The representation and processing of vowel duration as a perceptual cue for word-final voicing in native and non-native English

Zoe K Fletcher

St Cross College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2020

ACKNOWLEDGMENTS

First and foremost, I want to express my sincere gratitude to my supervisor, Professor Aditi Lahiri. I first became Aditi's student in 2013 when I began my MPhil in General Linguistics and Comparative Philology at Oxford. Throughout the past seven years, I have received a great deal of support and advice from her, culminating in this DPhil thesis. It has been a great privilege to be taught by Aditi, and I am grateful for all of her encouragement and reassurance throughout my time as a graduate student at Oxford.

I also extend my grateful thanks to my two examiners, Dr. Mary Baltazani and Professor Mathias Scharinger. I greatly enjoyed discussing my research with Mary and Mathias, and I thank them for taking the time to consider my thesis with such insight, and for making the examination process so enjoyable.

In addition, I am enormously thankful to the members of the Language and Brain Laboratory at Oxford, particularly Dr. Colin Brooks and Professor Henning Reetz. Colin has provided me with many hours of moral support, cheerful emails, and proof-reading for which I am so grateful. My fieldwork experience was also made considerably smoother by Henning's kind guidance throughout my time in Frankfurt. To other members of the lab, for the moral support, lovely lunches, and ample cups of tea, thank you. Stephen, Kim, Yoolim, Swetlana, Cong, Ana, Emily D., Emily L.S, Yaxuan, Holly, Beinan, Hilary, and Sandra- I feel very privileged to have worked with such talented group of researchers.

I am also very appreciative of the support granted to me by Goethe University, Frankfurt during my fieldwork. Thank you also to St. Cross College and the Faculty of Linguistics at Oxford for helping to fund this fieldwork. To those participants who took part in the experiments in Germany, and to those in Oxford, thank you. My sincere thanks also go to the speakers who were recorded

for the production studies in Experiment 1, and particularly those who were recorded again to make the stimuli for the subsequent experiments. Without this data, there would be no thesis.

Moreover, without the generous funding from both the Economic and Social Research Council and the Basant Kumar and Sarala Birla Graduate Studentship, undertaking my Doctorate would simply not have been possible. To both organisations, thank you for allowing me to pursue my love of learning and make it my 'job' for the past three and a half years.

I would also like to thank the Faculty of Linguistics at Oxford for giving me the opportunity to teach undergraduate Linguistics. It has been such an enriching and enjoyable experience; my students have been brilliant, bright, and it has been a joy watching them make their first foray into Linguistics.

Finally, and perhaps most notably, this thesis is the result of a great deal of love and support from my family and friends. My most heartfelt thanks go to my wonderful parents, Kay Fletcher and Laurence Fletcher, and grandparents, Joyce Vincent and Thomas Vincent, who have poured their lives into supporting my aspirations and have always believed in me. Enormous thanks also to Tracey Hall and Ian Hall, for their unconditional support, hours of childcare, and love. To Daniel Hall, the best man I know, for your endless and unfaltering encouragement, for keeping me going and loving me even in my most stressed out moments, thank you.

Most importantly of all, to my little boy, Harry. I thought that beginning my Doctorate when you were just ten months old would increase the challenge, but the opposite has been true. You have made everything easier. Thank you for allowing me to experience the world though your own wonder, for helping me to realise that I shouldn't always be working, for reminding me what is important in life, keeping me on track, and for being my biggest and proudest achievement. I only hope that watching me write this thesis has instilled in you the knowledge that you shouldn't ever have to choose between having a family and achieving your own aspirations. That said, of the many 'hats' I wear, being your mother will always be my favourite.

ABSTRACT

The relationship between vowel duration and word-final voicing in English is well attested. In a minimal pair, contrasting in word-final obstruents, the production of a longer vowel precedes the voiced obstruent (for example [bæd] *bad*) relative to the voiceless obstruent (for example [bæt] *bat*). The research presented here focuses on the underlying representation of vowel length, and the manner in which listeners process this representation to characterise word-final voicing. A total of five experiments were run examining a range of both recorded and synthesised monophthongs and diphthongs (/æ/, /eɪ/, /əʊ/, /aɪ/), thereby assessing the representation of both underlyingly single and bimoriac vowels.

In two of the five experiments presented, data obtained from native and non-native English speakers are compared. The aim here is to assess how the representation and processing of this fine-grained perceptual cue differs for L2 English speakers for whom this positional contrast does not exist in the surface form of their native phonology. Here, we consider German wherein the voicing contrast is neutralised word-finally, and only voiceless word-final obstruents are permitted in the surface form. Lexical status is also considered, alongside the extent to which lexical and acoustic information interact when there a mismatch between vowel duration and voicing. Most notably, results from a series of forced choice identification tasks and reaction time data obtained from a lexical decision task with fragment priming consistently support the notion that vowel duration acts as a primary perceptual cue for word-final voicing in native English speakers. For non-native speakers, an increased sensitivity to this perceptual cue is present in speakers who have had more exposure to spoken English, though this tendency remains governed by native phonology. Specifically, where an underlying voicing contrast does exist in German, listeners are primarily guided by their L1 phonology; perceiving a majority of voiceless responses. However, where an underlying voicing contrast does not exist in German, participants are guided by L2 phonology. Here, a lexical effect is found wherein German participants who have received less exposure to spoken English do not adhere to this tendency for nonwords. In the case of production, participants adhere to their L1 phonology for nonwords containing word-final sounds in which there is no underlying voicing contrast in German, regardless of the amount of exposure to spoken English received by the speakers.

TABLE OF CONTENTS

CHAPTER THREE

Vowel duration in the context of *voiced* and *voiceless* obstruents in English: production studies for native and non-native speakers

CHAPTER FOUR

The effect of vowel duration on the perception of English word-final voicing: identification tasks for native English speakers

CHAPTER SIX

An exploration into the relationship between vowel duration and the lexicon in English word-final voicing:

a lexical decision task for native speakers

CHAPTER SEVEN

Final discussion and conclusions

LIST OF ABBREVIATIONS

Abbreviation	Definition
CVCC	Consonant Vowel Consonant Consonant
CVNC	Consonant Vowel Nasal Consonant
CV	Consonant Vowel Fragment
CV?	Consonant Vowel Ambiguous Consonant
EMG	Electromyography
FUL	Featurally Underspecified Lexicon
L1	Level 1 (first language)
L2	Level 2 (second language)
V	Voiced
VL	Voiceless
VOT	Voice Onset Time

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

Introduction

1.1 Introduction

This thesis is primarily concerned with furthering our understanding of the ways in which listeners use variable acoustic signals to distinguish between word-final minimal pairs. Jones (1944) defines a minimal pair as two words or phrases that differ only in terms of one phonological element: their voicing characteristic. Voicing refers to the presence or absence of vibration at the vocal cords, in which *voiced* consonants are characterised by the presence of vibration and *voiceless* consonants are characterised by the absence of vibration. Both acoustic and lexical information is made available to listeners during speech. The acoustic information is dependent on both the articulation of a phoneme, as well as its surrounding environment. The mental lexicon, the brain's word-store, also supplies information which may influence categorisation. For example, phonemes which prime words of higher frequency are more likely to be perceived than those which activate lower frequency words; here, context also plays a part. The central research questions in this thesis investigate the variable contexts in which word-final phonemes are produced and perceived. More specifically, this thesis addresses the effect of vowel duration and of lexical status on the categorisation of word-final voicing in English.

The five experiments conducted within this thesis therefore seek to address the extent to which acoustic cues from preceding segments can provide information that aids the listener in identifying a word-final segment, along with furthering our understanding of the role of the lexicon and its relationship with this process. In addition, the interaction between Level 1 (hereafter, L1) and Level 2 (hereafter, L2) phonology will be investigated. Here, *L1* refers to a

speaker's native language, and *L2* refers to any languages which the speaker has subsequently acquired. As such, if a speaker is operating in a second language, are their native perceptual cues likely to inhibit their sensitivity to different cues that exist in the phonology of their second language? Here, this thesis considers German, in which the word-final voicing contrast is neutralised.

Drawing on support from existing literature, this chapter aims to introduce each of these areas of interest, before presenting the key research questions that this thesis seeks to answer. A more in-depth review of specific papers that have influenced this thesis will be presented in Chapter Two in the form of a detailed literature review.

1.2 Defining and modelling speech perception

The study of speech perception is concerned with the ways in which a listener maps variable acoustic signals onto perceptual objects. No single speaker can ever articulate the same word more than once in an identical fashion. Consequently, listeners must be able to use all the phonetic information available to them to gauge cues that will allow them to categorise phonemes and process speech. In their paper *Perception of the speech code*, Liberman et al. (1967) write:

> 'Man could not perceive speech well if each phoneme were cued by a unit sound... [instead] many phonemes are encoded so that a single acoustic cue carries information in parallel about successive phonemic segments.'
>
> (Liberman et al., 1967:431)

Mann (1980) illustrates this concept in a study that investigated the [da] to [ga] continuum and found that, when asked to categorise sounds in the middle of the continuum, listeners

perceived [ga] when preceded by [ɑl], and [da] when preceded by [ɑr]. Mann (1980) posits that this difference in perception may have been due to coarticulation effects between the final consonant of [ɑr]/[ɑl] and the corresponding initial consonant of [da]/[ga]. However, there also may have been an auditory contrast effect resulting from the formant frequencies. The relative frequency of F3 varies between [da] and [ga]. It is higher in the former, and lower in the latter. As such, the frequency of F3 in [ɑr]/[ɑl] may have affected the perception of F3 in [da]/[ga]. In [ɑr] F3 is low, therefore F3 in [da]/[ga] may sound contrastively higher, thereby cuing a [da] response. Conversely, F3 in [ɑl] is high which may lead to F3 in [da]/[ga] seeming lower and therefore cuing a [ga] response. This exemplifies the influence that successive segments may have on the overall perception and categorisation of phonemes.

There are two primary models often discussed regarding the nature of perception and the way in which listeners categorise speech sounds. They are the continuous model and categorical model. Continuous perception is the concept that phonemes are identified along a continuum. Conversely, categorical perception is the notion that phonemes are categorised into distinct groups. As such, discrimination of sounds within the same group proves more difficult than the discrimination between sounds from different groups.

> 'Research with some of the encoded phonemes has shown that they are categorical, not only in the abstract linguistic sense, but as immediately given in perception.'
>
> Liberman et al. (1967:442)

Experimental evidence for categorical perception can be found in studies where durational cues are synthetically varied to form steps along a continuum that differ in small, acoustically equal gates. For example, results from an experiment conducted by Liberman (1996:213) involving three stops; /b/, /d/, and /g/ suggest that listeners do not perceive step-like differences

between the stops, '*but essentially quantal jumps from one perceptual category to the other'*. Multiple experiments have investigated the linguistic cues that may contribute to the identification of phonemes by synthetically altering a wide range of variables, such as the Voice Onset Time (hereafter, VOT) and the transition lengths of different speech sounds. Here, *VOT* refers to the period of time between the release of a stop consonant and the subsequent onset of voicing, whereas the *transition length* refers to the time required to move from the production of one sound to the next.

Eimas (1963) compares the perception of consonants and vowels, stating that consonants are perceived categorically whilst vowels are perceived continuously. Eimas (1963) conducted an ABX identification and discrimination task using three series of duration stimuli, one series of reflectance stimuli, and then a series of speech sounds; the stop consonants /b/, /d/, and /g/. Participants were requested to listen to the auditory stimuli, and decide whether the duration stimuli were long or short by writing *L* or *S*. For reflectance stimuli, participants indicated whether the stimuli were light, medium, or dark by writing *L*, *M*, or *D,* and for speech stimuli, participants were asked to identify the speech sounds orthographically as *B*, *D*, or *G*. The results suggest that:

> 'Identification of consonants was marked by abrupt transitions between phonemic categories and small context effects…Thus, perception of the consonants was essentially categorical. Vowel perception, however, was continuous: the identification functions showed steep boundaries.'
>
> (Eimas, 1963:206)

So, does perception differ when humans are distinguishing between speech and non-speech sounds, and if so how is this difference likely to be modelled? Research has suggested that the perception of speech and non-speech sounds happens in different hemispheres of the brain.

Speech sounds are primarily processed in the left cerebral hemisphere, and non-speech sounds such as melodies and sonar signals are primary processed in the right cerebral hemisphere (Broadbent and Gregory, 1964; Bryden, 1963; Chaney and Webster, 1965; Kimura, 1961, 1964, 1967). Remez et al. (1981) conducted an experiment using sine wave analogues as opposed to *normal* sounding speech, as characterised by Johnson (2010):

> 'The fact that we have a more categorical response to speech signals than to sine wave analogs of speech suggests that there is something special about hearing formant frequencies as speech versus hearing them as nonspeech video game noises.'
>
> Johnson (2010:125-126)

A possible explanation for this is that humans can indeed extract intended, categorical gestures from the phonetic information given to them through the speech stream. However, this information can only be recovered from the speech stream in a learned manner, and so categorising non-speech sounds is not possible. Another explanation could be that humans perceive speech in ways that have been refined through linguistic experience. Johnson (2010) writes that:

> 'As speakers, we have somewhat categorical intensions when we speak- for instance, to say "dot" instead of "got".'
>
> Johnson (2010:126)

So, it follows that listeners may perceive speech according to the categories that they have had prior linguistic experience of. This theory is supported by the findings from this thesis as non-native speakers are guided more by their native phonology where a category exists in their native language.

