent more time in the vicinity of the correct response option.

For xk_2, a greater amount of curvature in the trajectory means it generally follows less of a straight line – in other words, less like the 'easy' idealised straight line described in the method for the Area Under the Curve analysis (Figure 5.6). Therefore, a greater curve translates to more perceptual difficulty when the listener hears a particular word type.

The significant three-way interaction for xk_2 was Coda by Blocktype by Class (Figure 5.14). In the Mixed block, there was a significant difference between middle class and working class *hurt* words (solid red and blue boxes on the right of the graph), such that the working class stimuli elicited a greater curvature. In contrast, there was no difference between the middle class and working class *hut* words. This is yet another example of the difficulty of hearing the words in the Mixed block, and shows that the curvature coefficient xk_2 can reveal an interesting element of the dynamics of a response. In fact, Figure 5.14 may be compared with the earlier Figure 5.7 for Area Under the Curve, which is conceptually similar to degree of curvature – as the two figures depict very similar patterns of results, it may be said that AUC and DCT measures are an effective 'test' for each other, as well as companion analyses.

The final coefficient, xk_3, shows how well the trajectory corresponds to a sinusoid curve with exactly two changes in direction (i.e. the bottom-right panel in Figure 5.8 (left)), so a greater xk_3 – that is, more 'noisiness' – indicates more changes in direction of the participant's mouse movements. This can be taken to mean that there is more indecision when choosing the response, as the participant may move towards the incorrect competitor, then back to the correct response to make their final decision.

The three-way interaction of Coda by Blocktype by Class (Figure 5.15) was significant for xk_3. This interaction showed relatively little difference between the noisiness of either middle class or working class *hut* words, in either Single or Mixed blocks (boxes on the left of the graph), but a large amount of variation between all types of *hurt* stimuli (boxes on the right of the graph). For both middle class and working class *hurt* stimuli, the participants' indecision increased massively in the Mixed block, meaning that hearing more than one talker or accent affects listeners' confidence in identifying the word they heard. The indecision was especially pronounced for the working class *hurt* stimuli in the Mixed block, confirming earlier results in this vein. Furthermore, the fact that the middle class *hurt* words promoted so little noisiness (especially in the Single talker block: red boxes with solid outlines on the right of the graph) is confirmation that words with highly rhotic variants are very easy to perceive in comparison to the other words in this experiment.

The first y-coordinate coefficient yk_0 relates to the mean 'height' of the participant's cursor, as they move up the screen. This means that a greater mean should be interpreted as relatively more time spent near the top of the screen. In other words, this can be taken to indicate increased difficulty or indecision for a particular stimulus type, as the participant is not 'heading straight for the target' upon hearing the word.

There were only two significant effects for yk_0: Class and Blocktype. For Class, working class stimuli evoked trajectories which spent longer near the top of the screen than for middle class stimuli. For Blocktype, the same pattern was found for stimuli in the Mixed block, causing more time to be spent at the top of the screen than for stimuli in the Single blocks.

These two results go hand-in-hand with the results for xk_3, above, which represents noisiness, or indecision due to multiple changes in direction. This is because the more changes in direction a participant makes, the more likely they will be to be spending time near the top of the screen, moving back and forth between the two response buttons, which are at the top corners of the display.

This is an interesting pattern to consider in relation to the competing motor impulses in arm movements, which may relate to 'competing cognitive representations' (Farmer, Anderson & Spivey 2007: 573). Overall, these yk_0 results add to the general pattern in Experiment 3, that the more difficult stimuli (working class) or more challenging listening conditions (Mixed block) result in more complicated trajectory paths, meaning more perceptual difficulty and indecision for the participants.

Some important information was collected in the post-experiment questionnaire (Appendix 10), and it may serve as a partial explanation for some of the complex results described above, especially those pertaining to extra perceptual load, e.g. increased response times for some stimuli. One of the questions asked participants 'How many speakers did you hear in the experiment?' There was a surprising variety of answers to this question. Out of the 51 listeners, three reported hearing only one speaker in the whole experiment. This may have been because they thought it was one speaker producing different accent varieties, even though there were notable differences between the voice qualities of the middle class and working class speakers. It is possible that these three listeners did not pick up on these differences due to a lack of attention to the voices, or a highly tuned "ear" for voices or accents. A further 25 listeners reported hearing two speakers in the whole experiment; this is the 'correct' answer.

Intriguingly, twelve listeners thought there were three speakers in the experiment, eight listeners thought there were four, and three listeners thought they heard five speakers in the experiment. That is, almost half of the participants (23 out of 51) reported hearing more talkers than were actually present. This raises the possibility that in the absence of any obvious factors which impede processing, such as noise in the signal, listeners may attribute a high cognitive load to the existence of multiple talkers, even when the number of talkers they think they are hearing is much greater than what they actually heard. In fact, the challenging listening conditions in Experiment 3 were created by the randomisation of the tokens (produced by more than one talker and accent in the Mixed block), coupled with a difficult mouse tracking task. This points to a similar pattern of cross-dialect lexical processing costs as found by various other studies (e.g Clopper, Pierrehumbert & Tamati, 2010; Clopper 2017; Floccia, Goslin, Girard & Konopczynski, 2006; Sumner and Samuel, 2009). Again, further investigation of this issue is warranted.

5.4.1 Summary

This chapter has presented Experiment 3, which used the Mouse Tracking methodology to investigate the dynamics of perception of phonemic contrasts in minimal pairs such as *hut* and *hurt*.

Analyses of Response Time, Area Under the Curve, and Discrete Cosine Transformation coefficients all showed that speaker class was a highly important predictor of processing difficulty, such that discrimination between middle class *hut* and *hurt* stimuli is less challenging than discrimination between working class *hut* and *hurt* stimuli. This addresses the first research question for this chapter: ‘How does experience relate to the learning of ambiguous fine phonetic detail for a contrast?’

A further finding was that when listeners heard the stimuli in the Mixed talker block, this generally made discrimination harder than when they appeared in one of the Single talker blocks. This effect was not as striking as the one for Class. This addresses the second research question for this chapter: ‘How do experienced listeners process ambiguous fine phonetic detail for a contrast?’

In summary, the overall finding was that middle class *hurt* words were the easiest to process, followed by middle class *hut* stimuli in the Single talker presentation condition. This was partnered by the finding that middle class *hut* stimuli in the Mixed talker presentation condition were relatively hard to identify, most likely because of their potential confusion with working class pronunciations of both *hut* and *hurt* words.

This chapter has also demonstrated the usefulness of the mouse tracking methodology, as it enables the use of detailed analysis techniques showing dynamic characteristics of the responses which could not be investigated using more common techniques such as response time. These characteristics include the participants' tendency to head more directly towards – and spend more time in the vicinity of – the correct response for the easier stimuli, as revealed by the ‘slope’ measure xk_1, as well as the ‘noisiness’ measure xk_3, which can be seen as a way to quantify the amount of indecision a participant experiences on a set of trials. Such nuanced analysis techniques are certain to find use beyond the scope of this investigation.

Part III

Discussion and Conclusion

Chapter 6

General Discussion

6.1 Introduction

This thesis has described a set of speech perception experiments, and the results of these experiments provide valuable information about a number of factors that can affect the perception of phonetic detail, including long-term accent experience, short-term learning of an accent, the detail of online perception, and perception under difficult listening conditions.

The accent under investigation was Glaswegian, and the perceptual testing ground was the socially-stratified realisation of postvocalic /r/ (e.g. in *car*, *hurt*). The particular focus of the experiments was listeners' ability to perceive ‘derhoticised /r/’, an audibly weak rhotic variant, with ambiguous acoustic properties, that is typically produced by working class speakers.

There were four primary research questions addressed in this thesis:

What is the role of experience in the perception of fine phonetic detail for a contrast?

How does experience relate to the learning of ambiguous fine phonetic detail for a contrast?

How do experienced listeners process ambiguous fine phonetic detail for a contrast? and,

Do harder listening conditions affect the online perception of ambiguous fine phonetic detail for a contrast?

They were addressed by running and analysing three perceptual experiments, making use of a number of analysis techniques.

Experiment 1 primarily examined the role of long-term experience or learning on the ability to distinguish minimal pairs which only vary across a very fine phonetic contrast.

Experiment 2 then moved closer in the temporal domain, addressing the important question of what happens in the very early stages of learning a new accent feature, in other words, perceptual adaptation to this fine phonetic detail.

Finally, Experiment 3 zoomed right in on the detail of the perception of this phonetic contrast, inspecting the dynamics of the most familiar listeners' perception of the contrast.

This chapter will now present each research question in turn (ordered by the experiments they influenced), looking at how the experimental results answered the question, and comment on how the patterns might be explained by theoretical positions in speech perception. Following this is an in depth discussion of the relation between the theories of speech perception introduced in Chapter 1, and the overall pattern of all the experimental results presented in this thesis.

6.2 Experiment 1

The research question for Experiment 1 was:
'*What is the role of experience in the perception of fine phonetic detail for a contrast?*'

The analysis presented in this experiment showed a very clear effect of listener experience on the perception of derhoticised /r/, in that the Glasgow listeners, who had the most experience with hearing the Glaswegian accent, were much more sensitive to subtle differences in fine phonetic detail, and they also showed the least response bias, revealing that they were highly attuned to the phonemic categories intended by the working class speakers in the experiment.

In contrast, Cambridge listeners were much less sensitive to difference between *hut* and *hurt* words, and they showed a large bias towards hearing *hut* words, even when the speaker's intention was to produce a word with /r/. The low sensitivity result shows that these listeners' lack of experience with working class Glaswegian speech severely affects their ability to interpret fine phonetic cues to a distinction, and the strong *hut* bias suggests that, in the absence of knowledge about the variation of such phonetic detail (due to their inability to distinguish *hut* from *hurt*), they tend towards organising stimuli into categories which are known to them. In other words, the similarity of the derhoticised *hurt* words to plain-vowel *hut* words encourages unfamiliar listeners to categorise them as plain-vowel *hut* words.

Theoretical positions such as exemplar models and Bayesian inference would likely state that this patterning shows a clear benefit of experience for the Glasgow listeners, due to a vastly increased number of exemplars of derhoticised /r/, compared with the Cambridge listeners. This is almost certainly linked to the increased contextual and situational knowledge of the Glaswegian listeners, that speakers intended to produce words with /r/.

Showing a more interesting pattern was the Intermediate listener group, representing English listeners who had lived in Glasgow for around three years. As expected, their sensitivity to difference between *hut* and *hurt* words was between those of the Glasgow and Cambridge listener groups. However, their bias had 'overshot' the ideal pattern that the Glasgow listeners displayed (i.e. little or no bias), and were now showing an effect of 'perceptual hypercorrection', where they were over-reporting the presence of /r/ in the ambiguous stimuli.

This suggests that the Intermediate listeners had accrued knowledge about the *existence* of derhoticised /r/ during their time in Glasgow, learning that it is a 'device' used by working class speakers to signify the presence of /r/. They may have been over-influenced by this relatively recently learned knowledge about a fairly unusual phonetic feature, and applied it to linguistic environments where they knew it could appear. This use of knowledge, expectation, and evidence, is a fundamental part of Bayesian inference, and it appears that the Intermediate listeners in Experiment 1 were part of the way along their journey of accruing enough evidence to effectively adjust their prior beliefs, for more accurate perception of derhoticised /r/.

One of the bias results in Experiment 1 warrants further discussion. On the extreme right of Figure 3.7, the response bias in the Glasgow listeners' (blue) responses to both middle class (solid lines: $c = -0.0905$) and working class (dotted lines: $c = 0.0758$) stimuli can be seen. Glaswegian listeners were very slightly biased towards reporting hearing /r/-less words when hearing the working class speaker. This was unsurprising – despite the Glaswegian listeners' high sensitivity to difference between the working class pairs (as shown by their *d'* results), the vowel-like formant structure of *hut* and *hurt* words produced by the working class speakers made it likely that there would be a slight bias towards reporting the absence of an /r/. However, they were also slightly biased to report hearing the middle class stimuli as more /r/-ful, regardless of whether the stimulus did in fact contain an /r/. The difference between these results was not large, but it was approaching significance, at $p = .08$. This result is in line with the bias of the other listeners: both the Cambridge and Intermediate listener groups were biased towards reporting more /r/-ful words in the middle class stimuli. However, the result was most surprising in the Glaswegian listeners, as it was predicted that their increased experience with the accents in question would help them to show almost no bias in the 'easy' distinction between *hut* and a highly-rhotic *hurt*.