Several proposals have been devised attempting to model speech perception. Werker and Logan (1985) characterise the nature of speech perception as:

> 'A continuous controversy… [of] the question of whether speech perception can be best explained by positing a specialised linguistic processor (Liberman [et al.], 1967), a generalised psychoacoustic processor (Pastore et al., 1977), or a dual-factor processor (Fujisaki and Kawashima, 1969).'
>
> Werker and Logan (1985:35)

Categorical perception provides evidence for a specialised linguistic processing mechanism given that the processing of speech sounds is characterised by much of the literature as being categorical. However, this perception is not limited to humans. Results from Kuhl and Miller (1975a) found that chinchillas both demonstrate labelling functions and position phonetic boundaries for the sounds /d/ and /t/ in a similar manner to English adults. Additionally, Kuhl and Padden (1983:1009) conducted research which demonstrates that animals '*can partition [a] continuum into categories consistent with the phonetic identity of sounds'*. These studies support the theory of a generalised acoustic processor. Equally, studies that illustrate context effects when presented with non-speech stimuli support the notion of a single-factor psychoacoustic mechanism (Miller and Liberman, 1979). These effects have found in both adult (Pisoni et al., 1983) and infant (Jusczyk et al., 1983) listeners.

Finally, a listener's ability to distinguish between stimuli and categorise phonemes one way under certain test conditions, and then perceive finer discriminations under alternative test conditions, supports a dual-factor model (Fujisaki and Kawashima, 1969, 1970; Pisoni, 1973). Relating to dual-factor models, evidence for a *phonetic module* came about in the 1960s (Liberman et al., 1967; Liberman and Mattingly, 1985). Fodor (1983) defines a *module* as a neural system that has evolved to perform a specific task, reliant on processing that is special

to its own domain. Duplex perception provides evidence for the existence of a phonetic module. Duplex perception occurs when a listener perceives an identical acoustic fragment in two different ways at the same time. Galantucci et al. (2006) report on the methodology of one study investigating duplex perception (Mann & Liberman, 1983; Whalen & Liberman, 1987). They write:

> 'Most of a three-formant syllable (the base) is presented to the listener's left ear. The remaining part of the syllable- either of two third-formant transitions that, when integrated with the base, sound like /da/ or /ga/- is presented to the right ear.'
>
> Galantucci et al. (2006:5)

Listeners reported hearing an unambiguous [da] or [ga] in the left ear and a non-speech *chirp* in the right ear. This chirp is the transition. Categorical perception of speech sounds is elicited when a continuum of transitions between /da/ and /ga/ are presented in the opposite ear to the base. Here, the *chirp* discrimination is not categorical. One explanation of these findings is that there are two distinct perceptual systems at play, one which is the phonetic module and one which is the auditory system. As such, dual-factor models suggest that acoustic information is stored separately in both an auditory and phonetic code. Given that *'the auditory code decays rapidly relative to the acoustic code (Fujisaki and Kawashima, 1969, 1970; Pisoni, 1973)) [and]…the auditory code decays more rapidly than the phonetic code'* (Werker and Logan, 1985:35), dual-factor models predict that one can determine an acoustic level of processing only when stimuli can be compared to one another after a short time-lag. As such, experiments which incorporate longer intervals between stimuli presentation or have higher memory demands are unlikely to facilitate retained access to the acoustic code (Pisoni, 1973; Crowder, 1982).

An additional model was proposed by Werker and Logan (1985). Here, native English and native Hindi speakers were incorporated in three same-different (AX) discrimination tasks. The findings were three-fold. Firstly, the participants were able to categorise syllables into the phonemic categories with which they were familiar. Secondly, they showed evidence of the perception of category boundaries between phonemes which were not native to their own language. Finally, participants were able to *'discriminate syllables on the basis of any acoustic variability between individual exemplars'* (Werker and Logan, 1985:43). Given this cross-linguistic speech perception, this paper posits the theory that:

> 'A three-factor model, including auditory, phonetic, and phonemic processing, may be necessary to accommodate existing findings…the combined results from [the] experiments provide support for the existence of three distinct speech-perception factors.'
>
> Werker and Logan (1985:35)

These findings are particularly relevant for this thesis, as one of the primary areas of research that it seeks to address centres on the interaction between native and non-native phonology, and the effect that this may have on the categorisation of word-final phonemes.

Finally, Repp (1982) reviewed models which investigated the notion that a human's ability to distinguish between and subsequently categorise phonemes is reliant on multiple acoustic cues. This paper *'reviews a variety of recent experimental findings'* and highlights the importance of the surrounding phonetic context on this perception (Repp, 1982:81). Regarding cues for voicing, Repp (1982:90) highlights several perpetual cues that listeners may use. Most significantly for the research presented in this thesis, vowel duration was found to be important for the perception of voicing for stop consonants in a word-final position. In addition, *'the properties of the release burst, and the duration of the preceding closure'* were significant.

The factors influencing the perception of speech sounds are therefore of long-standing interest to researchers, and a broad understanding of the nature and models of speech perception proves crucial for forming a theoretical basis for the research conducted in this thesis.

1.3 The role of segmental duration in language

1.3.1 Phonetics, phonology, and segmental duration across languages

Segmental duration is one piece of phonetic information that speakers vary. However, in terms of contrast, usually only two levels of durational variability are used meaningfully in language; long versus short. Vowels may be considered to be long or short across the languages of the world, and durational differences in consonants are termed as either singleton (short) or geminate (long). Phonetic attributes are in turn constrained by the underlying phonological rules of the languages in which they are found. The role of duration, long versus short, is therefore apparent within both the phonetics and phonology of any given language.

Though differences in vowel duration are more common, contrasts in consonant duration are also distinctive in some languages such as Arabic, Danish, Finnish, and Italian. In Finnish, for example, duration plays a crucial role in contrasting vowels as well as consonants: /taka/ ‘taka' *back*, / tak:a/ ‘takka’ *fireplace*, and /ta:k:a/ ‘taakka’ *burden*. Conversely, languages such as English do not have distinctively long and short consonants. Consonant gemination may act as a distinctive feature within the phonology of those languages in which they occur. For example in Italian, Northern and Southern dialects can be distinguished from one another by the presence or absence of gemination. In Southern Italian dialects, /fato/ ‘fato’ *fate* and /fat:o/ ‘fatto’ *fact* are distinguishable through gemination, whereas in Northern Italian dialects the two become homophonous due to degemination neutralising this distinction (Chang, 2000). Duration contrasts lead to contrasts in syllable weight, which in turn also plays a role in

phonological rules. In Old English, for instance, all high vowels are deleted after heavy syllables which include syllables with geminates as well as long vowels.

Interactions between consonant and vowel durations across the world's languages are common. In Italian, shorter vowels appear before geminates as opposed to singleton consonants (Esposito and Di Benedetto, 1999; Pickett, Blumstein, and Burton, 1999). For example, 'bevve' *he/she drank*, is /ˈbev:e/, while 'beve' *she/she drinks/is drinking* is /ˈbe:ve/. Similarly, in Hindi (Ohala, 1983; Shrotriya et al., 1995), /pat:a/ *leaf* is pronounced with a notably shorter initial vowel than /pata/ *address*. Conversely, languages such as Finnish (Lehtonen, 1970) and Persian (Hansen, 2004) demonstrate the lengthening of vowels before geminates. Then, there are languages that do not demonstrate a notable difference in the duration of the preceding vowel with regard to geminates or singletons. These languages include Egyptian Arabic (Norlin, 1987), Estonian (Engstrand and Krull, 1994), and Hungarian (Ham, 2002).

A crucial lengthening and shortening phenomenon is found in the context of coda consonants which differ in voicing; in particular, vowels before voiceless consonants are shorter than vowels before voiced consonants. Vowel duration as a perceptual cue for voicing is considered by some to be a language-universal (Chen, 1970). Chen (1970) conducted a production study in four languages; English, French, Korean, and Russian. He found that across all four languages, *'a vowel is invariably longer before a voiced consonant than before an unvoiced one'* (Chen, 1970:135). This phonological tendency forms the crux of this thesis. However, the extent to which this proves true appears to vary, with English showing the greatest contrast in vowel duration, followed by French, Russian, and Korean. Given that only four languages were investigated, three of which were Indo-European, Chen (1970) acknowledges that he cannot make any sweeping statements regarding this phonological tendency. However, he writes:

> ‘Such invariable variations of vowel duration depending on the voicing of the following consonant can hardly be regarded as accidental. In fact, if we look beyond our own data and examine some of the evidences furnished by published sources, we would see that similar phenomenon is observed in a number of other languages as well.’
>
> Chen (1970:135).

Beguš (2017) supports Chen’s (1970) assertion that this phonological tendency occurs across language families, more specifically:

> ‘English, French, Russian, Korean (Peterson and Lehiste, 1960; House, 1961; Chen, 1970; Luce and Charles-Luce, 1985; Abdelli-Beruh, 2004; de Jong, 2004), German, Swedish, Icelandic (Port, 1996), Hindi (Maddieson and Gandour, 1976; Lampp and Reklis, 2004; Durvasula and Luo, 2014), Arabic (Port et al., 1980; Jong and Zawaydeh, 2002), Bengali, Hungarian, Italian, Norwegian, Spanish, Danish, Persian, and Dutch (reported with references in Maddieson and Gandour, 1976; Kluender et al., 1988), among others.’
>
> Beguš (2017:2168)

Much of the research on this topic centres around vowels in closed syllables, however evidence suggests that this interaction between vowel duration and voicing also occurs in open syllables:

> ‘For example, in English, Korean, French, Arabic, Spanish, and Norwegian (Port, 1981; Chen, 1970; Abdelli-Beruh, 2004; Port et al., 1980; Fintoft, 1961; Zimmerman and Sapon, 1958). In the absence of evidence to the contrary, it is assumed that the same general mechanism is responsible for vocalic durational differences in both open and closed syllables. The magnitude of the voicing effect, however, can differ according to whether the affecting consonant is in coda position (closed syllable) or in the onset position of the following syllable (open syllable).’
>
> Beguš (2017:2168).

Jongman et al. (1992) investigated the interaction between vowels and voicing in Dutch and aimed to decipher the extent to which the phonological representation of the [VOICE] feature is able to influence the identification of individual words. Their paper explains that within Dutch there are contrasts in both obstruent voicing and vowel duration, but that these contrasts in voicing are neutralised when they appear in a syllable final position. However, Jongman et al. (1992) suggest that within Dutch *'both long and short vowels are lengthened by some 25ms when followed by a medial voiced obstruent'* (Jongman et al., 1992:137). The investigation aimed to look at vowel category boundaries and decipher *'whether this vowel length cue influenced listeners when hearing stimuli with ambiguous vowel duration in an identical, neutralised consonantal context in which the underlying representation of the obstruent following the vowel differed in voicing'* (Jongman et al., 1992:137). Investigating purely real word contexts, this experiment involved creating a series of ambiguous vowel lengths along the [at]~[a:t] continuum which were then placed before an initial consonant in order to form the following pairs of words /stad/ [stat]~/sta:t/ meaning *city* and *state*, and /zat/~/za:d/ [za:t] meaning *drunk* and *seed*. These were words for which the underlying word-final consonants are either voiced or voiceless, yet the surface representation is always realised as voiceless. The long vowel [a:] was digitally shortened by removing twelve glottal pulses. This reduced the vowel from 197ms to 83ms, after which the closure and release phase of the final consonant in the word /stad/ [stat] was added so that stimuli ended in a voiceless stop [t]. The initial consonants [z] (144ms) and [st] (244ms) were then appended onto the twelve-step continuum such that two word pair continua were formed. Fifteen participants in total were played the stimuli in a soundproof booth and were told that *'the continuum endpoints were existing Dutch words and that they were required to identify the vowel of each stimulus word as short or long by pressing one of two clearly marked buttons [labelled 'a' or 'aa'] on a response box placed*

in front of them' (Jongman et al., 1992:143). A categorical shift was found within the results such that short vowel responses shifted to long vowel responses mid-way through the twelve-step continuum. The results therefore found that listeners are not being guided by the surface form alone and that the perception of vowel length is affected by the underlying phonological feature of [VOICE]. The researchers therefore conclude that:

> 'There is indeed an underlying abstract phonological representation differing from the surface pronunciation which influences the recognition and identification of words. The listener's perception is guided not just by available acoustic information but also by underlying phonological representations of words.'
>
> Jongman et al. (1992: 149).

The findings of this paper therefore proved significant in giving cross-linguistic evidence for the importance of the relationship between vowel duration and voicing. As with Dutch, German phonology contains a word-final devoicing rule in which the voicing contrast is neutralised word-finally. The results from Jongman et al. (1992) therefore heavily influenced the research within this thesis. Particularly whether the underlying nature of voicing in German is likely to interact with the surface form of English word-final consonants when a native German speaker is operating in L2 English.

Thus, vowel and consonant duration are crucial not only in establishing underlying contrasts, but additionally, lengthening and shortening processes occur across many contexts which need not necessarily be contrastive. Although underlying phonological contrasts may guide perception, this thesis focuses on the extent to which non-phonemic contextual lengthening plays a role in the processing and identification of sounds. We therefore turn to English.

1.3.2 Allophonic vowel lengthening

In English, word-final voicing is acoustically very clear. Research refers to the relationship between vowel duration and voicing outlined above, and supports the notion that a voiced consonant is preceded by a notably longer vowel than a corresponding voiceless consonant when it occurs word-finally. Thus, minimal pairs such as [meɪd] *made* and [meɪt] *mate* differ not only in word-final voicing but also phonetically in terms of the duration of the vowel. This is known as pre-fortis clipping. Klatt (1976) argues that, in English, this preceding vowel duration acts a primary phonetic cue for word-final voicing. He writes that:

> 'If a change in duration is predictably larger than the just-noticeable differences for segmental duration, then this change has the potential of carrying useful information from the speaker to the listener.'
>
> Klatt (1976:1208).

Listeners use these fine-grained phonetic cues from neighbouring segments to influence their categorisation of word-final phonemes.

Beguš (2017) found that this phonological tendency also occurs in Georgian when the word-final consonant is an ejective stop. He conducted a production study which measured vowel duration before stops in relation to three laryngeal features, in addition to considering the effect that closure duration and VOT have on the length of the vowel. This experiment consisted of twelve native speakers. All speakers were living in the US and reported speaking at least two other languages proficiently. The participants all read out an identical list of six hundred and seventy-five sentences which contained nonce words. This list was read at a controlled pace, and speakers were asked to repeat the sentence if they felt that they had made a mistake. Beguš (2017) reports that:

> ‘The results show that vowels have significantly different durations before all three series of stops, voiced, ejective, and voiceless aspirated, even when closure and VOT durations are controlled for. The results also suggest that closure and VOT durations are inversely correlated with preceding vowel duration’.
>
> Beguš (2017:2168)

However, he acknowledges that ‘*the causes of this phonetic generalisation are…poorly understood’* (Beguš, 2017:2168). He writes:

> ‘General theories accounting for this suggest that vowel duration differences can be either automatic mechanical results of articulatory origin or part of the active phonology, i.e. actively controlled by speakers.’
>
> Beguš (2017:2169)

He references five theories that offer possible explanations for what he calls this ‘voicing effect’:

I. *Articulatory Energy Expenditure*: a greater ‘physiological’ force is necessary to articulate voiceless stops, which shortens the preceding vowel (Belasco, 1953).

II. *Compensatory Temporal Adjustment*: if the timing of the VC sequence is constant, vowels preceding voiced stops should be longer as these vowels have a shorter closure and VOT. He writes that this should also work *‘vice versa via “compensatory temporal reorganization of sequential motor commands” (Chen, 1970; Kozhevnikov and Chistovich, 1967; Lindblom, 1967; Port, 1981; de Jong, 1991; Port et al., 1980, 1987)’*.

III. *Rate of Closure Transition*: instead of a greater ‘physiological force’, this theory characterises voiceless stops as needing a greater velocity during closure due to being articulated with a higher intraoral pressure. This results in a shorter vowel duration preceding this consonant (Chen, 1970; Öhman, 1967).

IV. *Laryngeal Adjustment*: significant adjustments are required in vocal fold positioning to maintain voicing during closure. These adjustments take time to position, and this results in a longer vowel duration when the following consonant is voiced (Halle et al., 1967; Chomsky and Halle, 1968).

V. *Perceptual Distance*: this theory posits that closure duration provides a primary perceptual cue for voicing. It suggests that speakers therefore exploit the duration of the vowel, making it shorter prior to a voiceless consonant, to maximise the closure duration. As such, '*short closure is perceived as even shorter after a longer vowel (Denes, 1955; Lisker, 1957; Javkin, 1975; Kluender et al., 1988)*'.

Beguš (2017:2169)

The results of Beguš's (2017) study on ejective stops therefore support proposals which suggest that *'laryngeal gestures, temporal compensation and closure velocity affect vowel duration, [whereas] some explanations, especially perceptual and airflow expenditure explanations, are considerably weakened by the results'* (Beguš, 2017:2168).

The reason for so many competing proposals is likely because so little is understood about the relationship between the vowel duration preceding consonants and other laryngeal features. The research in this thesis aims to build upon current literature which attributes the perception of word-final voicing to preceding vowel duration in English. In doing so, this thesis does not deny that the nature of speech perception is complex, and that there are many possible cues aside from vowel duration that may influence phoneme categorisation word-finally. Instead, the findings of this thesis are focused on a more precise understanding of the extent to which the underlying representation of the vowel is necessary for the characterisation of word-final voicing. Indeed, the results of the five experiments presented in this thesis support the notion that vowel duration does act as both a primary and necessary cue for word-final voicing distinctions in English.

1.4 Building a model of speech perception

The results presented in this thesis support a model of perception based on the Featurally Underspecified Lexicon (hereafter, FUL) model (Lahiri and Reetz, 2002). As previously discussed, speech is variable and no word can ever be articulated in an identical fashion more than once, even by the same speaker. As such, the FUL model posits that individual phonemes possess *features* which are unique to them, and that speakers and listeners use these features to interpret the speech stream.

Specifically, this thesis is concerned with the [VOICE] feature when it appears word-finally. This feature simply refers to the nature of voicing for any given phoneme. As such, rather than supposing that word-final voicing is represented in the minds of speakers in a variable form, for instance; *somewhat voiced, no voice, consistent voicing,* this thesis considers it a more functional suggestion that the presence of a monovalent [VOICE] feature is likely. The variable nature of vowel duration, which has no associated feature according to the FUL model, is therefore judged by the listener to be either *short* or *long* in relation to the nature of the following [VOICE] feature of the word-final sound.

The FUL model therefore motivated the experimental design of this thesis in that the five experiments were designed to directly test the salience of the word-final [VOICE] feature, even when the ending of a stimulus was omitted or made ambiguous. In order to achieve this, participants were tested to determine the extent to which they were able to extract the acoustic information made available to them through the preceding vowel duration, and use it to guide the assignment of the binary [VOICE] feature word-finally. This was tested using both recorded speech and computer-generated speech, and the findings of the thesis suggest a model of perception that works as follows. Let us first consider the speaker. The speaker conceptualises an idealised representation of their desired utterance. In turn, they are able to anticipate the

nature of the word-final [VOICE] feature, and as such they lengthen the preceding vowel in their speech production according to whether they intend it to be short or long; short in the case of a voiceless word-final sound, and long in the case of a voiced word-final sound. Here, the production data from Experiment 1 demonstrated that native English speakers produced significantly different vowel durations preceding word-final voiced and voiceless sounds in line with this tendency.

For the listener, the acoustic information first provided to them in anticipation of word-final voicing is the vowel length. Recall that, according to the FUL model, the duration of a vowel is not specified for a feature, and so listeners must interpret the vowel segment of the utterance as underlyingly long or short according to the acoustic information made available to them. From this acoustic signal, the listener can extract the relevant information needed to assign the [VOICE] feature to the word-final obstruent. If the listener construes the vowel duration as long they interpret the word-final obstruent as being specified for the [VOICE] feature, and if short, this feature is not assigned. For Experiments 2-5, the specific voicing characteristics of the word-final obstruents were omitted or made ambiguous. Despite this, listeners were able to consistently use the vowel duration to guide their perception as expected; perceiving voiced consonants word-finally after a long vowel and voiceless consonants word-finally after a short vowel. These results therefore consistently support the notion that English is sensitive to fine-grained phonetic differences in vowel duration, and this plays an important role in the perception and salience of the underlying word-final [VOICE] feature. Listeners were able to extract the relevant information from the vocalic acoustic signal in order to categorise the word-final obstruents in this way.

A further discussion of the FUL model will be presented in Chapter Two.

1.5 The role of the lexicon in speech perception

Of course, phonology and phonetics are not the only factors at play when it comes to interpreting the speech stream. In addition to the role of acoustic information in speech perception, this thesis is additionally concerned with the influence of the mental lexicon. Specifically, this research investigates the manner in which the lexicon may interact with both the phonetic information being provided to listeners in the surface form, and the underlying phonological information of which the listener has knowledge. Much of the literature detailing previous studies in this area will be discussed in Chapter Two. However, here we will introduce the beginnings of the theory surrounding the role of the lexicon in speech perception. A focus on the quality of stimuli will also be discussed regarding this area of research (Burton et al.,1989; McQueen, 1991).

Crucially, the role that the lexicon plays in speech perception becomes all the more relevant when phonemes sound ambiguous. For instance, if a speaker is having difficulty perceiving the difference between a word-initial [b] or [p], they are more likely to disambiguate the phoneme as a [b] if it is followed by [aɪt] *bite* and [p] if it is followed by [ɪk] *pick*. Here, we see the influence of lexical status on the process of phoneme categorisation. The interactive model, one interpretation of the mechanism of speech perception, posits that lexical information affects pre-lexical representation. McClelland et al. (2006) characterise this model as predicting that:

> 'Lexical information actually reaches down and reshapes the mental representation of the sound that is heard. This view contrasts with other proposals, in which perceptual processing is [viewed] a strictly autonomous, bottom-up process, with the influence of lexical and semantic contextual information arising only at a later decision stage.'
>
> McClelland et al. (2006:363)

The lexicon therefore has the capacity to cue perceptual bias. Bias caused by the lexicon becomes relevant in situations in which there is a minimal pair of phonemes where one phoneme would construct either a real word or a word of higher frequency, and the opposing phoneme would construct a lower frequency word or a nonword. In both cases, a listener may be biased to perceive a phoneme based on lexical information that, were they using purely acoustic information, would be categorised differently. For example, if a listener is having to decide between hearing a [d] or [t] word-finally in a CV sequence such as [kae?], the interaction of the lexicon with the acoustic information provided by the vowel duration may lead the listener to hear [t] word-finally to create the higher frequency word *cat*, as opposed to the lower frequency word *cad*. This may be the case even if the vowel is perceived as underlyingly long, which would be expected to cue a [d] response. Naturally, other factors are also taken into consideration in every-day speech, for example the context and subject of a conversation. However, when tested under experimental conditions, deciphering more information about the effect of the lexicon and its interaction with acoustic information can shed a great deal of light onto the ways in which humans perceive speech sounds.

Burton et al. (1989) explore the relationship between phonetic categorisation and the lexicon by focusing on acoustic parameters. Much of the research into speech perception focuses on the interaction between higher and lower levels of processing. This paper acknowledges that experimental evidence from across a range of studies have suggested that the perception and categorisation of phonemes can be influenced by the lexicon (Ganong, 1980; Grosjean, 1980; Marslen-Wilson and Welsh, 1978; Rubin, Turvey, and Van Gelder, 1976; Warren, 1970). These studies all reference theories which suggest a top-down processing model, in which higher levels of speech processing such as sentence and word-level effects affect lower levels. Conversely, competing research has suggested that a bottom-up approach is more feasible, with

acoustic information forming the basis for higher levels of processing to act upon (Foss and Gernsbacher, 1983; Fox, 1984). Burton et al.'s (1989) paper therefore aims to '*consider the extent to which top-down processes affect speech perception, specifically by exploring the influence of the lexical status of a stimuli on phoneme identification'* (Burton et al., 1989:567). This paper highlights the importance of the quality of the methodology and stimuli chosen. Two experiments were conducted, the first was based on two test stimuli that contained synthesised versions of natural speech. These continua varied in VOT, with the first ranging from the word /du:k/ *duke* to the nonword */tu:k/ **tuke*, and the second ranging from the nonword */du:t/ **doot* to the word /tu:t/ *toot*. Native speakers of American English partook in this experiment. Subjects were played the stimuli and asked to decide whether they had heard a [d] or [t] word-finally. The results of this experiment '*replicated earlier studies that showed a lexical effect in the phonetic categorisation of speech continua. The overall functions demonstrated a small but significant shift in the locus of the phonetic boundary as a function of the lexical status of the stimulus'* (Burton et al., 1989:571). This research complemented the findings of Fox (1984) and Miller and Dexter (1987) in that a '*significant lexical effect emerged in the slowest reaction time range*' (Burton et al., 1989:571). The second experiment utilised stimuli that resembled natural speech more strongly. The stimuli from the first experiment were reused, except that the '*amplitude of the burst and aspiration was systematically varied across the continuum'* (Burton et al., 1989:571). The lexical effect resulting from the first experiment disappeared in the second. Stimuli closer to the approximation of natural speech led to both '*overall identification functions and the reaction time range identification functions [failing] to demonstrate a shift in the locus of the phonetic boundary as a function of the lexical status of the stimulus*' (Burton et al. 1989:572). The results of this paper therefore suggest that the acoustic structure of stimuli is of paramount importance in studying the relationship between

phonetic categorisation and the lexicon. Most significantly, the role of top-down processing in categorisation may be overemphasised in contexts where the acoustic structure of experimental stimuli has not been considered. It would appear from these results that more natural sounding stimuli may result in the reduction of lexical effects.

The findings of this thesis report that lexical effects are found in the case of both natural and synthesised stimuli, though this is dependent on word frequency, and vowel duration remains the primary perceptual cue. Specifically, the results of this thesis do not show vastly different lexical effects dependent on whether the stimuli consisted of natural or computer-generated speech, and a clear shift in the phoneme boundary was not clearly or consistently supported across the experiments regardless of the structure of the stimuli. However, incorporating both methods of stimuli formation into the experiments encompassed in this thesis acted as a control for this factor. Future research may consider running an identical experiment with two groups of participants, one responding to human recorded speech and one responding to computer-generated, synthesised, speech to directly test the influence of this on speech perception.

The findings from this thesis also consider the importance of stimuli quality. McQueen (1991) addresses this issue. Two experiments were conducted to test the quality of stimuli in which participants were required to categorise syllable-final fricatives. The first study showed no evidence of lexical shifts between the categorisation of word-final fricatives in word/nonword stimuli such as /fɪʃ/~*/fis/ *fish*~**fiss* and /kɪʃ/~/kis/ **kish*~*kiss*. However, the stimuli in the second experiment underwent a process of stimuli degrading by low-pass filtering. In this instance, lexical shifts did emerge. This paper concludes that:

> ‘Stimuli quality rather than stimulus ambiguity per se determines the extent of lexical involvement in phonetic categorisation. Furthermore, the lexical shifts were limited to fast [reaction time] ranges, contrary to the interactive model’s predictions. These data therefore favour an autonomous bottom-up model of speech recognition.’
>
> McQueen (1991:433)

This research therefore complements the aforementioned study conducted by Burton et al. (1989), in that the effects of top-down processing may be overemphasised, and the quality of the stimuli presented to participants is an important consideration when analysing the results and inferring generalisations. Again, the current thesis aimed to mitigate issues with stimuli quality by employing a range of different methodologies across the four perception experiments. Recall that this thesis has incorporated both recorded and synthesised speech, alongside including both the presence of an ambiguous word-final sound along with the absence of a word-final sound. The consistency across the findings, regardless of the different methodologies, support the notion that vowel duration is a primary cue for word-final voicing and that clear lexical shifts are not consistently evidenced regardless of the quality and formation of the stimuli.