This result may have important implications for the wider literature on Scottish rhoticity. Lawson, Scobbie & Stuart-Smith (2011b & 2014) report the results of both auditory and articulatory data, collected across the central belt of Scotland. They write that their data suggests a correlation between the auditory realisations of /r/ in both middle class and working class speakers (Figure 6.1), and the same speakers' articulatory configurations, taken from tongue-spline measurements in Articulate Assistant Advanced (Wrench 2007) (Figure 6.2).

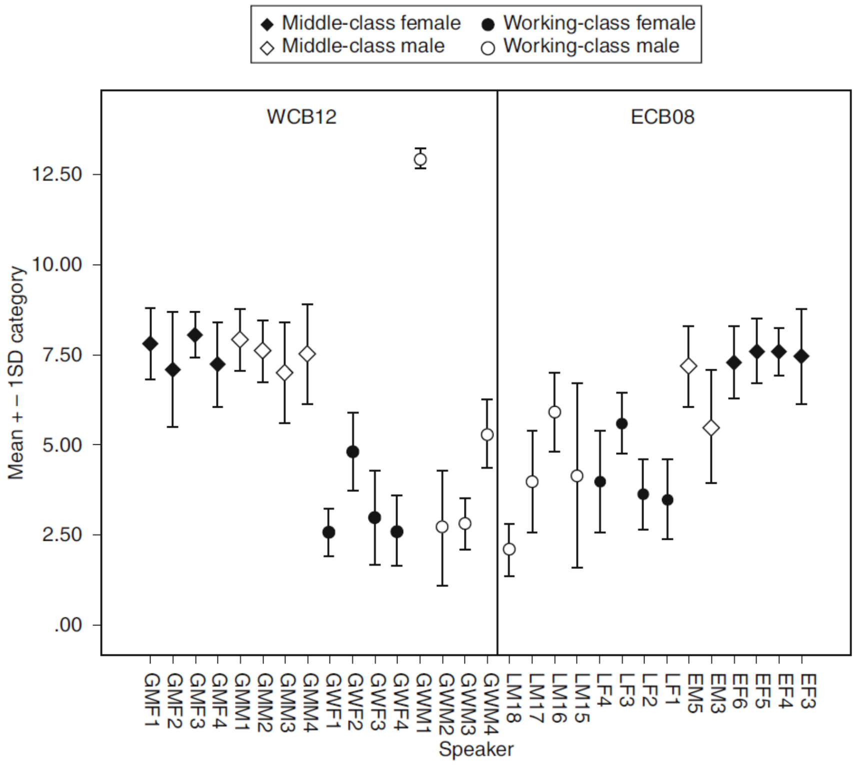


Figure 6.1: Phoneticians' auditory ratings of /r/-strength in speakers in/near Glasgow (WCB) and Edinburgh (ECB). Individual speaker /r/-index score means +/− one standard deviation. WCB12, N = 394, ECB08, N = 136 (from Lawson et al. 2014: 67)

These two graphs each show a clear split between the ‘strong’ and ‘weak’ /r/ productions of middle class and working class speakers in the Scottish central belt, and the same split can be seen across both graphs. Middle class speakers (diamond symbols in Figure 6.1) have stronger auditory /r/ ratings (higher on the y-axis), corresponding to more front- and mid-bunched articulations for /r/, and working class speakers have much weaker auditory /r/ ratings, corresponding to more front- and tip-up (i.e. derhoticised) articulations for /r/. The strong correlation between the patterns in the two analyses is confirmed by the authors. Interestingly, there appears to be slightly more class polarisation in the Western Central Belt auditory data (i.e. Greater Glasgow) than in the Eastern Central Belt auditory

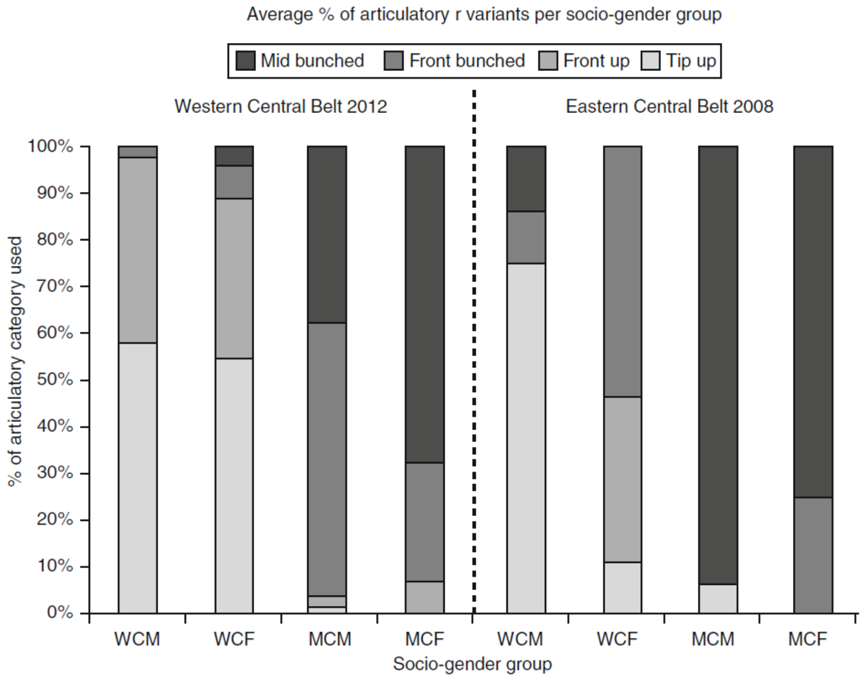


Figure 6.2: Percentage of articulatory /r/ variants used by each socio-gender group in the western and eastern Central Belt. WCB12, N = 394, ECB08, N = 136. Shades from lightest to darkest represent TIP UP, FRONT UP, FRONT BUNCHED and MID BUNCHED configurations respectively (from Lawson et al. 2014: 70)

data (i.e. Edinburgh and surrounding areas). The pattern is perhaps not as notable in the articulatory data (Figure 6.2), but companion perceptual experiments in the East and West Central Belt would shed more light on the perceptual correlates of these data.

The auditory judgements were made by phoneticians who were very experienced in hearing different variants of /r/ in the central belt of Glasgow, so these results can confidently be taken as accurate representations of the auditory quality of the /r/ variants in middle class and working class speakers. Experiment 1 is the first perceptual study which tests this in the wider population. The bias pattern of the Glaswegian listeners in Experiment 1, as described above, appears to support the existence of a split between how middle class and working class /r/ variants are perceived. In other words, Glaswegian listeners know about the difference between middle class and working class /r/ variants, and judge the speakers of the two sociolects differently *based on the indexical information they have inferred about the speakers.*

The clearly co-dependent or intertwined nature of social and phonological information is most easily accounted for by using the kind of representations as put forward in exemplar theory. Since the response bias for Glaswegian listeners is *towards* /r/ for the strongly rhotic middle class speakers (solid blue boxes on the right of Figure 3.7), but *away from* /r/ for the weakly rhotic working class speakers (dotted blue boxes on the right of Figure 3.7), this means they are classifying the middle class speakers as being more /r/-ful in general, even in their *hut* words, and the opposite for the working class speakers. If the Glasgow listeners are altering their judgements of individual *hut* and *hurt* tokens based on the identity of the speaker who produces them, this is firm evidence that they are using the speaker's identity to shape their expectation of which /r/ variant they are likely to hear. This mirrors the ‘integrated talker and phoneme processing’ that Mullennix & Pisoni (1990) found in their influential experiment, which was one of the catalysts for the growth in popularity of exemplar theory.

Crucially though, the Glasgow listeners still primarily treat both the middle class [ɚ] and the working class [ʌˁ] as /r/, to a much greater degree than either of the English listener groups (Intermediate and Cambridge), underlining the benefit of increased exposure. Moreover, the Glaswegian listeners have acquired knowledge of the system in an entirely different way. One issue which has not been raised as yet in this thesis is the possibility of the speaker's ‘own accent’ having an effect on their perception. In general terms it is likely that a listener who hears the same accent as their own will benefit from this fact, but it is undoubtedly very difficult to tease this apart from their accent exposure, that is, the accents they hear around them.

In the three perceptual studies described here, the vast majority of Glaswegian listeners spoke with middle class accents, due to the fact that recruitment took place in the University of Glasgow. In the scope of this work this issue was unavoidable, and the same issue likely affects many perceptual studies which are conducted on university campuses. This issue, in combination with the results of the perceptual experiments in this thesis could promote the question: ‘Why do middle class Glaswegian listeners perform so well when hearing working class derhoticised /r/?’ It is possible that they have an acute awareness of the articulatory configurations and timings that are required to produce derhoticised /r/ with a tip-up gesture in Glaswegian, as that strongly correlates with pharyngealisation in the working class accents in the central belt of Scotland (Lawson, Stuart-Smith & Scobbie 2017), which is what they actually hear. This could of course still be purely a matter of perceptual experience, rather than an effect of the speaker's own accent, but further experiments along these lines would help to explore this.

6.3 Experiment 2

Experiment 1 considered the effect of long-term learning, then the focus was adjusted to short-term learning, which was tested in Experiment 2, whose aim was to determine what happens when listeners have the opportunity to learn phonetic detail. The research question for Experiment 2 was:
'*How does experience relate to the learning of ambiguous fine phonetic detail for a phonemic contrast?*'

The analysis once again showed an effect of listener experience on the perception of derhoticised /r/, with the Glasgow listeners again being the most sensitive of the three groups to stimulus difference. The response bias results were replicated, but with one interesting difference which is described below. Response time analysis also showed a benefit for increased experience, though this effect was not as strong as that seen for sensitivity.

The short-term adaptation element of this experiment showed that, where the stimuli were in the Natural exposure listening condition (which presented listeners with resynthesized but acoustically unmanipulated stimuli in the Exposure story between Pretest and Posttest) the Glaswegian listeners were also the ones who benefited the most from hearing the *hut* and *hurt* word types in the context of a passage, read by the same speaker. This was shown in the sensitivity and response time analyses, which showed significant differences between Pretest and Posttest, most of all for the Glasgow listeners. In contrast, the Altered exposure condition (with some acoustic differences between *hut* and *hurt* words 'neutralised') did not help any of the listeners improve, whether this was measured in sensitivity or response time.

Response bias for Cambridge listeners shifted from *hut* to *hurt* in both listening conditions, thus matching, after short-term learning, the long-term bias of the Intermediate listeners. This suggests a very fast adaptation in the listeners' perceptual systems, and appears to signal a rapid change in their vowel and rhoticity categorisation criterion for this speaker. Presumably they would then be able to expand this newly altered categorisation criterion to other speakers of working class Glaswegian they encountered, but it would be difficult to predict how long this effect might last. Experiments by Kraljic & Samuel (2005), Eisner & McQueen (2006), and others, have shown that this type of knowledge is retained after 25 minutes, and 12 hours, respectively.

Since the short-term bias of the Cambridge listeners is virtually the same as the long-term bias of the Intermediate listeners, it may be assumed that previously unfamiliar listeners very quickly learn about the existence of fine phonetic detail such as derhoticised /r/ and the phonetic environments in which it may appear (for that speaker, or for that accent), but they do not markedly improve their ability to correctly apply that knowledge. This can be seen in the fact that their sensitivity shows a much smaller change (difference from solid to dotted lines in both red and green boxes on the right of Figure 4.9) than the big shift in their response bias (difference from solid to dotted lines in both red and green boxes on the right of Figure 4.10). Indeed, the response bias for Intermediate listeners shifted from a bias for *hurt*, to an even greater bias for *hurt*, in both listening conditions, possibly suggesting that the Exposure story boosted their knowledge about the *existence* of derhoticised /r/, which resulted in a tendency to perceptually hypercorrect to an even greater degree.