Nonetheless it would appear from this research that the lexicon does have the capacity to affect speech perception, and this forms an important aspect of the research encompassed within this thesis. In particular, how the lexicon interacts with acoustic information and results in the perception and categorisation of English word-final phonemes on the basis of preceding vowel duration.

1.6 The effect of native and non-native language on speech perception

One of the additional areas of research that this thesis seeks to address is the manner in which the phonological and durational cues of native and non-native languages may interact

when a speaker is operating in a second language. Final devoicing is a linguistic phenomenon in which the voicing contrast is neutralised word-finally in the surface form. As previously discussed, this phonological process occurs in languages such as German and Dutch.

Pallier et al. (2001:445) acknowledge that '*correctly perceiving and producing the sounds of a second language are notoriously difficult tasks*'. For example, native Japanese speakers have difficulty distinguishing between English [r] and [l] as both are phonologically mapped onto Japanese /l/ (Goto, 1971; McClelland et al., 1999). Polivanov (1932) and Trubetzkoy (1939/1969) posit that:

> 'The native-language phonological system acts as a 'sieve', filtering out the properties in the speech signal that the first language (L1) system cannot properly accommodate.'
>
> Pallier et al. (2001:445)

The results of the study conducted by Pallier et al. (2001) suggest that non-native speaker's lack of sensitivity to phonemic contrasts that occur in a second language extends to the way in which L2 words are stored in the mental lexicon. They write:

> 'If listeners have difficulties perceiving an L2 phonemic contrast, they will represent L2 word pairs with that contrast as homophones.'
>
> Pallier et al. (2001:448)

Similarly, Best and Tyler (2006) posit that contrasts at the phonological level may lead to phonemes being categorised differently when operating in an L2 language. They give the example of a native English speaker who is learning French. French and English have contrastive /b/ and /p/ phonemes, and speakers may therefore develop two different categories for these phonemes; one for French and one for English. However, '*there is a mismatch*

between the phonological and phonetic properties of stop contrasts in these two languages. French short-lag unaspirated [p] as the phonetic category for the phoneme /p/ overlaps phonetically with the English phoneme /b/, which is often realised phonetically as short-lag unaspirated [p] word-initially'. Best and Tyler (2006:23-24) propose that a dissimilation may therefore occur between [p] in English and French *'and perhaps between the two interlanguage phonemes /p/ and /b/, which might cause a global shift of all related L1 and L2 categories at the phonetic level, for example, on dimensions other than primary voice onset time differences'*.

Regarding the relationship between L2 production and perception, Fledge (1999) found that there is a correlation. This was despite earlier research suggesting that there is no correlation between how late learners produce and perceive speech in a second language. Interestingly, Fledge (1999:1276) posits that:

> 'Late bilinguals who have established a category for a sound that is found in the L2 but not in the L1 would be expected to produce the L2 sound more accurately than those who have not established a category.'

As such, he suggests that:

> 'It may be useful in future research to examine the relation between production and perception using discrete tests of category formation instead of, or in addition to, continuous tests of subjects' use of specific perceptual 'cues' such as VOT.'
>
> Fledge (1999:1276)

Here, reference is made to one of Fledge's earlier studies (Fledge et al., 1995) which investigated native Spanish speakers' production and perception of English word-initial voiceless stops. Participants were required to give a goodness rating for *'preferred VOT values*

for /p/ shifted according to (simulated) speaking rate' (Fledge, 1999:1276). The results suggested that native speakers of Spanish who most closely resembled the behaviour of the native English speakers had established a category for the *'long-lag /p/ of English. The subjects who could be credited with having formed an English /p/ category were found to produce stops with English-like long-lag VOT values'* (Fledge, 1999:1276). Similarly, Fledge (1988) found that native speakers of Mandarin Chinese who spoke English as a second language did not produce a significant durational difference between English CVCs differing in word-final voicing. The results of this study were attributed to the fact that the native speakers were likely to be unfamiliar with final stop consonants as they do not occur in Mandarin Chinese.

Conversely, there are some studies which suggest that the production and perception of first and subsequent languages do not affect one another. Fledge and Port (1981) investigated native speakers of Saudi Arabian Arabic who spoke English as a second language. They employed stimuli that varied in word-final voicing, such as /pi:z/ *peas* and /pi:s/ *peace*. In Arabic, there is a phonemic contrast between vowels which are long and short. One would therefore expect that participants would be particularly sensitive to this phonological cue when operating in English. However, the results showed that speakers showed little difference in vocalic duration in the production of Arabic or English CVCs. However, they were sensitive to the perception of word-final voicing dependent on vowel duration when operating in English. These inconsistencies could be due to methodological considerations. For instance, Crowther and Mann (1992:712) acknowledge that '*Arabic speakers may not have perceived the tokens in a native-English-like manner. Rather, they may have applied a post-perceptual strategy of identifying tokens with that they perceive to be phonemically long vocalic segments as "peas" and short vocalic segments as "peace".* Equally, it could be that varying combinations of language-specific factors, such as the acceptance of word-final stop consonants or the presence

of phonemically long and short vowels, may lead to differing abilities for non-native speakers of English to emulate vocalic duration as a cue for word-final voicing in a native-like manner.

Crowther and Mann (1994) revisited this research and combined it with the concept that speakers may also use the offset frequency of F1 as an additional cue when operating in English. Crowther and Mann (1994) therefore investigated the role of both vowel duration and F1. Utterances containing longer vowel durations terminate with relatively low F1 offset frequencies which cue a voiced word-final stop. Conversely, utterances containing a shorter vowel duration end with higher F1 offset frequencies, leading to the perception of a voiceless word-final stop (Raphael, 1972; Summers, 1987, 1988; Walsh and Parker, 1983; Wolf, 1978). Primarily, this paper addressed remaining questions regarding the use of these two factors in languages that are not English. Knowledge regarding the effect of vowel duration on speech perception up to this point had largely originated from production studies. This study therefore drew upon Fledge and Port (1981) and investigated the influence of both vowel duration and F1 offset frequency on the perception of voicing. They incorporated native speakers of Arabic who were L2 English speakers. As referenced above, previous findings suggest that Arabic does not demonstrate a tendency to lengthen a vowel before a voiced stop (Fledge and Port, 1981). Despite this Arabic is *'a language that includes a phonemic vowel length distinction and word-final stops as well. Although vocalic duration is probably not used as a voicing cue in Arabic consonant-vowel-consonant…syllables (Fledge [and] Port, 1981), the phonemic vowel length account nevertheless predicts that native Arabic speakers should use vocalic duration as a voicing cue in English, because Arabic includes a phonemic vowel length distinction'* (Crowther and Mann, 1994: 514). Munro (1990) had found that native speakers of Arabic could learn to distinguish between the French front rounded vowels /y/ and /ø/ which do not occur in either English or Arabic. As such, Crowther and Mann (1994:514) write *'that*

the native Arabic speakers are better able to learn to distinguish a novel vowel pair on the basis of duration suggests that they may readily learn to make voicing distinctions on the basis of vocalic duration differences in English'. The first experiment aimed to recreate an effect found by Fledge and Port (1981) that demonstrated that native Arabic speakers produced a small difference in vocalic duration between English CVCs ending in voiced and voiceless stop consonants, with slightly longer vowel durations being produced before a voiced stop. The results of from this experiment were in support of Fledge and Port's (1981) findings. Vowel differences produced prior to voiced and voiceless word-final endings, such as /pɒd/ *pod* and /pɒt/ *pot*, were small, but significant. Turning from production to perception, the second and third experiments were designed to replicate Fledge's (1984) findings. Here, an asymmetry was present as the findings from Fledge and Port (1981) show only a small durational contrast in the production of vowel duration as a basis for voicing, and yet Fledge (1984) found a significant perceptual sensitivity to vowel duration as a perceptual cue for voicing. This could be due to native Arabic speakers adopting a 'response strategy' in order to complete the perceptual task, meaning that they were unable to perceive the difference between /pi:z/ *peas* and /pi:s/ *peace*. For the second experiment, the researchers '*employed a factorial combination of [the] two cues and a finely spaced vocalic duration continuum. Arabic speakers did not appear to be very sensitive to vocalic duration, but they were about as sensitive as native English speakers to F1 offset frequency'* (Crowther and Mann, 1994:513). Additionally, in the third experiment, native Arabic speakers showed a near-native-like sensitivity to vowel duration in English. This experiment incorporated a one-dimensional continuum of stimuli which were spaced more widely apart from another. These stimuli varied in vowel duration. Despite these somewhat contradictory perceptual findings, Crowther and Mann (1994) agree

with Fledge and Port (1981) in that Arabic speakers do not seem to have the capacity to perform native-English-like perception of word-final voicing based on these two cues.

Varying evidence therefore exists in support of the extent to which a listener's native phonology may interact with their L2 phonology, along with the extent to which a they can override their native cues and adapt to more native-like L2 perception. The results of this thesis demonstrate an interaction between L1 and L2 phonology, with German speaker's perception of English word-final voicing being constrained by their L1 phonology. This introduction to the existing research on this topic will be further elaborated on in Chapter Two.

1.7 Summary and introduction to research questions

Of the five experiments presented in this thesis, Experiment 1 consisted of two production studies investigating vowel duration, voicing, and the role of the lexicon for both L1 and L2 English speakers. Experiments 2 and 3 look to transfer these findings from production to perception and consisted of two identification tasks focused on native English speakers. Experiment 4 encompassed an additional identification task, extending the results of the previous experiments by incorporating both native English and native German speakers operating in L2 English. Finally, Experiment 5 consisted of a lexical decision task with fragment priming involving native English speakers. The stimuli used for the experiments also varied between recorded and synthesised speech. The effect of the lexicon was considered in Experiments 1, 2, 4 and 5.

To summarise, the overarching aim of this thesis is to provide a more nuanced understanding of the importance of vowel duration as a primary perceptual cue for word-final voicing in English. We additionally consider the role of the lexicon and its interaction with acoustic information in this perceptual process. Incorporating L2 speakers into this thesis allows for

inferences to be made regarding the extent to which non-native speakers of English are sensitive to this fine-grained cue. For instance, if allophonic length is used by native speakers as an important cue, to what extent is this important in an L2 English speaking population where the native language does not have such a cue? As previously stated, German has a word-final devoicing rule, and therefore German phonology differs from English phonology in that there is no allophonic voicing contrast. When operating in L2 English, therefore, to what extent will L1 German speakers be influenced by their native phonology?

The five experiments reported in this thesis contain their own more specific research questions and predictions. However, each of these experiments relate to three overarching research questions that this thesis aims to answer. Firstly, can a speaker of English anticipate the [VOICE] feature in a word-final obstruent and consequently lengthen the preceding vowel in production? Secondly, to what extent does the listener interpret the preceding vowel length as long or short and thereby anticipate and assign the corresponding word-final [VOICE] feature? Finally, to what extent are non-native speakers of English able to anticipate and interpret these tendencies?

In line with the FUL model, the predictions of this thesis suggest that English speakers will vary vowel duration in accordance with the nature of the word-final [VOICE] feature, due to its salient representation. The predictions of the FUL model also posit that English listeners will demonstrate the ability to extract the relevant information from the acoustic signal, specifically the preceding vowel duration, to assign the appropriate word-final [VOICE] feature in stimuli where the word-final obstruent has been omitted or made ambiguous. Finally, the underlying nature of the [VOICE] feature and its representation in an L1 is predicted to impact how the perception of voicing is manifested in an L2 in relation to vowel duration, thereby constraining the perception of non-native speakers of English.

Before the findings from each of the aforementioned experiments are presented, let us first consider a more in-depth exploration of the literature.

CHAPTER TWO

Literature review

This chapter aims to give a review of the previous research that has informed this thesis, building on the discussions introduced in Chapter One. This chapter will address the varying approaches that have been carried out in the past to address this area of study, and discuss the ways in which the different methodologies and findings from these papers have influenced this research.

Given the wealth of literature on this topic, this chapter has been divided into several key areas relating to the research questions that this thesis aims to answer. Firstly, it is important to further investigate some of the specific models that have been suggested to account for the ways in which humans perceive and decode the speech-stream. Both autonomous and interactive models are discussed. We then turn to focus on the relationship between vowel duration and word-final voicing specifically, looking first at the role of vowel duration in the production of word-final voicing. This section will incorporate literature that considers the differences in this production between native and non-native speakers of English. Here, we explore the extent to which L1 and L2 voicing cues may interact in speech production. Turning then to perception, several papers will be discussed which either support or refute the hypothesis that vowel duration is a primary phonological cue that listeners use in predicting and perceiving word-final voicing in English. In the case of the latter, several papers will be discussed that attribute perceptual cues to factors other than vowel duration. The role of the lexicon will then be considered. In particular, the interaction that it has with acoustic information in providing perceptual cues. Finally, vowel duration in relation to word-final voicing in languages other than English will be explored. As part of this discussion, the

interaction between native and non-native language when perceiving word-final voicing in an L2 language will be evaluated.