The very small improvement in sensitivity to stimulus difference between Pretest and Posttest for the Cambridge listeners shows a very low benefit for short-term learning, and the fact that, after three years, the Intermediate listeners are also not at the level of the Glasgow listeners in sensitivity (clearly seen in both Experiment 1, Figure 3.6, and in Experiment 2, Figure 4.5), suggests that sensitivity to this particular aspect of fine phonetic detail *does not significantly improve* over a long period of time.

Interestingly, response bias for the Glasgow listeners in the Natural exposure condition did not change, suggesting that the knowledge they had about derhoticised /r/ from all their previous experience from living in Glasgow was simply confirmed, or rather supported, by hearing the expected stimuli in the Exposure story. However the Glasgow listeners who heard the Altered (acoustically 'neutralised') stimulus condition changed their bias significantly in the direction of reporting even more *hurt* tokens, mirroring the perceptual hypercorrection seen in the Intermediate listeners. This may be because the lack of difference between the *hut* and *hurt* words in the story altered the listeners' expectations of that particular speaker, believing him to be idiosyncratic in his pronunciations of derhoticised /r/. This may therefore have changed some of the prior beliefs or expectations the listeners had about the speaker, causing them to behave more like the Intermediate listener group.

It is possible that, like the less experienced listeners, these Glasgow listeners reverted to using knowledge about the existence of derhoticised /r/ and the environments in which it can appear, but their sensitivity in identifying exactly *when* the speaker was producing an /r/ was negatively affected by the lack of acoustic difference in the Exposure story. This putative explanation is once again supported by the general principles of Bayesian inference, which suggest that listeners accrue knowledge to build categories, and then match incoming linguistic data to those categories. If those categories were altered during the course of the experiment, it makes sense that the listeners in the Altered exposure condition have trouble when matching incoming phonetic information to the new categories.

Traditional views of exemplar theory may struggle with this, as there is no specific mechanism or structure for categorisation through building up and organising knowledge. More recent exemplar-based models (e.g. Hay & Foulkes 2016) attempt to answer this problem – in a similar vein as hybrid models – by assuming the existence of higher-level categories.

Another comparison to make is between Figure 6.3, which shows the response bias of each listener group to both middle class (left) and working class stimuli (right) in Experiment 1, and Figure 6.4, which shows response bias of each listener group to only working class stimuli, both before hearing the Exposure story (solid lines), and after (dotted lines), in Experiment 2.

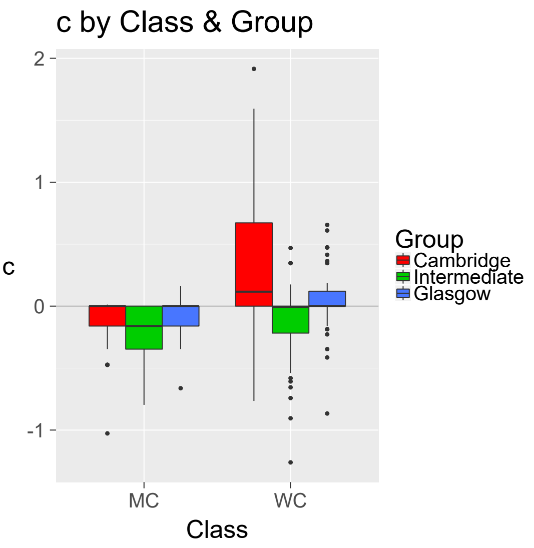


Figure 6.3: Experiment 1: Response bias *c* by Class & Group. Positive values of *c* indicate a bias towards responding HUT.

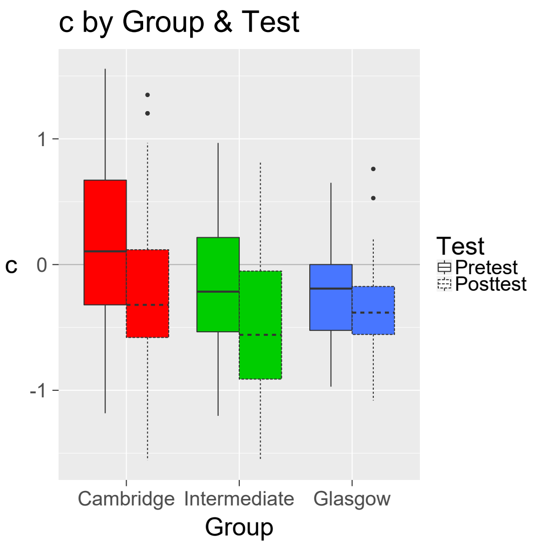


Figure 6.4: Experiment 2: Response bias *c* by Group & Test. Positive values of *c* indicate a bias towards responding CUT.

Comparison of the red box on the right of Figure 6.3 with the red box with solid lines in Figure 6.4, shows the same *positive* bias of Cambridge listeners towards responding ‘hut’ when choosing between *hut* and *hurt* stimuli, in both experiments. This is hypothesised to display their lack of experience with derhoticised /r/, leading them to misclassify both *hut* and *hurt* words as being /r/-less.

In a similar fashion, comparison of the green box on the right of Figure 6.3 with the green box with solid lines in Figure 6.4, shows the same *negative* bias of Intermediate listeners towards responding ‘hurt’ when choosing between *hut* and *hurt* stimuli, in both experiments. This is hypothesised to represent the aforementioned perceptual hypercorrection, whereby roughly three years of living in Glasgow appears to increase listeners' awareness of the existence of derhoticised /r/, but because they know that it has vowel-like formants, their relative lack of sensitivity to difference (compared to the native Glaswegian listeners) causes them to over-report hearing /r/, even when the speaker did not intend to produce one.

These patterns match across the two experiments, which is extremely useful for supporting their validity and replicability. However, close inspection of the third pair of results, relating to the Glasgow listeners, seems to tell a different story. The Glasgow listeners responding to the working class stimuli in Experiment 1 (blue box on the right of Figure 6.3) appear to have very little bias, and in fact are slightly biased towards reporting *hut*. In contrast, the Glasgow listeners responding to the working class stimuli in the Pretest task of Experiment 1 (blue box with solid lines in Figure 6.4) show a relatively strong bias towards reporting *hurt*, in a similar way to the Intermediate listeners (green boxes with solid lines in Figure 6.4).

The explanation for this seemingly odd pattern may be that, while both experiments used Glaswegian speakers, each task in Experiment 2 only presented listeners with one speaker, a working class male, whereas each task in Experiment 1 presented listeners with four speakers randomised together, and these speakers were split across working class and middle class accents. Therefore the tasks in Experiments 1 and 2 are not completely comparable, since they present either one speaker/accent or multiple speakers/accents. A more controlled comparison between these factors is described in Experiment 3.

The Glasgow listeners' negative pretest bias in Figure 6.4 (blue boxes with solid lines), it is similar to the Intermediate listeners' negative bias (green boxes with solid lines), indicating that the Glasgow listeners also appear to be perceptually hypercorrecting when hearing the working class listener, to report more /r/ productions than are actually intended. However, Figure 6.3 shows that the Glasgow listeners are slightly biased towards reporting /r/-less words for the working class speakers (blue boxes on the right), and slightly biased towards reporting /r/-ful words for the middle class speakers (blue boxes on the left). Indeed, these values for bias in Experiment 1 are much closer to zero than any other listeners across both experiments, *including* the Glasgow listeners in Experiment 2 (blue boxes in Figure 6.4), who only heard one speaker.

The fact that the native Glasgow listeners showed least response bias when hearing both accents together suggests that the context of hearing both the middle class and working class /r/ variants assisted them in their perception of derhoticised /r/. This appears to be yet more evidence for the benefit of indexical information in speech processing, as the Glasgow listeners in Experiment 1, who heard a greater range of indexical information (across accents *and* speakers), were the most accurate in their perception of derhoticised /r/ variants. This lends support to exemplar-based theories of speech perception, proponents of which generally claim that listeners compare incoming phonetic and other linguistic information to stored categories, and goodness-of-fit judgements are made for each new exemplar. It is possible that the existence in the same task of lots of indexical information allowed the Glasgow listeners in Experiment 1 to perform goodness-of-fit comparisons armed with much more information in their immediate and very short-term memory than the listeners in Experiment 2 had, hearing as they did only one speaker and accent. Furthermore, Glasgow listeners are the only listener group of the three who demonstrate this apparent benefit from indexical information, so it is likely that the native listeners are able to put their large repository of exemplars of Glaswegian /r/ (of all types, strong and weak) to very good use when categorising new linguistic data.

One possible mechanism of exemplar-based speech perception could be that listeners build up their collection of exemplars for a (e.g.) word, say, *yesterday*, into a most-likely pattern. This could be how listeners construct categories for words or phonemes, over time. The listener also can apply knowledge of things like speech rate, the presence of Lombard speech due to noisy conditions, interference due to the noise itself, and other conditioning factors, to their pattern recognition process. This means that they are acting like the 'ideal adapter' of Kleinschmidt and colleagues' (2015) Bayesian model of speech perception.

This allows for the possibility that, if a listener had for some reason only heard *yesterday* as *yeshay*, a reduced form which usually arises due to fast speech (see Ernestus 2014), they might be unaware that it is indeed a reduced form, and may be surprised when they hear the 'correct' form of *yesterday*. This 'surprise' is arguably evident in, for example, the Cambridge listeners' rapid change in response bias from Pretest to Posttest in Experiment 2. Perhaps more realistically, if the listener had only heard a certain word (or vowel, or consonant, etc.) produced in a certain way by listeners speaking with a particular accent, then they heard a different realisation of that word, they might experience difficulty in recognising it as intended by the speaker. The listener would then have to activate a number of conditioning factors that they may estimate to be in play, that they estimate may be affecting their recognition of the word. This might account for at least some of the extra processing cost in some of the results reported in this thesis. Further speech perception experiments could go some way to identify which components of increased response times can be attributed to particular difficulties a listener might experience with a linguistic phenomenon.

Of course, the conditioning factors that are available to a listener are only those which they have had the need to apply in the past – in other words, if a listener has experience of travelling to many different locations where speakers have different accents, they have a lot of experience with the fact that certain pronunciations vary due to location, in such-and-such a way. Moving to a different accent area may then alter their group of perceived exemplars for that word (though see Evans & Iverson 2007, where this does not reliably happen), with the category perhaps becoming expanded for a while, then changing to the new location. They would presumably still have the old exemplars in their memory, so they would likely still be able to easily adapt to hearing that pronunciation.

6.4 Experiment 3

Experiment 2 built upon the knowledge gained in Experiment 1 by assessing how listeners learned the fine phonetic detail which made discrimination difficult. Both experiments raised very interesting questions surrounding this detail, which would then be specifically addressed in Experiment 3. Both experiments showed that perception and learning of derhoticised /r/ could be very complex, provoking the need for a much more detailed understanding of the online perception of this contrast. Therefore the first research question for Experiment 3 to address was:
‘*How do experienced listeners process ambiguous fine phonetic detail for a phonemic contrast?*’

The second research question was formulated to address the apparent difficulty that arose as the Glaswegian listener group heard the mixed talkers in Experiment 1. It was thought that this may be due to extra cognitive load on the listeners, who may have been working hard to process the identity of the talker as well as making lexical decisions. Thus, the second research question was:
‘*Do harder listening conditions affect the online perception of ambiguous fine phonetic detail for a phonemic contrast?*’

The mouse tracking methodology chosen for this experiment afforded a close inspection of the dynamic changes in online processing. The response variables (i.e. the response buttons at the top left and right of the screen) are outputs which can be analysed in different ways, including signal detection analysis. It was decided early in the analysis stage of this experiment that signal detection would not be conducted. Response time, area under the curve (AUC), and discrete cosine transformation (DCT) analyses were completed.

The significant main effect of Class was found in all of these analyses, again highlighting the extreme difference between the /r/ realisations of middle class and working class Glaswegians, even for native Glaswegian listeners.

The other experimental factor which appeared in all of the analyses, though not always as a main effect, was Blocktype. This suggested that it was easier for listeners to distinguish word pairs if there was only one speaker, than if they were hearing two speakers at once, with two different accents. However a closer look at the response times in Figure 5.3 reveals that listeners experience a higher processing cost when they are distinguishing the middle class stimuli in the Mixed talker block than in the Single talker block, but it also shows that there is no difference between block types for listeners hearing the (presumably harder) working class stimulus pairs. This was surprising, but could be explained by noting that the working class stimuli already had a response time deficit compared to the middle class stimuli, meaning the listeners' responses to the working class stimuli were already suffering in the Single talker block.

The xk3 and yk0 coefficients, which referred to the ‘waviness’ on the x-coordinate dimension, and the mean y-coordinate dimension, respectively, can be taken together to indicate a listener's indecision about the identity of the word. This is because increased ‘waviness’ as reflected in higher xk3 coefficients corresponds to at least one change of direction in the trajectory, and a greater y-coordinate mean as indicated by a greater yk0 coefficient indicates a greater amount of time spent near the top of the screen. If a listener were indecisive about which response to make, they may spend time near the top of the screen, moving between the two response buttons, before making their final decision. This would result in high values for both xk3 and yk0, which is indeed what was found with Class and Blocktype, which were both significant main effects in xk3 and in yk0. These main effects provide further confirmation of the difficulty of both the working class stimuli and of the Mixed talker block.

The difficulty that the listeners experienced with the middle class stimuli in the Mixed block highlights the extra cognitive load that is placed on the perceptual system when more than one talker is speaking in close temporal proximity. Middle class *hut* and *hurt* are ordinarily easy to distinguish, but when the extra load of trial-to-trial variation in speaker identity exists this discrimination suffers, as evidenced by the significantly poorer results in response time and xk0 (mean x-coordinate) for the middle class stimuli in the Mixed block, compared to the Single talker block. This is the case even though no instruction or task about speaker identification was given in the experiment.

These results can be explained by exemplar theory, such that the listeners appear to be processing the identity of the speaker *as well as* making a lexical or phonological decision. This represents another result from this thesis that suggests listeners undertake integrated talker and phoneme processing (Mullennix & Pisoni 1990; see also e.g. Goldinger 1996; Cole, Coltheart & Allard 1974).

6.5 Relation of theory to findings

It is important to now reflect on the pattern of results from all the experiments together, and the relation of the theoretical positions introduced in Chapter 1 to these findings.

Chapter 1 first introduced abstractionist theories, noting that they started to fall short when faced with evidence suggesting that variation is important for speech perception. The general pattern of results from Experiment 3 in this thesis – that is, the consistent main effect of Class in all analyses, with working class stimuli the most challenging – suggest that variation is indeed important for the listener. If it were the case that variation is simply stripped away, converting the signal into discrete categories (e.g. McClelland & Elman 1986; Studdert-Kennedy 1976), or phoneme strings (Halle 1985, cited in Smith 2013: 6), then the variation presented to the listeners in Experiment 3 as a difference in speaker class might be expected to have little effect upon the results of this perceptual experiment. However, it seems that such variation does matter to the listener, so this appears to be further evidence supporting a more nuanced view of speech perception than the abstractionist theories can provide.

The final point in the previous section of this chapter states that exemplar theory may explain some of the results in this thesis, due to the suggestion by some theorists that fine-grained detail about the talker affects perception of linguistic units. The exemplar position put forward by Johnson (1997, cited in Evans and Iverson 2004), as introduced in Chapter 1, suggests that listeners may be able to perform speaker normalisation, comparing the incoming signal with their stored exemplars, with accent normalisation potentially being a similar process (e.g. Nygaard and Pisoni 1998, cited in Evans and Iverson 2004, 2007). In Experiments 1 and 3, which presented the listeners with different speaker classes, there was frequently an effect of class in the results, with the middle class stimuli eliciting better performance from the listeners than the working class stimuli. This suggests that if speaker or accent normalisation takes place it does not do so on an equal basis for all talkers.

Of course, this pattern of results may be entirely in line with the broadest interpretation of the exemplar approach, which holds that incoming exemplars are matched against the inventory of previously stored exemplars, enabling a probabilistic judgement about the phoneme's identity. The fact that the listeners in

Experiments 1 and 3 are better at identifying the words spoken by the middle class Glaswegians is a possible indicator that they have more stored exemplars for the middle class /r/ than for the working class /r/. However, in Experiment 1 this can only be true for the Glasgow and (to a lesser extent) Intermediate listener groups, as the Cambridge listeners probably do not have a large enough inventory of middle class exemplars with which to perform an effective matching process, as their familiarity with Glaswegian is low. Despite their lack of familiarity, the Cambridge listeners performed almost as well as the other groups when responding to middle class stimuli, as can be seen in the red bars with solid outlines in Figure 6.5, which is a repetition of the sensitivity *d'* graph in Figure 3.6, showing the listeners' ability to detect difference between stimulus pairs. A similar pattern of performance between the groups can be seen in the response bias *c* data, in Figure 6.6.

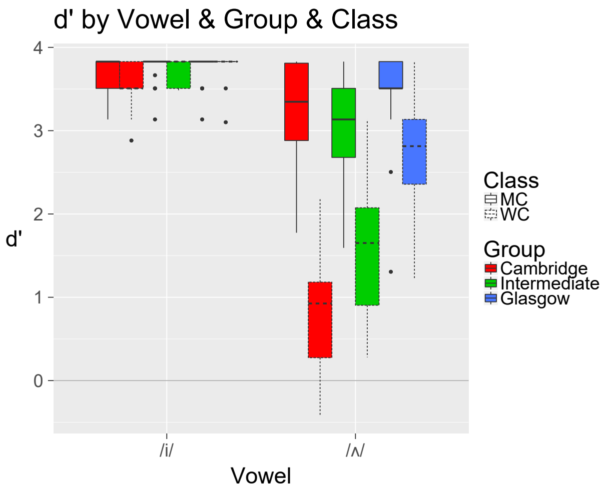


Figure 6.5: Experiment 1 *d'* by Vowel, Group, & Class

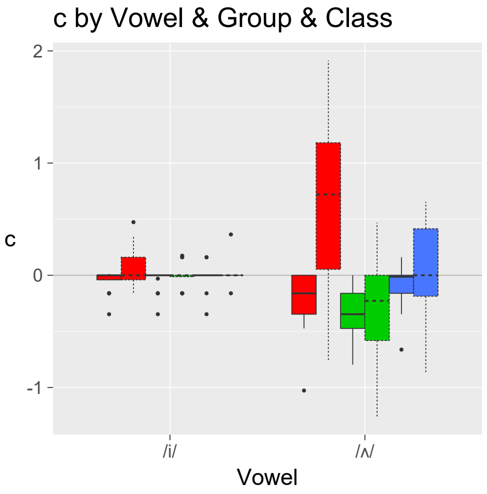


Figure 6.6: Experiment 1 *c* by Vowel, Group, & Class

It is important to remember that accent familiarity is likely to be uneven between middle class and working class Glaswegian (although this is very difficult to measure), leading to different sizes of exemplar inventories for a middle class Glaswegian accent and a working class Glaswegian accent. If it were the case that the size of the inventory of stored exemplars is responsible for the listeners' ability to match and identify the stimuli, then the difference between the *d'* for middle class stimuli (solid outlines, high performance) and working class stimuli (dotted outlines, poorer performance) in Figure 6.5 would support this. However, the Cambridge listeners' high performance with the middle class stimuli is once again evidence against this position, as is the high performance of all listener groups when responding to /i/ stimuli, compared with /ʌ/ stimuli. It therefore seems very likely that the main factor contributing to the pattern of performance seen in Experiment 1 is the relative acoustic similarity of the working class /ʌ/ stimulus pairs, compared with the middle class /ʌ/ pairs as well as the /i/ pairs for both classes.

It is also possible to interpret these results in terms of hybrid approaches, which hold that abstract, symbolic representations exist alongside talker-specific and other indexical information (Schacter & Church 1992, cited in McQueen 2005: 264). It could be the case that the Cambridge listeners in Experiment 1 are mapping the incoming exemplars of e.g. [hʌˤt] (working class *hurt*) to the category /hʌt/, which constitutes their abstract representation for the form. The fact that they also correctly categorise [hʌt] (working class *hut*) as /hʌt/ may explain their very low sensitivity to stimulus difference between WC *hut/hurt* words and their extreme bias to classifying them all as /r/-less (dotted red boxes on the right of both Figure 6.5 & 6.6).

A hybrid interpretation may also explain the trend for the Cambridge group to shift their bias towards reporting more WC *hut/hurt* items as /hʌrt/, as seen in the change between the red pretest box (solid outline) and the posttest box (dotted outline) in Figure 6.4. If it is the case that speaker-specific detail is stored by the listener when identifying words, the Cambridge listeners (and to an extent the Intermediate listeners, shown by the green boxes in Figure 6.4) may be using the speaker-specific information they heard in the exposure story in order to refine and encode the input from this particular speaker. This may then enable the listeners to more effectively assign the incoming [hʌt] and [hʌˤt] tokens to /hʌt/ and /hʌrt/ respectively, improving their ability to detect the phonemic contrast and apply it to new instances from this speaker. Not only is it possible that the listeners have begun to systematically associate allophonic details with words for this speaker (Pierrehumbert 2002: 19), but they may also have used the exposure story to improve the process of pattern-matching between signal and memories, using the rich hierarchical structures proposed by the Polysp model (Hawkins & Smith 2001; Hawkins 2003, 2010; Smith 2015).

The other theoretical stance introduced in Chapter 1 was the Bayesian approach to speech perception, which suggests that the listener makes decisions regarding the identity of e.g. phonemes, based upon a combination of evidence from the speech signal and their knowledge or expectation of which phonemes they are likely to hear (e.g. Smith 2013). One application of the Bayesian approach is Kleinschmidt and colleagues' ideal adapter model (Kleinschmidt & Jaeger 2015, 2018), whereby the listener has a set of prior beliefs about linguistic patterns, which are updated when new information arises.

Experiment 2 directly tested for the effects of short term exposure on the listeners' ability to perceive derhoticised /r/, so it is possible to examine the results in terms of a Bayesian approach. Figure 6.7 (a repetition of Figure 4.3 from Chapter 4), shows the three way interaction of response time (log(rt)) by Coda, Group, & Test.

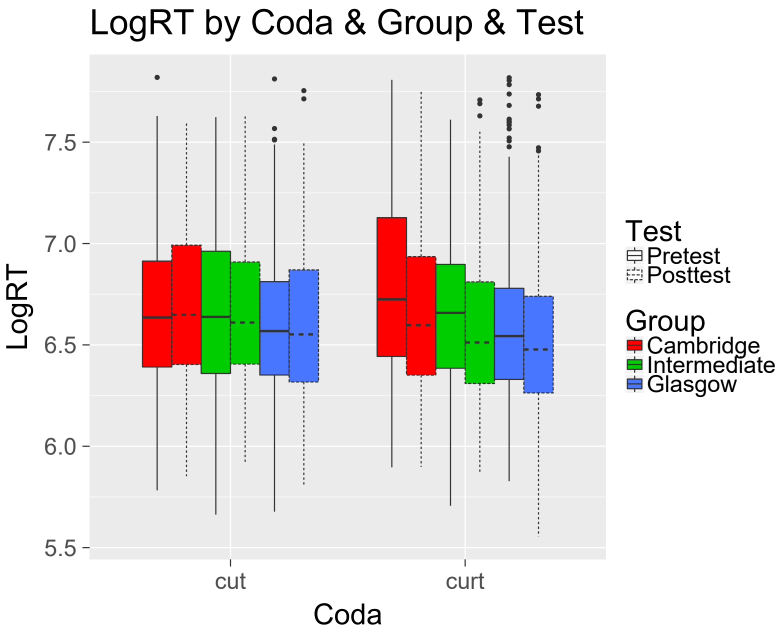


Figure 6.7: Experiment 2 log(rt) for responses to correct stimuli, by Coda, Group, & Test

In this interaction, there were no significant differences from pretest to posttest for any of the listener groups for the *cut* words, shown on the left of the graph, but all groups get significantly faster in posttest when responding to the *curt* words, shown on the right (recall that listeners did not hear the lexical items *hut* or *hurt* in the pretest or posttest, but did hear *hut* and *hurt*, and other similarly-structured items, in the exposure story). This shows that there is an imbalance in the way that the listeners update their expectations about derhoticised-r /ʌˤ/, compared with the plain back vowel /ʌ/. It seems that the listeners in Experiment 2 have begun to adapt by learning from the context of the story that (for this speaker) pharyngealisation means postvocalic /r/, but no such updating of the listeners' prior probability distribution was required for the plain vowel /ʌ/, as the speaker's productions of that phoneme fit their already-established expectation for /ʌ/.