2.1 Models of speech perception

As introduced in Chapter One, speech perception may be defined as the process of recognising and identifying individual phonemes or features within the speech-stream. Key questions surrounding speech perception focus on a human's ability to distinguish between phoneme sequences that differ only slightly, such as [bæ] *ba* and [pæ] *pa*. Do we perceive the phonemes in these sequences as belonging to distinct categories, or as points along a continuum? Experimentation has thus focused on deciphering the nature of perception, and whether it appears to be categorical or continuous. We reference Liberman et al. (1957, 1961a, b), who used synthetic CV syllables ranging in '*an acoustic parameter (e.g., the slope of the F2 transition) and ranging perceptually across several initial consonants (e.g. /bV/ -/dV/-/gV/)* Diehl et al., 2004: 155) to explore this. These stimuli were played to participants in a perceptual experiment and listeners were asked to identify or discriminate between the phonemes. As referenced earlier, categorical perception may be defined as the perceptual tendency to categorise phonemes that exist along a continuum into distinct groups. Specifically, this experiment reported two distinct patterns of results. They write:

> '[Firstly] labelling functions exhibited abrupt boundaries between phoneme categories… [and secondly] discrimination accuracy was close to chance for stimulus pairs within a phoneme category but nearly perfect for stimulus pairs that straddled an identification boundary.'
>
> Diehl et al. (2004: 155)

This indicates that when presented with stimuli which exist along a continuum, listeners are able to place boundaries between phonemes that belong to different categories more easily than phonemes that are within the same category.

The significance of categorical perception became apparent when it began to influence early ideas for constructing a model for speech perception. Initially, it was theorised that the explanation for categorical perception came from the anatomy of speech production (Harnad, 2003), and this gave rise to the motor theory. Research conducted in the 1950s by researchers at the Haskins Laboratories began to centre on the idea that *'perceived phonemes and features have a simpler (i.e. more nearly one-to-one) relationship to articulation than to acoustics'* (Diehl et al., 2004: 150). The motor theory of speech perception therefore hypothesises that listeners use the movements of the vocal tract to identify phonemes. This theory centres on the invariant motor commands supplied to the muscles needed to move the articulators and as such, the invariant nature of the speech articulators themselves. Consider once again that humans will tend to perceive [bæ] *ba*, [gæ] *ga*, or [pæ] *pa*, but nothing in-between. The motor theory argues that this is because human articulators are not capable of producing anything half way between two categories; our vocal folds are either vibrating, for instance word-initially in [bæ] *ba* or, they are not vibrating, such as word-initially in [pæ] *pa*. Equally, the articulators may either be positioned at the lips to produce bilabial sounds such as [bæ] *ba* or [pæ] *pa*, or at our velum as is the case for [gæ] *ga*. It is not possible to form two distinct positions of articulation at the same time. In addition, the motor theory infers that there is a specialised decoder or module that is unique to humans, and this mechanism allows us to identify and categorise phonemes.

However, the motor theory has faced several criticisms. The theory that the motor commands driving the anatomical structures are responsible for phonetic invariances was

proven to be improbable by three main developments in the literature at that time. Galantucci et al. (2006) writes:

> 'Turvey (1977) developed a theory of action in which the motor system was to be understood in terms of functional units—called coordinative structures (cf. Easton, 1972)— rather than anatomical structures. Next, Fowler, Rubin, Remez, and Turvey (1980) extended Turvey's theory to the domain of speech production. Finally, Browman and Goldstein's articulatory phonology (Browman and Goldstein, 1986) identified coordinative structures as fundamental linguistic units that they called phonetic gestures.'
>
> Galantucci et al. (2006:5)

Here, coordinative structures refer to groups of functional units, usually muscles, which signal the articulators to work together with the aggregate result being speech. Here, a key differential between coordinative structures and anatomical structures is the flexibility with which these coordinative functional units can operate. Whereas the anatomical model suggests that speech is organised according to a fixed input-output model, coordinative structures account for more varied cooperation from the articulators, allowing for adjustments to be made within speech in accordance with different phonetic requirements.

Another criticism of the motor theory is that it infers that one must understand how to produce speech in order to be able to perceive it (Harnad, 2003). However, Eimas et al. (1971) found that infants exhibit categorical perception even before they learn to speak as adults. Indeed, research has also found that categorical perception is not necessarily restricted to humans (Kuhl and Miller, 1975 a, b, 1978; Kuhl and Padden, 1983). Both Morse and Snowdon (1975) and Kuhl and Miller (1978) found that:

> 'In English-speaking adults, chinchillas, and monkeys, the category boundary between /ba/ and /pa/ is around 30ms…for VOT <30ms the phoneme is mostly perceived as /ba/, and for VOT > 30ms as /pa/.'
>
> Eggermont (2015:2)

In light of these developments in the theory of speech perception, Liberman and Mattingly (1985) reviewed the motor theory and incorporated Browman and Goldstein's (1986) idea of phonetic gestures. Here, a distinction was made between intended gestures, pre-vocal and at a linguistic level, and the actual movements that occur as part of articulation. As such, *'Liberman and Mattingly identified intended gestures, and not actual vocal tract actions, as the fundamental objects of speech perception'* (Galantucci et al., 2006:5).

The direct realist theory of speech perception was first developed in the 1980s, again through research being conducted at the Haskins Laboratories (Fowler, 1981, 1984, 1994, 1996). As an alternative to the motor theory, the direct realist theory *'asserts that the articulatory objects of perception are actual, phonetically structured, vocal tract movements, or gestures, and not events that are casually antecedent to these movements, such as neuromotor commands or intended gestures'* (Diehl et al., 2004: 152). To contrast the motor theory, the direct realist theory refutes the concept that there is a specific mechanism for speech perception. Instead, a speaker's gestures, as signalled by the movement of the articulators, structure the acoustic information received by the listener, and from this signal the listener can recover the intended gestures. Lotto and Holt (2006:181) criticise Fowler (2006) in relation to both the motor and direct realist theories, recognising that both speech contexts and speech targets influence perception. They write:

> ‘There are no coarticulatory influences across two distinct speakers. If listeners have a special system that has evolved a knowledge of vocal tract dynamics (motor theory) or can pick up these dynamics directly from the signal (direct realism), they should not compensate across independent vocal tracts.’
>
> Lotto and Holt (2006:178)

An alternative approach is the TRACE model of word recognition (McClelland and Elman 1986), which suggests that there is an interaction, or network, between speech perception and lexical perception. This challenged autonomous models. The TRACE model hypothesises that there are different levels of a perceptual network, and that speech units such as features, phonemes, and words inhabit different levels of this network. For example, regarding features, McClelland et al. (2006) write:

> ‘One feature bank represents the degree of voicing, which is low for unvoiced sounds such as /t/ and /s/ and higher for voiced sounds such as /d/ and /z/. At the phoneme and lexical levels, one unit stands for each possible phoneme or word interpretation of the input. These sets of units and the connections between them are duplicated for as many time slices as necessary to represent the input to the model.’
>
> McClelland et al. (2006: 365)

This model theorises that the speech units across different levels share excitatory activation; once a feature is activated it will activate the related phoneme, which will in turn activate the relevant units on the word level. To pinpoint an exact target, the activation of one speech unit within a level would inhibit the activation of other competing speech units on that same level. Norris et al. (2000) reported on several criticisms of the TRACE model, for example the lack of evidence it provides in support of lexical inhibition of phoneme recognition, and subcategorical mismatches. However, McClelland et al. (2006:366) acknowledge these

criticisms and present research containing counterarguments to these issues. Firstly, Norris et al. (2000) reference that work conducted by Frauenfelder et al. (1990) and Wurm and Samuel (1997) failed to show evidence of lexical inhibition as, in their study, no evidence of slowing effects occurred in the recognition of phonemes which were contextually inconsistent. This caused doubt regarding the nature of interactive processing. However, McClelland et al. (2006:366) reference Mirman et al. (2005) whose results supported the notion '*that the TRACE model…correctly expected the conditions required to show lexical inhibition'*. Secondly, the data acquired from Marslen-Wilson and Warren (1994) and McQueen et al. (1999) did not fit with the TRACE model regarding the influence of '*lexical status on subcategorical mismatch'* (McClelland, 2006:366). Subcategorical mismatches occur when there is an incompatibility between two pieces of linguistic data presented to a participant from within the same category. For example, if *'the initial consonant and vowel of the word jog was spliced onto the final consonant from the word job…a subcategorical mismatch between the coarticulatory information in the vowel of the spliced stimulus and the final consonant'* would be created (Yantis and Pashler, 2002:508). However, later research using eye-tracking (Dahan et al., 2001) and an analysis of global model behaviour (Pitt et al., 2006) found that the TRACE model was consistent with the data, and this once again does support the interactive model.

Conversely, the cohort theory (Marslen-Wilson and Welsh, 1978) suggests that perception operates based on a cohort of lexical entries. According to the cohort theory, all lexical entries within a cohort share the same acoustic information; provided by the speaker to the listener. As the speech stream unfolds and more acoustic information is obtained, members of the cohort which are no longer consistent with the input drop out. For example, if the initial input [kæp] *cap* of the word [kæptɪn] *captain* is provided, a cohort of lexical entries such as [kæp] *cap*, [kæptɪn] *captain*, [kæpsaɪz] *capsize*, [kæpʃən] *caption*, [kæpɪtl] *capital*, etc., would be

activated. Once the speech signal develops into [kæpt] *capt*, all the above words apart from [kæpɪtl] *capital* and the target word [kæptɪn] *captain* would be omitted from the cohort. This process continues until a single target word remains. This is a similar model to the neighbourhood activation model (Luce, 1986; Luce and Pisoni, 1998) which suggests that word recognition happens based on identifying a target from among a set of activated candidates. This theory is based on the probability of the target word compared with the other words in that neighbourhood, all of which are phonologically similar. This probability is also influenced by lexical frequency.

The final family of models that we will address are the exemplar models (Goldinger, 1997; Johnson, 2005). These offer a non-analytical approach, and suggest that mental representations do not need to be highly abstract. Instead, these models suppose that information such as the speech style of a particular speaker, or the environment in which the speech occurs, is stored. In order to categorise a word, it is compared with all previous instances of that category. Remembered examples are activated to a greater or lesser extent depending on their similarity to the incoming speech stream, and the stimulus with the greatest degree of activation determines categorisation.

Models of speech perception which centre on the recognition of whole words as opposed to single phonemes form an important part of the theory informing this thesis as this research is interested in the role of the lexicon and how this may interact with acoustic information. Here, a distinction should be made between these two approaches. Models, such as the TRACE model, which posit that speech perception begins with a 'lower' perceptual unit, such as a feature or phoneme, which then activates the relevant unit at the 'higher' word-level, imply that the target lexical item is not represented from the onset of speech. Alternatively, models such as the cohort model, suggest that lexical units are represented from the outset and as more

information is supplied from 'lower' levels, specifically the phonetic units, these whole word units drop out as they become incompatible with the acoustic information. As such, there is undeniably an interaction between speech perception and spoken word recognition. To summarise how we may reconcile these two approaches, Samuel (2010) writes:

> 'The field is now at a point when it is both possible and desirable to bring these two subfields together…to bring improved insights into how humans accomplish the extraordinary feat of understanding spoken language.'
>
> Samuel (2010:16.20)

As introduced in Chapter One, overall the findings of this thesis are in support the FUL model of perception (Lahiri and Reetz, 2002). Recall that this model posits that speech is organised according to sets of binary features. The FUL model suggests that [VOICE] is a specified feature and that the nature of word-final voicing therefore has a salient representation in the minds of listeners. Here, the FUL model brings a simplicity to the otherwise more complex representations of speech perception suggested by competing models. The binary nature of the [VOICE] feature and its presence or absence regardless of the acoustic variability of speech seems far neater when compared with representations of voicing on a variable scale, or in models which posit that listeners must retrieve all previous instances of speech from their store in order to interpret the nature of voicing each time they encounter a new utterance. As such, the FUL model provides a more efficient processing pathway for the discrimination of the variable speech stream, and the identification and categorisation of phonemes across languages.

In order to assign the appropriate features, the FUL model relies on acoustic information. For the [VOICE] feature, the FUL model takes into account several factors, including the

vibration of the vocal cords and also the duration of the preceding vowel. Of course, there is no absolute measure of acoustic information and therefore the FUL model posits that listeners obtain the maximum interpretation of the acoustic signal possible before accessing the lexicon where it is decided if there is a serious ambiguity, with certain features being considered more important than others. Consequently, any acoustic information that can activate the [VOICE] feature is crucial, and the implications of this thesis support this notion with regard to preceding vowel duration. Though vowel duration is not specified for a feature according to the FUL model, the acoustic information provided to a listener through the vowel duration proves fundamental in allowing the relevant information to be extracted from the signal and translated into the word-final [VOICE] feature in English. Wherein the vowel is interpreted as long, this feature is assigned. The findings from Experiments 1-5 support the salience of the [VOICE] feature and its relationship with preceding vowel duration in English. Additionally, the lexicon was found to have effects on the data which suggests that both lexical representation and bias are somewhat relevant in deciphering the nature of word-final voicing in English, though these lexical effects are not as influential as the acoustic information provided by the vowel duration.

To summarise, the research discussed here seeks to address both models of speech perception and models of units of speech perception. Naturally, these two approaches are linked in that the perception of individual units at a low level, such as features and phonemes, feed into the interpretation of speech at a higher level, such as whole word and sentence recognition. In particular, the categorical and continuous models of speech perception characterise the way in which the units of speech perception are represented and interpreted. Equally, models which address the organisation of the units of speech from lower to higher levels demonstrate the evidence needed to support models of speech perception. In this way,

the link between these two approaches become directly interdependent, for example the evidence in support of categorical perception deduced from the motor theory.

Given the complexity of speech perception, there is no clear way to determine whether a single model is unequivocally correct. Therefore, the research in this thesis therefore aims to provide a more nuanced view of the interaction between acoustic and lexical information, and will provide further insight into key variables that speakers take into consideration when decoding the speech stream.

2.2 Speech production: vowel duration and voicing

Before we turn our attention to the perception of word-final voicing in relation to vowel duration, it is important to consider the interaction between these two factors in the production of speech.