This updating process may explain the improvement in the listeners' response time for this particular phonetic feature, as they begin to form a set of expectations for the phonemic inventory exhibited by this speaker, and possibly also for the working class Glaswegian accent. On this point, it should again be noted that the listener groups have all heard this particular speaker for the same amount of time, but it is probable that their different long-term experience affects this speaker recognition in different ways. This could explain the fact that, unlike the Glasgow and Intermediate listener groups, the Cambridge listeners show a trend for slightly slower responses to the *cut* stimuli in posttest (compare solid (pretest) and dotted (posttest) red boxes on the left of Figure 6.7). It may be the case that, due to their lack of experience with Glaswegian speakers, the Cambridge listeners are more 'unstable' in their perception of the phonemic categories associated with this speaker, possibly because there are many unfamiliar accent features to contend with. By extension, the Glasgow listeners (and to a lesser degree the Intermediate listeners) have more exemplars of the working class Glaswegian accent than the Cambridge listeners, so their categories are more stable (this interpretation is also in line with hybrid models). It would be interesting to explore this issue in terms of Kleinschmidt and colleagues' suggestion that there may be a deep connection between sociolinguistics and psycholinguistics. However, as no explicit Bayesian analysis has been done in this thesis, no further comment can be confidently made about the patterns of results presented here, other than these brief points regarding short term listener experience.

6.6 Conclusion

This thesis has presented a set of experimental results, providing insight into how listeners perceive a complex and changing speech sound, derhoticised /r/. It appears to be the case that the majority of these results are best explained by exemplar theory, as they suggest that all listener groups in Experiments 1 and 2 display a degree of plasticity in their perception of phonemic categories, and the results of Experiment 2 appear to show that category changes can be relatively rapid. This is at odds with traditionally abstractionist views, as they may not allow for such a high degree of short term variability in their (presumably fairly rigid) categories. These results may also be explained by certain hybrid models of perception, which claim that there *are* abstract categories at some level, but the speaker and indexical information cannot be ignored.

In any case, these results are an argument against the abstractionist view that information about the talker is 'stripped away' and discarded to reveal the underlying abstract representations alone, as seen in the results of Experiment 3, which showed phoneme processing costs associated with hearing multiple talkers. This also likely applies to other indexical information such as the listeners' knowledge about how social class affects language in Glasgow (particularly so for the native Glasgow listeners), which could also play an important role in the increased response times or skewed response bias in the experiments where listeners were presented with more than one talker.

Bayesian inference may also go a long way to explaining the present results, as the theory could be applied in a flexible manner with regard to the existence of categories. One way of directly testing this would be to use the data that already exists from this project, and undertake an analysis based on the principles of Bayesian inference (e.g. Kleinschmidt & Jaeger 2015, Levitin 2016).

6.7 Future directions

A question for future research was alluded to earlier in this chapter, in the section on Experiment 1. In discussing the social aspect of speech perception, Smith asserts that listeners ‘are not mere passive receivers of information’, but fulfil a more active role (2013: 10). Since listeners are indeed very likely to actively apply at least some part of their own production/perception mechanism when they perceive speech, this could be tested and controlled for in future experiments. It will be extremely useful to apply the factor of characteristics including social class to the design of this type of perception experiment, so that the perception of working class speech by working class listeners can be compared to the perception of the same stimuli by middle class listeners. Of course, this may still not get to the question of the effect of a listener's ‘own accent’ on their perception of similar/dissimilar speech, but it would certainly provide much more detail about the role of experience. This type of experimental design has been successfully undertaken by e.g. Clopper 2017, who used multiple experimental factors to answer the complex question of dialect familiarity.

Another possible direction could be the investigation of real-time processing as a listener encounters new, unfamiliar speech stimuli. As the methodology for Experiment 3 was new to the researcher, the design was reduced from the three listener groups in the first two experiments to just one, for simplicity. Nevertheless, valuable information may be gained about online learning of derhoticised /r/ in future research. McQueen writes that, due to apparent feedback for learning, ‘the question for future research will be whether apparent demonstrations of feedback in on-line processing (i.e. feedback as the word is being heard) are in fact the result of longer-term learning effects, or are indeed true on-line effects that might arise epiphenomenally, that is, as a consequence of the need for feedback for perceptual learning.’ (2007: 47). A combination of the methodologies of Experiments 2 and 3 may therefore be a fruitful direction of future research into derhoticised /r/, to help address this issue.

Finally, it is known from other research into cross-dialect perception that cross-modal priming can have an effect on a listener's perception of an accent (e.g. Niedzielski 1999; Hay, Warren & Drager 2006; Hay, Nolan & Drager 2006; Hay & Drager 2010; Koops, Gentry & Pantos 2008; Robertson 2015; but see Lawrence 2015). It is therefore sensible to consider experimental designs in which the listener attends to speaker specific and other indexical information, as well as other stimuli such as visual clues to a speaker's class.

The wealth of data that has been presented in this thesis has only scratched the surface of the potential for perceptual information which can be explored using derhoticised /r/ in Glasgow. However, care must be taken when designing perceptual experiments. Smith (2013) writes that a paradox may ensue as various speech perception phenomena are better understood. If this increased understanding arises through the use of carefully produced and prepared (sometimes artificial) stimuli, the more researchers may wish to test for the complex effects of the perception of natural coarticulatory effects in speech, among other speech phenomena – in other words, the more ecologically valid the experiments must become. This issue must remain at the heart of the design of any behavioural experiment, in order for it to remain informative, yet relevant to the perception of speech in the real world.

Bibliography

Adank, P., Evans, B., Stuart-Smith, J., & Scott, S. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 520–529.

Adank, P. & McQueen, J. (2007). The effect of an unfamiliar regional accent on spoken-word comprehension. *Proceedings of the 16th International Congresss of Phonetic Sciences*, 1925–1928.

Aitken, A. & McArthur, T. (1979). *Languages of Scotland.* Edinburgh: Chambers.

Ashton, L. (2011). *Perception of /r/ in Scottish English.* Undergraduate Dissertation, Queen Margaret University.

Barden, K. & Hawkins, S. (2013). Perceptual learning of phonetic information that indicates morphological structure. *Phonetica, 70*(4), 323–342.

Beddor, P., McGowan, K., Boland, J., Coetzee, A., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America, 133*(4), 2350–66.

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 171–204.

Bladon, A. (1983). Two-formant models of vowel perception: Shortcomings and enhancement. *Speech Communication, 2*(4), 305–313.

Boersma, P. (2006). Praat: Doing phonetics by computer. *praat.org.*

Bond, A. (2013). *The phonetics and phonology of coda /r/ in Scottish English.* Masters Thesis, University of Cambridge.

Braber, N. & Butterfint, Z. (2008). Local identity and sound change in Glasgow: A pilot study. *Leeds Working Articles in Linguistics*, 22–43.

Bradlow, A., Akahane-Yamada, R., Pisoni, D., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Attention, Perception, & Psychophysics, 61*(5), 977–985.

Bradlow, A. & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*(2), 707–729.

Bradlow, A., Pisoni, D., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America, 101*(4), 2299–2310.

Brato, T. (2012). *A sociophonetic study of Aberdeen English: Innovation and conservatism.* Doctoral Thesis, University of Giessen.

Brown, J. & Carr, T. (1993). Limits on perceptual abstraction in reading: Asymmetric transfer between surface forms differing in typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(6), 1277.

Carey, E. (2010). *Cross accent perception of Standard Southern British English and Glasgow English.* Undergraduate Dissertation, University of Glasgow.

Carr, T., Brown, J., & Charalambous, A. (1989). Repetition and reading: Perceptual encoding mechanisms are very abstract but not very interactive. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(5), 763.

Chambers, J. (2008). *Sociolinguistic Theory*. Language in Society. Hoboken, NJ: Wiley-Blackwell.

Chistovich, L. & Lublinskaya, V. (1979). The center of gravity effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli. *Hearing Research, 1*(3), 185–195.

Chomsky, N. & Halle, M. (1968). *The sound pattern of English.* New York, NY: Harper & Row.

Cisek, P. & Kalaska, J. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action. *Neuron, 45*(5), 801–814.

Clarke, C. & Luce, P. (2005). Perceptual adaptation to speaker characteristics: VOT boundaries in stop voicing categorization. *Proceedings of ISCA Workshop on Plasticity in Speech Perception.*

Clayards, M., Tanenhaus, M., Aslin, R., & Jacobs, R. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition, 108*(3), 804–809.

Clopper, C. (2017). Dialect interference in lexical processing: Effects of familiarity and social stereotypes. *Phonetica, 74*(1), 25–59.

Clopper, C., Pierrehumbert, J., & Tamati, T. (2010). Lexical neighborhoods and phonological confusability in cross-dialect word recognition in noise. *Laboratory Phonology, 1*(1), 65–92.

Clopper, C. & Pisoni, D. (2004). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics, 32*(1), 111–140.

Clopper, C., Rohrbeck, K., & Wagner, L. (2012). Perception of dialect variation by young adults with high-functioning autism. *Journal of Autism and Developmental Disorders, 42*(5), 740–754.

Cole, R., Coltheart, M., & Allard, F. (1974). Memory of a speaker's voice: Reaction time to same- or different-voiced letters. *The Quarterly Journal of Experimental Psychology, 26*(1), 1–7.

Creel, S., Aslin, R., & Tanenhaus, M. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition, 106*(2), 633–664.

Creel, S. & Tumlin, M. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language, 65*(3), 264–285.

Creelman, C. & Macmillan, N. (1979). Auditory phase and frequency discrimination: A comparison of nine procedures. *Journal of Experimental Psychology: Human Perception and Performance, 5*(1), 146–156.

Dahan, D., Drucker, S., & Scarborough, R. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition, 108*(3), 710–718.

Dahan, D. & Mead, R. (2010). Context-conditioned generalization in adaptation to distorted speech. *Journal of Experimental Psychology: Human Perception and Performance, 36*(3), 704–728.

Dalmasso, D. (2012). *Diachronic change in the postvocalic /r/ in the Dutch of Amsterdam.* Masters Thesis, University of Amsterdam.

Delattre, P. & Freeman, D. (1968). A dialect study of American r's by x-ray motion picture. *Linguistics, 6*(44), 29–68.

Dimopoulou, T. (2014). *The role of discriminant reinforcement in task-irrelevant perceptual learning of acoustic-phonetic categories.* Doctoral Thesis, Athens University of Economics and Business.

Doyle, M. & Walker, R. (2001). Curved saccade trajectories: Voluntary and reflexive saccades curve away from irrelevant distractors. *Experimental Brain Research, 139*(3), 333–344.

Eisner, F. & McQueen, J. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America, 119*(4), 1950–1953.

Eisner, F., Melinger, A., & Weber, A. (2013). Constraints on the transfer of perceptual learning in accented speech. *Frontiers in Psychology, 4*.

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua, 142*, 27–41.

Espy-Wilson, C., Boyce, S., Jackson, M., Narayanan, S., & Alwan, A. (2000). Acoustic modeling of American English /r/. *The Journal of the Acoustical Society of America, 108*(1), 343–356.