The first experiment encompassed in this thesis consists of two separate production studies. As such, it is important to note the methodological difficulties and limitations of both the production studies in this thesis and of those which have been conducted by previous researchers. In 1992 van Santen conducted a production study involving two speakers. They write:

> 'Studying durational phenomena in speech through segmental duration has its limitations…segmental boundaries are to some extent arbitrary, not reflecting discrete articulatory events but being determined by conventions based on easy-to-detect acoustic features. Thus, segmental durations may not correspond to any direct way to underlying physical or neural processes.'
>
> van Santen (1992:515)

Therefore, difficulties arise from the complexity of agreeing on one single, correct method for dissecting the continuous flow of speech. Fant (1960:208) writes that *'the…trouble occurs if investigators insist on measuring or defining phoneme durations'*. The very nature of production studies makes avoiding this issue difficult, and taking accurate measurements for the vowel durations within this thesis forms an integral part of the research. To overcome this issue, consistency within experimentation is key, as is a having a nuanced view of the extent to which production studies may be compared with one another. As such, the recordings encompassed in Experiment 1 were measured with as much consistency as possible. The results were also only compared within each production study, as opposed to being compared between the two studies, or indeed to specific measurements that have been obtained from production studies within the literature. Standard errors are also incorporated into the analysis of Experiment 1 to account for any inaccuracies made throughout the measurement process. Despite these limitations, production studies still provide an important insight into the acoustic tendencies of speakers. As such, if these studies are conducted with stringent consistency between speakers and compared like-for-like, they can provide important information used answer research questions such as those in posited in Experiment 1.

Turning now specifically to the production of vowel duration, Van Santen and Olive (1990) characterise the nature of this by explaining that it depends on:

> 'Several factors, including the phonemic identity of a segment, speaking rate, lexical stress, syntactic stress, syllabic position, being at a…boundary, and phonetic context. The joint effects of these factors can cause vowel duration to be as short as 25ms and as long as 300ms (Crystal and House, 1988).'
>
> Van Santen and Olive (1990:359)

As such, the nature of vowel production can be attributed to an amalgamation of factors which, when combined in varying ways, result in durational differences. This thesis focuses on differences in vocalic duration where the vowel immediately precedes a word-final phoneme in English, and how the relationship between vowel duration and word-final voicing may be linked. As referenced in Chapter One, the phonological tendency in English is that, in word-final minimal pairs, the voiced word-final consonant is preceded by a notably longer vowel than the corresponding voiceless word-final consonant. However, it is important to reiterate that this thesis does not intend to suggest that the relationship with voicing is the sole factor driving this durational tendency in English. Instead, the research in this thesis endeavours to take into consideration the multiple factors referenced above, and without denying this complexity, support the notion that word-final voicing has a primary effect on the shortening and lengthening of the preceding vowel in English, both in the surface form and underlyingly.

There are several proposed theories for the relationship between the production of vowel duration and the nature of voicing in English, some of which were previously introduced in Chapter One, as detailed in Beguš (2017). These theories highlight the importance of the surrounding context on the production of the vowel, and the way in which neighbouring segments may influence one another during speech perception. In relation to this, Coleman (2003: 353) states that *'the 'final' voicing contrast…is sometimes reflected in slight differences of the word-initial consonant. In particular, the /l/ of lend may be slightly longer and darker (slightly more velarised) than that of lent'*. With regard to these observations, is favourable to conclude that the production of vowel duration in relation to voicing is the result of several physiological attributes, as detailed earlier (Beguš, 2017).

Raphael (1975) contributed to this discussion and addressed whether variations in vowel duration are physiologically prescribed to speakers of English. Raphael (1975) conducted two

electromyography studies (hereafter, EMG). He did this in order to further investigate the physiological nature of the muscular activity required by the articulation of CVC syllables containing identical vowels, differing only in the voicing characteristic of the syllable-final consonant. EMG is a technique used to record the electrical activity produced by muscles across different actions. For the initial experiment, participants were required to read out minimal pairs, for example [li:v]~[li:f] *leave~leaf*. For the second experiment, disyllabic minimal pair utterances were incorporated, such as [əpep]~[əpeb]. From the results, Raphael (1975) acknowledges the notion that vowels following voiceless consonants are generally shorter than those following voiced consonants. He attributes this to two potential hypotheses. The first of these hypotheses is the notion that there is a greater duration of muscular activity needed when a vowel precedes a voiced consonant. The second hypothesis suggests that an equal duration of muscular activity is needed for both vowels preceding a voiced consonant and vowels preceding a voiceless consonant. Reminiscent of the de Jong's (1991) hypotheses, the variation in duration would, in the case of the latter, be:

> 'Affected by a difference in the timing of the onset of muscular activity of the following consonants in relation to the offset of preceding vowel activity: relatively earlier in the voiceless case and relatively later in the voiced case.'
>
> Raphael (1975:26)

His results support the first of these hypotheses, and provide evidence for the notion that the variation in duration between vowels in English can be physiologically attributed to the motor commands given to the articulators.

However, one must not neglect influence phonology in speech production. If the production of vowel duration fell simply to physiological reasons such as the duration of the following

consonant, we would expect to see evidence of the same patterns of vowel lengthening and shortening happening across languages. Of course, this is not the case as there are many languages, such as French (Mack, 1998), which do not exhibit such a large voicing-dependent effect on vowel duration. House (1961) argued that vowel lengthening in English is therefore a combined result of being both learned aspect of the phonology of English, and a consequence of the process of articulation itself. In support of the results of this thesis, he concluded that the voicing characteristic of the following consonant had the primary effect on the production of vowel duration. House (1961) recorded native speakers of American English and found that the average duration of the vowels read out as part of this experiment were affected primarily by their phonetic environment, in particular the voicing characteristic of the consonant. The results from this paper suggest that:

> 'Th[is] primary influence [on duration] is contributed [to] by the voicing characteristic of the consonant, whereas the manner of production of the consonant shows a smaller effect [and] the place of consonantal production is shown to have a negligible influence on the duration of the vowel.'
>
> House (1961:1175)

Regarding the manner of production, one of the considerations in this thesis is the relationship between vowel duration and the articulatory manner of the subsequent word-final consonant. De Jong (1991:1) argues that '*if the following consonant [after the preceding vowel] is a fricative, the preceding vowel tends to be longer than if the following consonant is a stop'*. In this thesis, the experiments which incorporated non-native speakers contained stimuli ending in both stops and affricates. The results of this thesis suggest that this difference in the manner of production word-finally is a key factor driving the relationship between vowel duration and voicing, and in particular how this relates to the underlying phonology of L1 and L2 languages.

Klatt (1976) also discusses the influence of a post-vocalic consonant on vowel duration and once again attributes this to a combination of physiology and phonology. Here, Klatt (1976) presents the example of the [æ] diphthong in the word [bæg] *bag*, which he found to be 50ms longer than in the word [bæk] *back*. With further regard to the reasoning behind these differences in production, Klatt (1976:1214) suggests that this *'is probably a result of the natural tendency to make a slightly early glottal opening gesture for a postvocalic voiceless consonant in order to ensure that no low-frequency voicing cue is generated during the obstruent'.* This results in a shorter vowel. In addition to positing this physiological theory, he notes that in the English language, this tendency is likely to be perpetuated by a phonological rule of the grammar.

In light of this literature, this thesis carefully considers the role of underlying phonology in the salience of the word-final [VOICE] feature. As previously stated, Experiment 1 encompasses a production study which incorporates both native and non-native speakers of English. Studies have previously shown that it is difficult to replicate L1 phonological cues in an L2 language if the L1 language does not possess that same phonological cue. Edge (1991) details a study in which native speakers of Cantonese were asked to produce English word-final voiced obstruents. These productions were then compared with those produced by native English speakers under the same conditions. Edge (1991:379) writes *'Cantonese, unlike…English, has no voice contrast in any position, as it only has voiceless obstruents…English…has a voice contrast for obstruents in all positions'*. This study incorporated results from Cantonese speakers, and for comparison, base-line data was recorded from native English speakers. Three elicitations were conducted, the first was a storytelling task designed to elicit words ending in voiced obstruents, for example [rɛd] *red,* [hɛd] *head,* and [dɒg] *dog*. In Task 2 participants were required to read out a story which contained tokens ending in voiced obstruents. Similarly,

for Task 3, participants read out a list of randomly ordered words containing voiced and voiceless obstruents in all positions. Examples include [raɪdɪŋ] *riding*, [ðɪs] *this*, [græb] *grab,* and [ʤəʊk] *joke.* The results showed that Cantonese speakers devoiced a significant amount of the time, further illustrating that the production of an L2 language is directly influenced by the phonology of the L1 language. Importantly, this paper also highlighted that it is crucial to directly compare non-native speech samples with matched native speech samples. This influenced the methodology of Experiment 1 which contains direct comparisons between native and non-native speakers of English performing the same elicitations.

In summary, this literature on the production of vowel duration in relation to voicing proves relevant for this thesis as it provides a theoretical framework in which the present research may be situated. Developing a more nuanced view of the physiological and phonological reasons for the relationship between these two factors in production is crucial in understanding how this relationship may be perceived. Of course, the production of speech is complex, and there are many factors that affect the duration of any given phoneme. However, the literature outlined in this section is consistent in providing evidence in support of the influence of consonantal voicing on the production of preceding vowel duration. Additionally, the importance of the underlying phonology is widely recognised, alongside the phonetic attributes of the surface form. This is supported by the results of this thesis.

2.3 Speech perception: vowel duration and voicing

Having now considered the relationship between vowel duration and voicing in speech production, let us turn to perception. Previous literature has attributed preceding vowel duration as a primary perceptual cue for voicing in English to varying degrees. This section will discuss literature in support of vowel duration as a primary cue, and conversely, literature that attributes

perceptual voicing distinctions to other factors. Studies that address voicing in different positions in words will be discussed, contributing additional information and insight to the theoretical basis of this thesis.

As early as 1950, Jones (1950) alluded to the notion that in English *'words like heed and heat…are distinguished solely by the length of the vowel'* (Raphael, 1972:1296). This encouraged further investigation into the extent to which vowel duration can be held accountable as a primary cue for the perception of voiced and voiceless word-final phonemes in English. A multitude of previous studies have since attributed vowel duration as having the ability to provide perceptual information for voicing (Ainsworth, 1972; Denes, 1955; Klatt and Cooper, 1975; Raphael, 1972). Raphael (1972) explains that there appears to be a recognised relation between vowel duration and the voicing characteristic of the following word-final consonant. This means that the duration of the vowel appears to be providing the listener with acoustic information that directly affects their perception of word-final voicing.

Haskins Laboratories (1956) conducted an experiment investigating the word-final velar stops /g/ and /k/. The stimuli employed in this experiment were synthetic in nature, and all voicing cues had been neutralised. The experiment found that listeners perceived hearing [Eg] when the preceding vowel was longer than 200msec, and conversely [Ek] when the preceding vowel was shorter than 200ms. This suggested that there was a perceptual interaction between vowel duration and word-final voicing. Around this time, Denes (1955) conducted a study investigating the perception of voicing in word-final fricatives, depending on preceding vowel duration. Using the words [juz] *use* and [jus] *use,* Denes (1955) systematically shortened the duration of the /s/ phoneme whilst lengthening the duration of the /z/ phoneme. Vowel durations subsequently measured at 0.05, 0.1, 0.15, and 0.2 seconds, whilst the consonant durations increased in five 0.05 second steps from 0.05 up to 0.25 seconds, respectively.

Participants were asked to indicate whether they perceived hearing the phrase *to use* or *the use* by writing down a corresponding *Z* or *S*. The results suggested that, *'the perception of 'voicing' of the final consonant increases as the ratio of the duration of final consonant to preceding vowel decreases'* (Raphael, 1972:1296). There are some criticisms of this study, such as a potential orthographic bias towards participants responding with an [s] given that both *to use* and *the use* end orthographically with an *S*. However, despite these criticisms, this study strongly supports the hypothesis that the perception of word-final voicing in English appears to be significantly influenced by preceding vowel duration.

Raphael (1972) extended the work of Denes (1955) by investigating a variety of word-final stops, fricatives, and clusters in the hope of supporting or refuting the motor theory. In order to achieve this, Raphael (1972) conducted a forced choice identification task in which participants were asked to identify members of a minimal pair, for example [gs]~[kz] *pigs~picks*. A second study followed an oddball format in which participants were asked to identify the odd sound out of three phonemes, two of which were identical. The results from these studies found that *'preceding vowel duration is a sufficient…cue to the perception of the voicing characteristic of a word-final stop, fricative, or cluster'* (Raphael, 1972: 1301). The results also supported the notion that the presence of voicing during the closure duration of a word-final consonant *'does have some cue value* (Raphael, 1972:1299). Though, on balance, this information was considered to be less influential than the acoustic information provided by the vowel duration. The current thesis supports these results and concurs that preceding vowel duration is a sufficient and primary cue for word-final voicing in English. Interestingly, the results from Raphael (1972:1301) also suggest that '*the cue of preceding vowel duration is more effective before stops and clusters than before fricatives'*. This is surprising considering that the findings from de Jong's (1991) work on speech production suggest that vowel duration

is longer, and therefore more exaggerated, before a fricative than before a stop. Finally, the results suggest that *'the perception cued by the preceding vowel duration is continuous, [not] categorical'* (Raphael, 1972:1301).

Wardrip-Fruin (1982) provided a comprehensive overview of the literature on vowel duration and its relation to voicing. Three main papers are referenced; Mermelstein (1978), Greenlee (1978), and Hogan and Rozsypal (1980). As opposed to the support given in favour of vowel duration as a primary cue for voicing in the aforementioned literature, more inconsistency is presented in the experimental results of these papers. Particularly regarding the extent to which vowel duration can be considered a primary (Klatt, 1976) or necessary (Raphael, 1972) cue. Mermelstein's study varied both the vowel duration and first formant frequency in a series of synthesised CVC syllables such as [bɛd] *bed*, [bæd] *bad*, [bɛt] *bet*, and [bæt] *bat*, and produced evidence in support of vowel duration cuing a perceptual difference in the voicing distinction between word-final [d] and [t]. Conversely, Greenlee's experimental results only supported a crossover in the relationship between vowel duration and voicing in 50% of their stimuli. This study focused on six syllables and found that only three showed a change in the perception of voicing due to vowel duration. Comparatively, Hogan and Roszypal (1980) *'found that changes in vowel duration before voiced stops did not effect a category change of voiced to voiceless for most stops tested'* (Wardrip-Fruin, 1982:187).

Indeed, Hogan and Rozsypal (1980: 1764) support the notion that in CVC or CVCC configurations of English words, *'the vowel nucleus in words ending in a voiced consonant is longer than the vowel nucleus followed by its voiceless consonant cognate'*. However, they also recognise that other factors may provide additional cues for the voicing distinction, such as the tenseness of the preceding vowel, along with vowel height and following consonant type; more specifically whether it is a stop, fricative, or affricate. Their experiment aimed to

'evaluate quantitively the relative contribution of vowel duration and some additional spectral and temporal cues to the perception of voicing of the following final context in natural speech' (Hogan and Rozsypal 1980: 1764). Taking influences from Malécot (1970), the acoustic cues being investigated were the *'durations of the final formant transitions, voice bar duration, and the duration of the silent interval between the voice bar and consonant release'* (Hogan and Rozsypal, 1980: 1764). Examples of the stimuli include [bʌd]~ [bʌt] *bud~but*, [hɪz]~ [hɪs] *his~hiss*, and [rɒdz]~ [rɒts] *rods~rots*. The vowels were synthetically shortened using a digital gating method and employed in a forced choice identification task in which participants were required to circle the cognate that they had perceived hearing. The results from Hogan and Rozsypal's (1980) experiment do not attribute vowel duration alone as a cue for voicing, but instead argue that listeners perceive differences in word-final consonant voicing using a combination of complex acoustic cues, namely *'vowel duration, voice bar duration, silent closure duration, and burst/frication duration'* (Hogan and Rozsypal, 1980: 1770).

Wardrip-Fruin's (1982) own study employed monosyllabic words, such as [biːd]~[biːt] *bead~beat,* which then underwent both analogue-to-digital processing and linear predictive coding. Participants were asked to judge the nature of word-final voicing. The results provide counter-support for the notion that vowel duration is a primary cue in the perception voicing, and instead suggest that the nature of voicing during closure is *'required to disambiguate final voiced stops'* (Wardrip-Fruin, 1982:187).

Alongside closure duration, nasal duration has also been suggested to contain cues for word-final voicing. Raphael et al. (1975) conducted a production study based on a series of elicitations made by adult speakers. The utterances contained minimal pairs, for example [pɛnd]~ [pɛnt] *pend~pent.* As expected, vowel duration was longer before a voiced word-final phoneme than a voiceless word-final phoneme. However, nasal duration and vowel plus nasal

duration was also longer before a voiced as opposed to a voiceless phoneme. The results from this study suggest that *'assuming the voiceless context was the base, the increase in nasal duration in the voiced case was proportionally greater than the increase in vowel duration. This outcome suggest that nasal duration is a more powerful cue to the voicing characteristic of the following consonant than is vowel duration'* (Raphael et al., 1975:389). In a second part of the study, this hypothesis was tested in a perception experiment in which listeners were presented with stimuli and asked to label the final consonant as a [d] or [t]. The stimuli consisted of synthetic CVNC utterances *'in which the nasal and vowel segments were independently varied in duration over a range of 40ms to 200ms'*. The results suggested that, when testing CVNC utterances, listeners were more indeed sensitive to changes in nasal duration as opposed to vowel duration.

In addition, Pisoni and Lazarus (1974) published a study in which native English speakers were tasked with identifying and discriminating between stimuli which varied in the duration of VOT. This experiment focused on the word-initial voicing continuum between /bæ/~/pæ/ *ba~pa*, and aimed to establish whether the sounds along this voicing continuum were processed in accordance with a categorical or continuous model of perception. The participants were split into two groups, one of which heard a randomised list of stimuli varying in VOT. The other received an ordered sequence of stimuli in which the respective VOTs were presented consecutively. This study therefore consisted of two discrimination tasks, an ABX discrimination test *'in which X was identified with A or with B'* and a 4IAX test *'of paired similarity in which two pairs of stimuli- one pair always the same and one pair always different- were presented on each trial'* (Pisoni and Lazarus, 1974:328). The group who heard both the list of stimuli in a consecutive order and who experienced the 4IAX discrimination task showed a higher level of discrimination within phonetic categories that reflected a

'noncategorical perception of the voicing distinction' (Pisoni and Lazarus, 1974:328). The paper concludes that speech sounds may be processed according to either an auditory, or phonetic model. The 4IAX task provides the listener with auditory information not present within the ABX task. Therefore, in the ABX task listeners may be forced to rely purely on phonetic input. The results of this paper were interpreted as suggesting that in speech perception, there are two separate levels for both auditory discrimination and phonetic discrimination. Importantly, this study also concluded that categorical and continuous methods of processing speech sounds *'may not be completely dichotomous. Rather, these two modes of responding to speech stimuli may represent processing of information at two, among many levels, of perceptual analysis for speech sounds'* (Pisoni and Lazarus, 1974:333). Despite considering word-initial voicing in relation VOT as opposed to word-final voicing in relation to vowel duration, this paper provided important theoretical insight into the voicing cues held within the VOT, and the way in which speech perception may be modelled.

Soon after this time, Keating and Blumstein (1978) considered the effect that varying the duration of transition lengths can have on the perception of stop consonants. They investigated the ways in which listeners can categorise stop consonants that have a range of different transition lengths. The study encompassed two experiments, the first of which consisted of forming three continua made up of transition durations timed at 45ms, 95ms, and 145ms, respectively. These continua represented the different phonetic categories between /dæ/ and /gæ/. The stimuli were then presented to participants in a labelling and discrimination task. The results of this experiment showed that:

> 'Although there was a significant change in identification performance from 95-145ms, the shape of the functions, and the locus and slope of the phonetic boundary did not significantly vary across transition lengths. In addition, discrimination of within-category stimulus comparisons was significantly better at 95ms transition length than at 45 or 145ms.
>
> Keating and Blumstein (1978:57)

The second experiment within this study therefore aimed to investigate the acoustic information available to listeners with the adaptation paradigm. The adaptation paradigm was believed to be more acoustically sensitive to the attributes of each of the stimuli, otherwise known as selective adaptation. Keating and Blumstein (1978:61) write that *'selective adaptation involves a comparison of identification performance on a test continuum before and after repeated presentation of a particular adapting stimulus'*. From this, one is usually able to see a shift in the phonetic boundary along a continuum that is geared towards the category of the adapting stimulus. In this part of the experiment, participants were tasked with labelling the 45ms series before and after the stimuli had been adapted with 45ms, 95ms and 145ms [dæ] stimuli. The results of this experiment ultimately found that transition length did not have an effect. Instead, the findings suggest that *'the slope and duration of formant transitions seem to contribute minimally to the perception of place of articulation in stop consonants'* (Keating and Blumstein, 1978:57). This paper therefore provided further insight into the role of another potential acoustic cue which may have had an impact on the results obtained from the experiments within the thesis.

More recently, Herd et al. (2010) conducted an experiment looking in part at the effect of vowel length on word-final [d] and [t] wherein these sounds are flapped in American English. Herd et al. (2010) investigated the perception, categorisation, and acoustics of [d] and [t] flaps in positions where they appear word-finally or intervocalically. This paper combined both an

acoustic and perceptual study with the aim of deciphering whether /d/ and /t/ become neutralised when produced in a flapped environment. Again, the researchers acknowledge that when a vowel proceeds a [d], it is longer in duration than when it precedes a [t]. Despite this well attested-to notion within the English language, this paper was motivated by studies such as Charles-Luce (1997) who found that *'neither vowel duration nor word duration were statistically significant independent cues'* (Herd et al., 2010:3). The methodology for the initial acoustic phase of the experiment aimed to discover how often, and in which environments, speakers of American English produced a flap. As such, the participants were asked to produce a range of words such as [raɪd] *ride*, [greɪt] *grate*, [bʌdɪŋ] *budding*, and [pɛtl] *petal*. The results appeared to show that word frequency and morphological complexity did not affect the frequency of flaps within the data. A subsequent perceptual study followed on from this. A set of four naturally produced word pairs of varying word frequency were utilised, [li:də]~[lɪtə] *leader~litter*, [wɛdɪŋ]~[wɛtɪŋ] *wedding~wetting*, [taɪdl]~[taɪtl] *tidal~title*, and [mædə]~ [mætə] *madder~matter*. A forced choice identification task was then constructed in which participants were required to choose between two words, for example [li:də] *leader* and [lɪtə] *litter*. The stimuli had been manipulated for *'underlying representation of (/t/ or /d/), vowel duration preceding the flap, and word frequency'* (Herd et al., 2010: 1). The results from this experiment concluded that the effect of vowel duration alone did not have a significant impact, but that instead several other cues were at play. Interestingly, in the case of the perceptual study, they found a word frequency effect on the perception of the consonants, with higher frequency words being correctly identified more often than lower frequency words. The effect of word frequency will be reviewed in more detail in the subsequent section of this literature review, wherein the role of the lexicon in speech perception is considered. However, the notion that higher frequency words were perceived more accurately than lower frequency words illustrates

that there is an interaction between acoustic and lexical information, and suggests that lexical information has the capacity to activate or inhibit the perception of phonemes. Here, some methodological differences must be considered in that, in this thesis, participants were required to identify single phonemes as opposed to whole words in the identification tasks. This limits the extent to which these studies can be directly compared. Nevertheless, the results from this thesis suggest that higher frequency words did demonstrate the expected outcomes more so than lower frequency words in Experiment 2 and 4, and therefore the effect of word frequency is considered as part of the analyses.

Indeed, in addition to the influence of vowel duration, changes in the perception of voicing have been attributed to multiple other perceptual cues. Here, we have discussed a number of these additional cues including the closure duration, nasal duration, VOT, transition length, and word frequencies of stimuli. The overarching conclusions of each of the papers mentioned within this section support the notion that manipulating fine-grained phonetic quantities will significantly alter speech processing and perception. In summary, it is clear from the literature that vowel duration has been attributed as a perceptual cue for voicing to varying degrees. Other factors have also been extensively evaluated, and considered to be varyingly influential alongside vowel duration. In order to gain a fuller picture of this research, and elaborate on the importance of lexical representation and word frequency, let us now turn to the additional effect of the lexicon.

2.4 The effect of the lexicon on speech perception

Ganong (1980) investigated the notion of categorical perception in relation to the lexicon. Regarding the role that lexical status has on perception, Ganong (1980) references earlier work carried out by Rubin et al. (1976) which had already discovered that when a phoneme appears

in a word it is recognised more quickly than if it appears in a nonword. These faster reaction times provide evidence that there is an interaction between auditory information and lexical knowledge. With respect to the above concept, Ganong (1980) discusses two potential models, the categorical model and the criterion shift model. The categorical model describes a process in which lexical status does not affect processing until after phonetic categorisation. On the other hand, the criterion shift model encompasses the idea that lexical status may affect the way in which acoustic information is interpreted before phonetic categorisation, thereby producing a shift in the location of the phoneme boundary. He gives the example*, 'a subject whose phoneme boundary for a [da-ta] continuum was at 35msec VOT might require 40msec VOT to hear a [t] in the environment -ash (because dash is a word and tash is not)'* (Ganong, 1980:113). This model assumes that for acoustically ambiguous stimuli that lie near the phoneme boundary*, 'a change in criterion would produce large effects on categorisation'* (Ganong, 1980: 113). Here it is acknowledged that:

> 'Such a change could shift a stimulus from the ambiguous region to the word region or from the part of the nonword phonetic category near the boundary into the ambiguous region.'
>
> Ganong (1980: 113)

Despite focusing on varying VOT as opposed to vowel length, and word-initial as opposed to word-final voicing, the findings from Ganong (1980) proved significant in developing the hypotheses encompassed in this thesis. Specifically, whether the desire to create a word versus a nonword in speech perception may override acoustic information when processing and categorising individual phonemes.

Ganong (1980) conducted two experiments. The first incorporated seven pairs of synthesised continua, for example [dæʃ]~ *[tæʃ] *dash~*tash* and *[dɑːsk]~[tɑːsk] **dask~task*. Three other alveolar pairs based on [dʌst]~[tʌft] *dust~tuft*, [dɜːt]~[tɜːrf] *dirt~turf*, and [dəʊs]~[təʊst] *dose~toast* were also included, in addition to three velar continua which were synthesised based on [gɪft]~[kɪs] *gift~kiss*, [giːs]~[kiːp] *geese~keep*, and [gʌʃ]~[kʌsp] *gush~cusp.* Each continuum contained seven varying VOTs of 15ms, 25ms, 30ms, 35ms, 40ms, 45ms, and 55ms, respectively. Participants were presented with separate experimental blocks for stimuli containing alveolar and velar stimuli. They were required to listen to the stimuli and write down which sound they perceived hearing word-initially. In a second condition, the end points of the stimuli were presented to participants, who were then required to spell out the words and nonwords that they perceived hearing. For the second experiment, twenty-eight stimuli pairs were incorporated; four labial pairs, four alveolar pairs, and six velar pairs. Examples include [bæʃ]~[pæst], *bash~past*, [dɑːk]~[tɑːrp] *dark~tarp*, and [gʌlp]~[kʌlp] *gulp~culp*, respectively. Six stimuli were produced per continuum, one with the shortest possible VOT as determined by the token-base, four designed to span the phoneme boundary in 5ms steps, and the final one containing the longest possible VOT as determined by the token-base. Again, participants were asked to identify the word-initial consonant in the first instance, and identify the stimuli as a word or nonword alongside spelling it out in the second instance. In summary, both experiments identified a lexical effect at the phoneme boundary. This effect was greater for acoustically ambiguous stimuli, those found near the phoneme boundary, than those which were acoustically unambiguous and found at the end-points. The findings from Ganong's (1980) paper suggest that lexical information precedes acoustic information within ambiguous contexts. This shift in perception in the direction of a word as opposed to a nonword became known as the 'Ganong Effect'. These findings support a top-down model of speech

perception. However, it must also be acknowledged that these findings do not rule out the possibility that some acoustic information may be available before lexical status is considered.