Evans, B. & Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in Northern and Southern British English sentences. *The Journal of the Acoustical Society of America, 115*(1), 352–361.

Evans, B. & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America, 121*(6), 3814–3826.

Farmer, T., Anderson, S., & Spivey, M. (2007). Gradiency and visual context in syntactic garden-paths. *Journal of Memory and Language, 57*(4), 570–595.

Farmer, T., Liu, R., Mehta, N., & Zevin, J. (2009). Native language experience influences the perceived similarity of second language vowel categories. *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*, 2588–2593.

Feldman, N., Griffiths, T., & Morgan, J. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review, 116*(4), 752.

Flege, J. (1995). Second language speech learning: Theory, findings, and problems. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research, 92*, 233–277.

Floccia, C., Butler, J., Goslin, J., & Ellis, L. (2009). Regional and foreign accent processing in English: Can listeners adapt? *Journal of Psycholinguistic Research, 38*(4), 379–412.

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance, 32*(5), 1276–1293.

Forster, K. & Forster, J. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, 35*(1), 116–124.

Fowler, C. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Status Report on Speech Research*, 139–169.

Fowler, C. & Rosenblum, L. (1991). The perception of phonetic gestures. *Modularity and the Motor Theory of Speech Perception, 335*, 33–59.

Franco-Watkins, A. & Johnson, J. (2011). Applying the decision moving window to risky choice: Comparison of eye-tracking and mousetracing methods. *Judgment and Decision Making, 6*(8), 740.

Freeman, J. & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods, 42*(1), 226–241.

Garrett, P., Coupland, N., & Williams, A. (1999). Evaluating dialect in discourse: Teachers' and teenagers' responses to young English speakers in Wales. *Language in Society, 28*(3), 321–354.

Goldinger, S. (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.

Goldinger, S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251–279.

Goldinger, S. (2000). The role of perceptual episodes in lexical processing. *ISCA Tutorial and Research Workshop (ITRW) on Spoken Word Access Processes*, 2–5.

Goldinger, S. (2007). A complementary-systems approach to abstract and episodic speech perception. *Proceedings of the 16th International Congress of Phonetic Sciences*, 49–54.

Green, K., Kuhl, P., Meltzoff, A., & Stevens, E. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. *Attention, Perception, & Psychophysics, 50*(6), 524–536.

Halle, M. (1985). Speculations about the representation of words in memory. *Phonetic Linguistics*, 101–114.

Harrington, J. (2010). *Phonetic Analysis of Speech Corpora*. Hoboken, NJ: Wiley-Blackwell.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics, 31*(3), 373–405.

Hawkins, S. (2010). Phonetic variation as communicative system: Perception of the particular and the abstract. *Laboratory Phonology, 10*, 479–510.

Hawkins, S. & Smith, R. (2001). Polysp: A polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics, 13*, 99–188.

Hay, J. & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics, 48*(4), 865–892.

Hay, J. & Foulkes, P. (2016). The evolution of medial /t/ over real and remembered time. *Language, 92*(2), 298–330.

Hay, J. & Maclagan, M. (2012). /r/-sandhi in early 20th century New Zealand English. *Linguistics, 50*, 745–763.

Hay, J., Maclagan, M., Preston, D., & Niedzielski, N. (2010). Social and phonetic conditioners on the frequency and degree of 'intrusive/r/' in New Zealand English. *A Reader in Sociophonetics, 219*, 41–69.

Hay, J., Nolan, A., & Drager, K. (2006). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review, 3*, 351–379.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics, 34*(4), 458–484.

Hayward, K. (2014). *Experimental phonetics: An introduction*. London: Routledge.

Heeger, D. (1998). Signal detection theory. *cns.nyu.edu/david/handouts/sdt/sdt.html*.

Heselwood, B. (2009). Rhoticity without F3: Lowpass filtering, F1-F2 relations and the perception of rhoticity in 'NORTH-FORCE', 'START' and 'NURSE' words. *Leeds Working Papers in Linguistics & Phonetics*, *14*, 49–64.

Heselwood, B. & Plug, L. (2011). The role of F2 and F3 in the perception of rhoticity: Evidence from listening experiments. *Proceedings of the 17th International Congress of Phonetic Sciences*, 867–870.

Heselwood, B., Plug, L., & Tickle, A. (2010). Assessing rhoticity using auditory, acoustic and psycho-acoustic methods. *Proceedings of Methods XIII*.

Huettig, F. & McQueen, J. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482.

Huettig, F., Rommers, J., & Meyer, A. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171.

Jackson, A. & Morton, J. (1984). Facilitation of auditory word recognition. *Memory & Cognition*, *12*(6), 568–574.

Jauriberry, T., Sock, R., Hamm, A., & Pukli, M. (2012). Rhoticité et dérhoticisation en Anglais Écossais d'Ayrshire. *Proceedings of the Joint Conference JEP-TALN-RECITAL*, *1*(1), 89–96.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. *Talker Variability in Speech Processing*, 145–165.

Johnson, K., Strand, E., & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, *27*(4), 359–384.

Johnston, P. (1997). Regional variation. In Jones, C. (Ed). *The Edinburgh History of the Scots Language. Edinburgh: EUP*, 433–513.

Jones, C. (1989). *A history of English phonology*. London: Longman.

Kleinschmidt, D. & Jaeger, T. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203.

Kleinschmidt, D., Weatherholtz, K., & Jaeger, T. (2018). Sociolinguistic perception as inference under uncertainty. *Topics in Cognitive Science*, 1–18.

Koops, C., Gentry, E., & Pantos, A. (2008). The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *University of Pennsylvania Working Papers in Linguistics*, *14*(2), 91–101.

Kraljic, T. & Samuel, A. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, *51*(2), 141–178.

Kraljic, T. & Samuel, A. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review, 13*(2), 262–268.

Kuhl, P. (1991). Human adults and human infants show a perceptual magnet effect for the prototypes of speech categories, monkeys do not. *Attention, Perception, & Psychophysics, 50*(2), 93–107.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). LmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13).

Labov, W. (1986). The social stratification of (R) in New York City department stores. *Dialect and Language Variation*, 304–329.

Labov, W. (1994). *Principles of linguistic change. Vol. 1: Internal features*. Oxford: Blackwell.

Labov, W. (2001). *Principles of linguistic change. Vol. 2: Social factors*. Oxford: Blackwell.

Ladefoged, P. (1975). *A course in phonetics*. New York: Harcourt Brace Jovanovich.

Ladefoged, P. (2003). *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Hoboken, NJ: Wiley-Blackwell.

Lass, R. (1997). *Historical linguistics and language change*. Cambridge: CUP.

Lawrence, D. (2015). Limited evidence for social priming in the perception of the BATH and STRUT vowels. *Proceedings of the 18th International Congress of Phonetic Sciences*.

Lawson, E., Scobbie, J., & Stuart-Smith, J. (2011a). A single case study of articulatory adaptation during acoustic mimicry. *Proceedings of the 17th International Congress of Phonetic Sciences*.

Lawson, E., Scobbie, J., & Stuart-Smith, J. (2011b). The social stratification of tongue shape for postvocalic /r/ in Scottish English. *Journal of Sociolinguistics*, 256–268.

Lawson, E., Scobbie, J., & Stuart-Smith, J. (2013). Bunched /r/ promotes vowel merger to schwar: An ultrasound tongue imaging study of Scottish sociophonetic variation. *Journal of Phonetics, 41*(3-4), 198–210.

Lawson, E., Scobbie, J., & Stuart-Smith, J. (2014). A socio-articulatory study of Scottish rhoticity. In Lawson, R. (Ed). *Sociolinguistics in Scotland*, 53–78.

Lawson, E., Stuart-Smith, J., & Scobbie, J. (2008). Articulatory insights into language variation and change: Preliminary findings from an ultrasound study of derhoticization in Scottish English. *University of Pennsylvania Working Papers in Linguistics, 14*(2), 102–110.

Lawson, E., Stuart-Smith, J., & Scobbie, J. (2017). The role of gesture delay in coda /r/ weakening: An articulatory, auditory and acoustic study. *Journal of the Acoustical Society of America, 143*(3), 1646–1657.

Lawson, E., Stuart-Smith, J., Scobbie, J., Yaeger-Dror, M., & Maclagan, M. (2010). Liquids. In M. Di Paolo & M. Yaeger-Dror (Eds.) *Sociophonetics: A student's guide*, 72–86.

Lennon, R. (2012). A real-time sociophonetic study of postvocalic /r/ in the speech of schoolchildren in Bearsden. *Undergraduate Dissertation, University of Glasgow*.

Lennon, R. (2013). The effect of experience in cross-dialect perception: Parsing /r/ in Glaswegian. *Masters Thesis, University of Glasgow*.

Levitin, D. (2016). *A field guide to lies: Critical thinking in the information age*. Dutton.

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

Liberman, A. & Mattingly, I. (1989). A specialization for speech perception. *Science*, *243*(4890), 489–494.

Lindau, M. (1978). Vowel features. *Language*, *54*(3), 541–563.

Lindau, M. (1985). The story of /r/. In V. Fromkin (Ed). *Phonetic Linguistics: Essays in Honor of Peter Ladefoged. Orlando: Academic Press*, 157–168.

Logan, J., Lively, S., & Pisoni, D. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–886.

Lohse, G. & Johnson, E. (1996). A comparison of two process tracing methods for choice tasks. *Organizational Behavior and Human Decision Processes*, *68*(1), 28–43.

Lourido, G. T. & Evans, B. (2015). Switching language dominance for ideological reasons: A study of Galician new speakers speech production and perception. *Proceedings of the 18th International Congress of Phonetic Sciences*.

Macafee, C. (1983). *Glasgow*. John Benjamins.

Macaulay, R. (1976). Social class and language in Glasgow. *Language in Society*, *5*(2), 173–188.

Macaulay, R. (2005). *Extremely common eloquence: Constructing Scottish identity through narrative*. Amsterdam: Rodopi.

MacFarlane, A. & Stuart-Smith, J. (2012). One of them sounds sort of Glasgow Uni-ish. Social judgements and fine phonetic variation in Glasgow. *Lingua*, *122*(7), 764–778.

Macmillan, N. & Creelman, C. (2005). *Detection theory: A user's guide*. Mahwah, NJ: Lawrence Erlbaum.

Maddieson, I. & Ladefoged, P. (1996). *The sounds of the world's languages*. Oxford: Blackwell.

Maye, J., Aslin, R., & Tanenhaus, M. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*(3), 543–562.

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

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

McMurray, B., Clayards, M., Tanenhaus, M., & Aslin, R. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review*, *15*(6), 1064–1071.

McQueen, J. (2005). Speech perception. *The Handbook of Cognition*, 255–275.

McQueen, J. (2007). Eight questions about spoken-word recognition. In M. G. Gaskell (Ed.) *The Oxford Handbook of Psycholinguistics. Oxford: OUP*, 37–53.

McQueen, J., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*(6), 1113–1126.

McQueen, J. & Viebahn, M. (2007). Tracking recognition of spoken words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology*, *60*(5), 661–671.

Mielke, J., Baker, A., & Archangeli, D. (2006). Covert /r/ allophony in English: Variation in a socially uninhibited sound pattern. *Oral paper at LabPhon 10.*

Miller, J., Connine, C., Schermer, T., & Kluender, K. (1983). A possible auditory basis for internal structure of phonetic categories. *The Journal of the Acoustical Society of America*, *73*(6), 2124–2133.

Mugglestone, L. (2003). *Talking proper: The rise of accent as social symbol.* Oxford: OUP.

Mullennix, J. & Pisoni, D. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, *47*(4), 379–390.

Mullennix, J., Pisoni, D., & Martin, C. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, *85*(1), 365–378.

Munro, M. & Derwing, T. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*(3), 289–306.

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

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*(2), 132–142.

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

Norris, D., McQueen, J., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238.

Norris, D., McQueen, J., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, *31*(1), 4–18.

Nygaard, L. & Pisoni, D. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*(3), 355–376.

Nygaard, L., Sommers, M., & Pisoni, D. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42–46.

Ohala, J. (1993). The phonetics of sound change. *Historical Linguistics: Problems and Perspectives*, 237–278.

Ohala, J. (1996). Speech perception is hearing sounds, not tongues. *The Journal of the Acoustical Society of America*, *99*(3), 1718–1725.

Palmeri, T., Goldinger, S., & Pisoni, D. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 309–328.

Paton, K. (2009). Probing the symptomatic silences of middle-class settlement: A case study of gentrification processes in Glasgow. *City*, *13*(4), 432–450.

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.) *Laboratory phonology VII. Berlin: Mouton de Gruyter*, 101–40.

Pierrehumbert, J. (2006). The next toolkit. *Journal of Phonetics*, *34*(4), 516–530.

Pisoni, D., Lively, S., & Logan, J. (1994). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In J. C. Goodman & H. C. Nusbaum (Eds.) *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words. Cambridge, MA: MIT Press*, 121–166.

Plug, L. & Ogden, R. (2003). A parametric approach to the phonetics of postvocalic /r/ in Dutch. *Phonetica*, *60*(3), 159–186.

Rathcke, T., Stuart-Smith, J., Timmins, C., & José, B. (2012). Trying on a new BOOT: Acoustic analyses of real-time change in Scottish English. *Poster presented at NWAV*, *41*, 26.

Robertson, D. (2015). *Implicit cognition and the social evaluation of speech.* Doctoral Thesis, University of Glasgow.

Romaine, S. (1978). Postvocalic /r/. *Scottish English: Sound Change in Progress*, 144–157.

Salverda, A. & Tanenhaus, M. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1108.

Samuel, A. (1982). Phonetic prototypes. *Attention, Perception, & Psychophysics*, *31*(4), 307–314.

Schacter, D. & Church, B. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 915.

Scharenborg, O., Norris, D., Bosch, L., & McQueen, J. (2005). How should a speech recognizer work? *Cognitive Science*, *29*(6), 867–918.

Scobbie, J., Sebregts, K., & Stuart-Smith, J. (2009). Dutch rhotic allophony, coda weakening, and the phonetics-phonology interface. *QMU Speech Science Research Centre Working Paper (QMU, Edinburgh)*, 3–24.

Smith, R. (2013). New directions in speech perception. In R. A. Knight and M. J. Jones (Eds.) *Bloomsbury Companion to Phonetics. London: Bloomsbury.*

Smith, R. (2015). Perception of speaker-specific phonetic detail. In S. Fuchs, D. Pape, C. Petrone and P. Perrier (Eds.) *Individual Differences in Speech Production and Perception. Peter Lang,* 11–38.

Smith, R. & Hawkins, S. (2012). Production and perception of speaker-specific phonetic detail at word boundaries. *Journal of Phonetics*, *40*(2), 213–233.

Sóskuthy, M. (2014). Formant Editor: Software for editing dynamic formant measurements. *github.com/soskuthy/formantedit.*

Spivey, M., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America, 102*(29), 10393–10398.

Strand, E. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology, 18*(1), 86–100.

Stuart-Smith, J. (1999). Glasgow: Accent and voice quality. In P. Foulkes and G. J. Docherty (Eds.) *Urban Voices: Accent Studies in the British Isles. Leeds: Arnold.*

Stuart-Smith, J. (2003). The phonology of Modern Urban Scots. In J. Corbett, D. McClure, J. Stuart-Smith (Eds.) *The Edinburgh Companion to Scots. Edinburgh: EUP*, 110–137.

Stuart-Smith, J. (2007). A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents. *Proceedings of the 16th International Congress of Phonetic Science,* 1449–1452.

Stuart-Smith, J. (2016). Social dynamics and phonological representations: Observations from speech and society in Scotland. In *Proceedings of the 15th international congress of laboratory phonology*.

Stuart-Smith, J., Lawson, E., & Scobbie, J. (2014). Derhoticisation in Scottish English: A sociophonetic journey. In C. Celata and S. Calamai (Eds.) (Advances in Sociophonetics), 57–94.

Stuart-Smith, J., Timmins, C., & Tweedie, F. (2007). 'Talkin' Jockney'? Variation and change in Glaswegian accent. *Journal of Sociolinguistics, 11*(2), 221–260.

Studdert-Kennedy, M. (1976). Speech perception. In N. J. Lass (Ed.) *Contemporary Issues in Experimental Phonetics. New York, NY: Academic Press,* 243–293.

Sulpizio, S., Fasoli, F., Maass, A., Paladino, M., Vespignani, F., Eyssel, F., & Bentler, D. (2015). The sound of voice: Voice-based categorization of speakers' sexual orientation within and across languages. *PloS one*, *10*(7), e0128882.

Sumner, M. & Samuel, A. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language, 60*(4), 487–501.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632–1634.

Team, R. D. C. (2013). R: A language and environment for statistical computing.

Tipper, S., Howard, L., & Jackson, S. (1997). Selective reaching to grasp: Evidence for distractor interference effects. *Visual Cognition, 4*(1), 1–38.

Trudgill, P. (1986). *Dialects in contact.* Oxford: Blackwell.

Tuinman, A., Mitterer, H., & Cutler, A. (2011). Perception of intrusive /r/ in English by native, cross-language and cross-dialect listeners. *The Journal of the Acoustical Society of America, 130*(3), 1643–1652.

Twist, A., Baker, A., Mielke, J., & Archangeli, D. (2007). Are 'covert' /r/ allophones really indistinguishable? *University of Pennsylvania Working Papers in Linguistics, 13*(2), 207–216.

Valbret, H., Moulines, E., & Tubach, J. (1992). Voice transformation using PSOLA technique. *Speech Communication, 11*(2-3), 175–187.

van Bezooijen, R. & Gooskens, C. (1999). Identification of language varieties: The contribution of different linguistic levels. *Journal of Language and Social Psychology, 18*(1), 31–48.

Video, L. (2013). Vlc media player.

Walker, A. & Hay, J. (2011). Congruence between word age and voice age facilitates lexical access. *Laboratory Phonology, 2*(1), 219–237.

Watson, C. & Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in Australian English vowels. *The Journal of the Acoustical Society of America, 106*(1), 458–468.

Watt, D., Llamas, C., & Johnson, D. (2010). Levels of linguistic accommodation across a national border. *Journal of English Linguistics*, *38*(3), 270–289.

Wells, J. (1982). *Accents of English.* Cambridge: CUP.

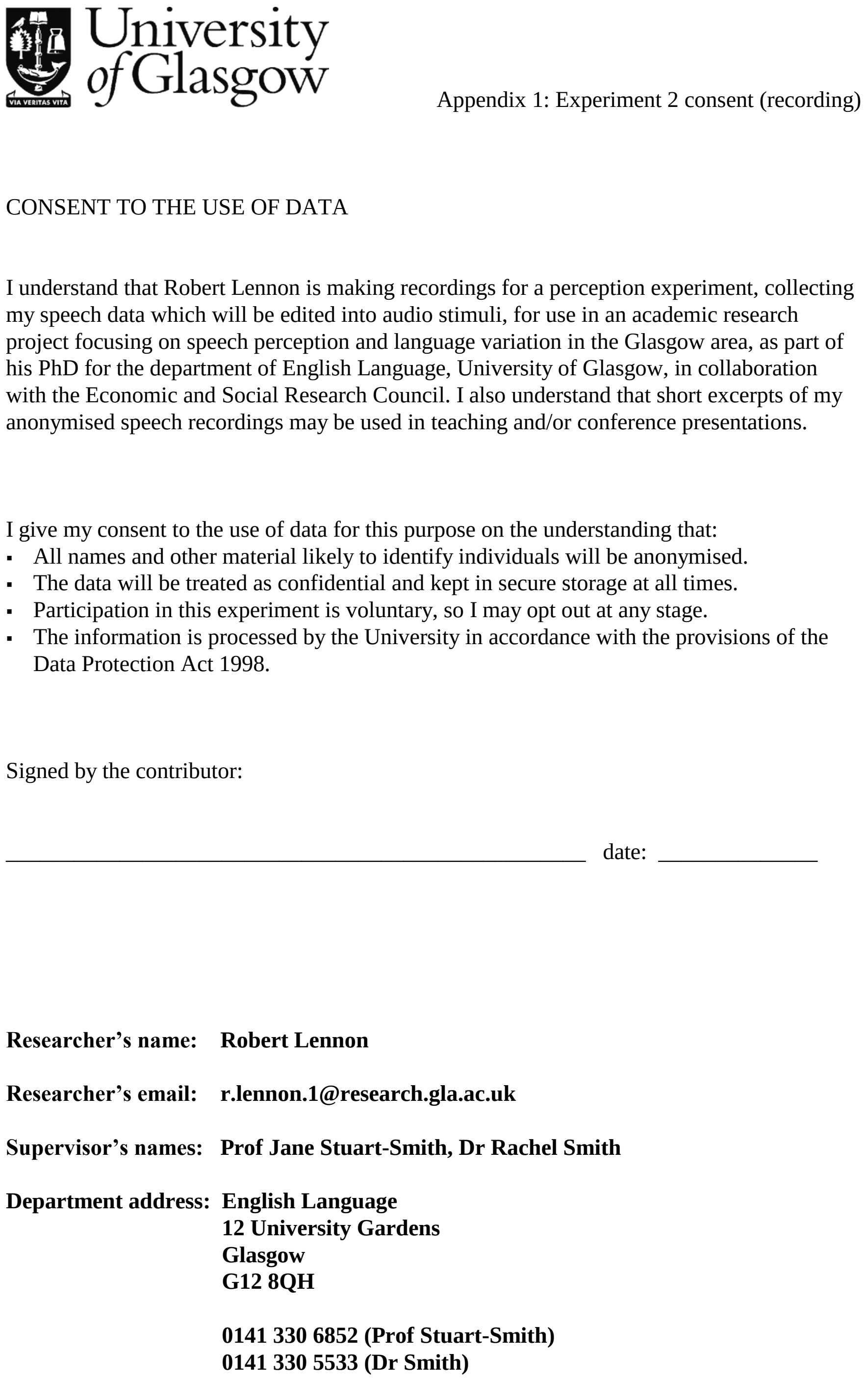
Wrench, A. (2007). Articulate Assistant Advanced user guide: version 2.07. *Edinburgh: Articulate Instruments Ltd.*

Yu, A., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and autistic traits. *PloS one*, *8*(9), e74746.

Zhou, X., Espy-Wilson, C., Boyce, S., Tiede, M., Holland, C., & Choe, A. (2008). A magnetic resonance imaging-based articulatory and acoustic study of 'retroflex' and 'bunched' American English /r/. *The Journal of the Acoustical Society of America*, *123*(6), 4466–4481.

Zhou, X., Espy-Wilson, C., Tiede, M., & Boyce, S. (2007). An articulatory and acoustic study of 'retroflex' and 'bunched' American English rhotic sounds based on MRI. *Eighth Annual Conference of the International Speech Communication Association*, 5–8.

Appendices



Appendix 1: Experiment 2 consent (recording)

CONSENT TO THE USE OF DATA

I understand that Robert Lennon is making recordings for a perception experiment, collecting my speech data which will be edited into audio stimuli, for use in an academic research project focusing on speech perception and language variation in the Glasgow area, as part of his PhD for the department of English Language, University of Glasgow, in collaboration with the Economic and Social Research Council. I also understand that short excerpts of my anonymised speech recordings may be used in teaching and/or conference presentations.

I give my consent to the use of data for this purpose on the understanding that:

- All names and other material likely to identify individuals will be anonymised.
- The data will be treated as confidential and kept in secure storage at all times.
- Participation in this experiment is voluntary, so I may opt out at any stage.
- The information is processed by the University in accordance with the provisions of the Data Protection Act 1998.

Signed by the contributor:

____ date: ______________

Researcher's name: **Robert Lennon**

Researcher's email: **r.lennon.1@research.gla.ac.uk**

Supervisor's names: **Prof Jane Stuart-Smith, Dr Rachel Smith**

Department address: **English Language**
12 University Gardens
Glasgow
G12 8QH

0141 330 6852 (Prof Stuart-Smith)
0141 330 5533 (Dr Smith)

It was the weekend, and John had no plans in the next couple of days. He was feeling tense since he'd had a stressful week in his job. His wife had gone away on business a couple of days ago, and he thought back to what she told him as she left, which was: "I wish I didn't have to go away this weekend, but I'll see you in a few days".

John recalled that his wife had shown him a nice seaside town a while ago, and it had quickly become a favourite of his. Straight away, he decided to pin all of his hope on enjoying his weekend at the town. He decided to drive to the town that day, and stay the night. He thought to himself that it wouldn't just be his second time visiting the place, it would be his third.

He knew the way so well, he had thrown the map in the bin. It was just a quick hop along the back roads to the town, and it didn't take him long to drive to the place.

When he was almost at the coast, the roads became narrower, and it became narrower still with every bend and every turn. Suddenly he had to slam on the brakes. Something was standing in the middle of the road, and it took him by complete surprise: it was a goat. He was glad he managed to stop, as the vehicle weighed a tonne. He realised just how lucky he'd got. He knew that anything that heavy would be difficult to stop, especially if it weighed about a tonne. The animal almost had a terrible fate.

He climbed out the vehicle and managed to guide the animal back into the field next to the road. Once it was in, he thought he should check that the gate was shut. He pushed it, and it closed with a thud. Looking across to the buildings in the field, he could see some animals. In a pen, next to the stables, was a pig. He thought to himself that it would be a good pet, as his friend Ben, who had one, said they can be really clean animals. He also saw a sheep in the next field. "They wouldn't be as good", he thought to himself. "The cost would be way too much to feed them if I bought a few of them, and to protect them from the wind I'd have to hollow out some kind of trench, like a pit".

Just beside the road was a hut. Inside it was a donkey. It stuck its head out to look at him and he stroked the mane on the back of its neck. "Usually they hate that", he thought to himself. So he stopped, as he didn't want to be mean to it, or cause it any hurt. Just then, a second donkey came round from behind the hut. "That must be its mate", he thought, and he left them alone.

He was about to continue his drive, but he was enjoying how brightly the sun shone on his face, so he decided to walk the small distance to the town. He thought he should keep in mind the highway code, so he didn't get hit by any traffic. He knew that walking on the right would be his best bet.

It was a pleasant walk, and at the roadside the daffodils were all in bud. He decided he would buy a bunch, when his wife arrived home in a few days.

He continued to walk down the road, and he looked at the trees as he walked past. He knew he was close to the seaside because a flock of seagulls flew by. Also, he noticed that in each tree, singing a melody, was at least one bird.

Because of fifteen minutes of walking in the heat, John could feel his skin beginning to burn. He knew that soon it would begin to hurt. Also, the sweat began to soak through his shirt. It was getting close to lunchtime, and he knew that in these hot conditions, without food and drink, he might take a funny turn. Each footstep felt like it was hitting the ground with a thud.

When he finally reached the town he was famished, so he found the closest takeaway and looked at the menu. He decided against chicken wings, as they didn't have much meat. Instead, he bought cod and chips, and some fruit punch. Then, he went to the bakery across the street and bought a hot cross bun. He also bought some of the cake in the cabinet. He didn't want the whole cake, so he just asked for a third. He wanted to go into Boots to get cream to stop his skin from continuing to burn, but at that time of the day it was shut. He quickly ate his fish and chips, then slowly bit into the cake, then polished off the hot cross bun.

When he arrived at his room his feet felt painful from walking all day. He took off his shoes, then he took off one sock at a time. He then took off his shirt. He lay back on the comfortable bed, which was really big. He looked out of the window and he got a fright when he saw a shadow just outside the window on the ledge, but it was only a bird. He was exhausted, and as he began to fall asleep the memories of the day began to fade, and he recalled the daffodils that were in bud.

He began to dream about the countryside. He was happy - he knew it would shape up to be a good weekend.

alarmed
bad
bait
bake
ban
bark
baste
bat
bead
beak
beard
beast
beat
beg
ben
bench
bet
big
bin
bird
bit
bizarre
blossom
boat
bowl
break
bud
bun
bunch
bunk
burst
butt
bump
bust
burn
cat
coast
cod
coat
con
code
cope
cone
cop
curt
cost
cud
cot
curse
crash
curd
cut
cuss
dot
dote
dog
dress

drink
dock
edge
face
fate
fear
fussed
fade
fall
feed
fifth
feet
first
fish
flower
food
fur
gate
gold
goat
green
grow
got
hate
hop
heat
hot
hope
hut
horse
house
howl
hurt
injure
kit
leaf
loch
make
mouse
mate
mane
mouth
mean
meat
mop
meek
mope
nope
note
nose
not
odd
pad
pan
park
pat
peg

pen
pet
pig
pier
pin
pit
punch
plant
punk
putt
port
pot
purr
rabbit
rat
road
roar
Ruchill
scared
same
sea
second
seem
shoes
shape
silver
shirt
sheep
smash
shut
shone
spun
sore
sock
snake
shown
sneak
Spain
spurn
soak
stem
strange
tonne
tyre
third
thump
those
thud
turn
weed
war
water
weird
wired

Appendix 4: Experiment 2 consent (experiment)

CONSENT TO THE USE OF DATA

I understand that Robert Lennon is undertaking a perception experiment, collecting data on my responses to audio stimuli, for use in an academic research project focusing on speech perception and language variation in the Glasgow area, as part of his PhD for the department of English Language, University of Glasgow, in collaboration with the Economic and Social Research Council. The research will be conducted as outlined in the accompanying information sheet. I also understand that my anonymised responses may be used in teaching and/or conference presentations.

I give my consent to the use of data for this purpose on the understanding that:

- All names and other material likely to identify individuals will be anonymised.
- The data will be treated as confidential and kept in secure storage at all times.
- Participation in this experiment is voluntary, so I may opt out at any stage.
- The information is processed by the University in accordance with the provisions of the Data Protection Act 1998.

Signed by the contributor:

____ date: ______________

Researcher's name: Robert Lennon

Researcher's email: r.lennon.1@research.gla.ac.uk

Supervisor's names: Prof Jane Stuart-Smith, Dr Rachel Smith

Department address: English Language
12 University Gardens
Glasgow
G12 8QH

0141 330 6852 (Prof Stuart-Smith)
0141 330 5533 (Dr Smith)

Appendix 5: Experiment 2 participant instructions

Thank you for agreeing to take part in this research. **Please read this data sheet before starting the experiment.**

The whole experiment will last no more than 25 minutes.

- There are 3 sections:
  1. A word identification task.
  2. A short passage to listen to, with a simple counting task.
  3. Another word identification task.
- At the start of sections **1** and **3**, there will be a short practice session: Please take this opportunity to adjust the volume to a comfortable level (top left of laptop keyboard).
- Please ask the researcher if there is anything you would like to have explained further.

Section 1:

- You will hear some recordings of words – your task is to choose the word you thought you heard.
- Please indicate this by choosing from the two options on the screen.
- Please use the labelled buttons on the keyboard to do this.
- **You will only have 2 seconds to make your choice, then the program will move to the next word.**
  (If you miss an item, don't worry – simply **continue with the next one**)
- After you have heard 25 words, you will be given a break, then when you are ready, press the spacebar to continue to the next session.
- You will hear 3 more sessions of 25 words, with breaks between them.

To start the first section, press the spacebar.

Section 2:

- Listen to the short story, played through your headphones. Don't worry about the sound quality – simply relax and listen!
- **Task:** As you listen to the story, listen out for **each time an animal is mentioned**.
- As you hear each one, tally them up (e.g.: 卌 || ...), then write the total beside it:

(counting space...)________________________________ Total: _____________

When you hear the beep at the end of the passage, please call the experimenter.

Section 3:

- This is the same kind of task as **Section 1** – please refer back to those instructions.

When you have finished, please call the experimenter.

Thank you!

University of Glasgow

- Where were you born?

____

- Please list the main places you have lived, and for how long:

____

____

- Where did your parents/guardians grow up?

____

- Which hand do you write with? LH / RH
- What is your age? ______________________
- Did you find it easy to understand the speaker in the story?
  Were there any sections/words you found difficult?

____

____

____

____

- How did you hear about this experiment?

____

Thank you!

Appendix 7: Experiment 3 consent (recording)

CONSENT TO THE USE OF DATA

I understand that Robert Lennon is making recordings for a perception experiment, collecting my speech data which will be edited into audio stimuli, for use in an academic research project focusing on speech perception and language variation in the Glasgow area, as part of his PhD for the department of English Language, University of Glasgow, in collaboration with the Economic and Social Research Council. I also understand that short excerpts of my anonymised speech recordings may be used in teaching and/or conference presentations.

I give my consent to the use of data for this purpose on the understanding that:

- All names and other material likely to identify individuals will be anonymised.
- The data will be treated as confidential and kept in secure storage at all times.
- Participation in this experiment is voluntary, so I may opt out at any stage.
- The information is processed by the University in accordance with the provisions of the Data Protection Act 1998.

Signed by the contributor:

____ date: ______________

Researcher's name:	**Robert Lennon**
Researcher's email:	**r.lennon.1@research.gla.ac.uk**
Supervisor's names:	**Prof Jane Stuart-Smith, Dr Rachel Smith**
Department address:	**English Language** **12 University Gardens** **Glasgow** **G12 8QH** **0141 330 6852 (Prof Stuart-Smith)** **0141 330 5533 (Dr Smith)**

CONSENT TO THE USE OF DATA

I understand that Robert Lennon is undertaking a perception experiment, collecting data on my responses to audio stimuli, for use in an academic research project focusing on speech perception and language variation in the Glasgow area, as part of his PhD for the department of English Language, University of Glasgow, in collaboration with the Economic and Social Research Council. The research will be conducted as outlined in the accompanying information sheet. I also understand that my anonymised responses may be used in teaching and/or conference presentations.

I give my consent to the use of data for this purpose on the understanding that:

- All names and other material likely to identify individuals will be anonymised.
- The data will be treated as confidential and kept in secure storage at all times.
- Participation in this experiment is voluntary, so I may opt out at any stage.
- The information is processed by the University in accordance with the provisions of the Data Protection Act 1998.

Signed by the contributor:

____ date: _____________

Researcher's name: **Robert Lennon**

Researcher's email: **r.lennon.1@research.gla.ac.uk**

Supervisor's names: **Prof Jane Stuart-Smith, Dr Rachel Smith**

Department address: **English Language**
12 University Gardens
Glasgow
G12 8QH

0141 330 6852 (Prof Stuart-Smith)
0141 330 5533 (Dr Smith)

University of Glasgow

Appendix 9: Experiment 3 participant instructions

Thank you for taking part in this research! **Please read this data sheet.**

- There are **3 tasks**, and the whole experiment will last around 30 minutes.
- You will hear recordings of words – **your task is to choose which word you heard**.
- You will use the computer mouse to make your choice, out of 2 words on the screen.

<u>Instructions (the same for each task):</u>

- There will be a short practise session before the start of each task.

- On the top-left and top-right of the screen will be 2 words.
  Take a second to familiarise yourself with the location of each word!

- After a second, a START button will appear at the bottom of the screen.

- When you click on the START button, one of the words will start to play over your headphones. At the same time, you should **immediately start to move the mouse**, and click on the word you heard.

- **Make your choice as quickly as you can! After <u>2 seconds</u> the program will move to the next word.** (Don't worry if you miss an item: the program will continue)

- After you have heard 25 words there will be a break: when you're ready, press Enter to continue to the next 25.

- Please ask the researcher if there is anything you would like to have explained further.

- **When you have finished, please call the researcher.**

Thank you!

- Where were you born?

____

- Please list all the places you have lived, and for how long:

____

____

- How long have you lived in Glasgow, and in which areas?

____

- Where did your parents/guardians grow up? (...if Glasgow, which areas?)

____

- How many speakers did you hear in the experiment?

____

- Any comments about the experiment in general? (e.g. how easy were the tasks, etc.)

____

____