Shortly after Ganong (1980) had produced his research, Fox (1984) attested to the importance of lexical status on the categorisation of phonemes. Despite rejecting the top-down model of speech perception, Fox's (1984) results support Ganong's (1980) theory that lexical status does appear to influence perception based on the place of articulation of any given sound, along with perceptual cues from the VOT continuum. Fox (1984) focused on place-of-articulation continua. Sets of [bVC]- [dVC] stimuli were created along a continuum in such a way that the end points of the tokens formed word/word, word/nonword, nonword/word, and nonword/nonword combinations. The four syllables ended with [æ], [æb], [æd], or [æg]. Participants were required to identify the initial syllables as a [b] or a [d]. The results demonstrated that ambiguous tokens were more often perceived as words than nonwords, thereby supporting the role of lexical bias. The second experiment employed different participants and encompassed a reaction time procedure using the same stimuli. The findings suggest that lexical effects are more profound when there is a response latency, demonstrating that *'the lexical effect represents a perceptual process that is separate from and follows phonetic categorisation'* (Fox, 1984:526). The third experiment aimed to separate both the lexical status and phonetic context variables. A further four pairs of continua were synthesised on a seven-step scale with [b]~[d] word-initially, ending in [ʌ], [ʌt], [ʌʧ] and [ʌz] to form the word/word, word/nonword, nonword/word, and nonword/nonword combinations. The procedure took on the same format as the first experiment. Indeed, it was found that lexical and phonetic levels of processing appeared to separately affect the identification of the word-initial stop. Fox (1984) therefore summarised the experiment by stating that when a stimuli is acoustically ambiguous, it is more likely to be affected by lexical status and that '*the degree to*

which lexical status affects a subject's perception varies as a function of the latency between stimulus presentation and the subject's key response' (Fox, 1984: 537). Here, Fox (1984) suggests that there is evidence that phonetic categorisation can take place in the absence of knowledge from higher levels of linguistic processing, such as lexical status. Fox therefore rejects the top-down model, as attested to by Klatt (1979, 1980), Morton (1969, 1979), and Warren (1976), among others. Fox (1984) concludes that *'it is premature to claim that the phonetic categorisation process is strictly autonomous until we can better understand the precise mechanisms underlying the lexical biasing effect'*. This, along with the earlier example from Burton et al. (1989) provide a counterargument to Ganong (1980).

When evaluating the role of the lexicon in speech perception, word-frequency is widely attested to as an important consideration. Here, Forster (1976) references search models, a suggestion of the ways in which the lexicon retrieves particular tokens. Search models propose *'a frequency-ordered search of the lexicon such that high frequency words are considered before low frequency words'*. On the other hand, activation models *'code frequency information in terms of differential levels of activation; high frequency words have higher base activation levels than do low frequency words'* (Connine et al., 1993:81). Both models assume frequency information is obtained at the level of lexical access. This means that frequency effects occur when initially accessing information about a word. In relation to this, Connine et al. (1993) conducted two experiments aiming to decipher the effects of word frequency in a phoneme identification task. Voicing continua were constructed such that one end of the continua consisted of a high frequency word, and the other endpoint was a low frequency word, for example, [bɛst]~[pɛst] *best~pest*. The first experiment manipulated intrinsic frequency effects, and pairs of words which differed in word-initial voicing were used. Of these pairs, half consisted of a low frequency voiced word and a high frequency voiceless word, and the

other half consisted of a high frequency voiced word and a low frequency voiceless word. For each of the voiced words, portions of periodic energy were omitted from the onset of the burst (/d/, /g/, or /b/). These portions were replaced with acoustic segments taken from the corresponding voiceless consonant, beginning at the end of the closure; /t/, /k/, or /p/. Participants were required to identify the word-initial phoneme that they perceived hearing. The findings showed that participants were more likely to label an ambiguous token such that a high frequency word was formed. The second experiment measured extrinsic word frequency effects. A mixed list was incorporated which contained the same stimuli as the first experiment, a second list was devised which presented ambiguous stimuli with only high frequency unambiguous stimuli, and a third list with only low frequency unambiguous stimuli. They found that *'a high frequency list bias produced an exaggerated influence of frequency; a low-frequency list bias showed a reverse frequency effect'* Connine et al. (1993: 81). Though this thesis does not specifically test for the effect of word frequency in perception, it has been a consideration in the formation of the experiments which address the role of the lexicon; Experiments 2 and 4 specifically. Therefore, CELEX; a database devised at Goethe University, Frankfurt, was used to obtain word frequency values for all relevant stimuli, and this information was subsequently used in the analysis to make appropriate inferences.

In summary, the precise nature of the interaction between lexical and acoustic information is disputed within the literature. Differing approaches to this research support either a top-down model in which lexical information supersedes phonetic categorisation, or a bottom-up model whereby phonetic categorisation must occur before the lexicon can be accessed. However, a strict model of either of these approaches is difficult to substantiate, and the interaction between these higher and lower levels is likely to be more nuanced. As with Fox (1984), the results of this thesis do not support a strict top-down model as the acoustic information from the

preceding vowel duration consistently influences the categorisation of the word-final phoneme, as opposed to a clear shift being prevalent in the direction of a known lexical item. Interestingly, the lexical effects which are evident in the findings from this thesis do appear to be constrained by the frequency of the lexical items being tested, with higher frequency words demonstrating the predicted outcomes more-so than lower frequency words. This complements findings from research such as Connine et al. (1993), and demonstrates the complexity of the interaction between lexical and phonetic information in speech perception.

2.5 Speech perception in non-native languages

As proposed by Diehl et al. (2004):

> 'In order to communicate proficiently, a listener must [be able to] discriminate acoustic variance in the speech signal that is linguistically relevant and…generalise across variance that is irrelevant [to their specific language].'
>
> Diehl et al. (2004:164)

Very early on in life, before the age of one, humans are able to distinguish and discriminate between sounds that are used contrastively in their native language (Eilers, 1977; Eimas, 1974; Eimas et al., 1971; Miller and Eimas, 1983). Chomsky (1965) first introduced the concept of a *language acquisition device* that is unique to humans and enables them to innately acquire language. However, deciphering the extent to which non-native speakers are able to acquire non-native cues in a learned manner proves an interesting area of research.

In relation to this, Broersma (2010) tested Dutch and English speakers to determine the effect of vowel duration on the categorisation of a word-final fricative as either voiced or voiceless. Recall that, as with German, Dutch only permits voiceless word-final obstruents in

the surface form. However, Dutch does encompass word-initial and intervocalic voicing contrasts, and it was therefore thought possible that Dutch listeners may be able to apply these tendencies to the word-final position when operating in English. As such, Broersma (2010) varied the vowel duration across the stimuli. This study employed nonwords across two continua ranging from /v/ to /f/ and from /z/ to /s/. There were nine steps between the natural voiced and voiceless end points. Broersma (2010) writes:

> 'To this end, the final /f/ and /s/ were shortened to match the durations of the final /v/ and /z/. They were shortened by 56-127ms (31%) for the /f/ and by 80-187ms (30%) for the /s/ by removing a portion from the centre of the fricative. Next, for the pairs of phonemes thus obtained, the amplitudes of the waveforms were added in varying proportions of 11 equally spaced steps.'
>
> Broersma (2010:1638)

The CV portion of the nonwords */ku:v/, */ku:f/, */fu:z/, and */fu:s/ were extracted, thus resulting in a two truncated sequences (/ku:/ and /fu:/). For each of the two sequences, two variants existed, one which originally ended in the voiced final fricative and contained the phonetically longer vowel, and one which originally ended in the voiceless final fricative and contained the phonetically shorter vowel. Broersma (2010) summarises this methodology by stating that:

> 'For each continuum, the two end points and the nine intermediate steps were spliced onto the end of the appropriate carriers. Thus, for the /v-f/ continuum, all steps were spliced onto both /ku:/ carriers, and for the /z-s/ continuum, all steps were spliced onto both /fu:/ carriers.'
>
> Broersma (2010:1638)

Participants consisted of Dutch speakers and English speakers, and they were tested using a goodness rating and a phonetic categorisation experiment. The Dutch speakers had received an average of seven years of English tuition, and the English speakers had no knowledge of Dutch. For the first experiment, participants were asked to indicate using four buttons labelled one to four, one being *good* and four being *poor*, how *good* the word-final sounds were in relation to a target word-final sound. For example, participants would be asked to what extent a fricative *sounded like a good [z]*. For the second experiment, participants were presented with the stimuli and asked to indicate whether they had heard a [v] or an [f], or a [z] or an [s] word-finally. Participants made their decision by pressing one of two response buttons which were labelled for each of the possible sound selections. Broersma (2010:1641) writes *'each participant heard both the phonetically long and the phonetically short vowel within the same experiment and within the same block…each block contained 20 repetitions of all combinations of 11 fricatives and two carriers, yielding 440 trials per block.'*

The findings of this study found that the Dutch listeners consistently used vowel duration as a cue for word-final voicing to a lesser extent than the English listeners did. However, Dutch listeners did use vowel duration to some extent, and the differences between the two groups were small. Broersma (2010:1643) concludes *'the findings…seem to reflect a robust difference between Dutch and English listeners' use of vowel duration that is not limited to a single phoneme contrast or to a particular experimental design or paradigm*. Crucially, this seems to suggest that phonological cues from L1 languages do not automatically translate into differing L2 contrasts. These results were supported by the overall findings from this thesis in which native English speakers demonstrated the most sensitivity overall to vowel duration as a cue for word-final voicing when compared with the non-native speakers.

A previous paper by Broersma (2005) also focused on vowel duration as a cue for word-final voicing in Dutch. This study contained two experiments. The first considered the voicing contrasts between word-final /z/~/s/, /v/~/f/, /b/~/p/ and /d/~/t/ on the basis of vowel duration, and found that Dutch listeners were able to accurately categorise these phonemes word-finally in the case of nonwords. The second experiment encompassed vowel durations which were uninformative and contained mismatches between the vowel duration and other information, for example a voiceless fricative preceded by a long vowel. As such, Dutch listeners outperformed native English speakers who were possibly misguided by the mismatched and uninformative vowel durations. Therefore, although there was no evidence to suggest that the Dutch speakers were using vowel duration, they did show evidence of attaining native-like accuracy. Broersma (2005) conclude that the native production of non-native contrasts in unfamiliar positions *'may hardly ever be attained'*. The results from this study proved significant for this thesis in informing the analysis for Experiment 4. Here, the German speakers occasionally outperformed the native English speakers in using vowel duration as a guide for word-final voicing. Based upon the results from Broersma (2005), this finding was attributed to potential processing difficulties for the native English speakers caused by the nature of the computationally-generated stimuli used in Experiment 4. It was thought that this issue did not impact the German speakers to as greater an extent as they were not as perceptually sensitive to these L2 cues, and were therefore not misguided by the acoustic input.

Crowther and Mann (1992) produced a study incorporating native American English speakers, native speakers of Japanese, and Mandarin Chinese speakers. Unlike English, there are no word-final stop consonants in Japanese or Mandarin Chinese, and as such this study encompassed three experiments incorporating speakers of these languages who were learning English as a second language. The phonology of Japanese employs an alternation between

phonemically long and short vowels. Crowther and Mann (1992:711) write '*the results of some recent work [regarding this phonemic difference] suggest that native experience with long/short vowels might be relevant to the use of vocalic duration as a cue to final consonant voicing (...Fledge, 1984; Fledge and Hillenbrand, 1986; Stevens et al., 1986)'*. Both vowel duration and F1 offset frequency were investigated. The first experiment *'measured the F1 offset frequency and vocalic duration in productions of 'pod' and 'pot''*. The second experiment *'assessed identification of natural tokens of 'pod' and 'pot' with and without closure segment bursts'*. Finally, the third experiment *'assessed categorisation of synthetic 'pod'[~]'pot' stimuli that systematically manipulated vocalic duration and F1 offset frequency'* (Crowther and Mann, 1992:711). Regarding vowel duration, native English speakers showed the greatest sensitivity, followed by Japanese, and then Mandarin Chinese speakers. These results are likely to be due to an unfamiliarity of the use of stop consonants word-finally in the L2 speaker's native phonologies. Japanese speakers were most likely more accurate than Mandarin Chinese speakers given their exposure to alternations in vowel duration in their native language. Once again, this paper demonstrates the role of underlying phonology in speech perception, and the ways in which this can constrain the mapping of L1 phonetics and phonology onto L2 phonetics and phonology.

Kennard and Lahiri (2019), argue that native phonology also constrains the phonetic perception of loan words. They posit that *'phonological features which contribute to allophonic alternations may become contrastive, but no new phonological feature should be introduced merely via loans'* (Kennard and Lahiri, 2019:1). The findings of this thesis support this notion, as native German speakers demonstrate an asymmetry in their perception of word-final stops versus affricates in English. Unlike stops, underlyingly voiced affricates are not part of the phonology of German, unless borrowed into the language as a loan word. Therefore,

when operating in L2 English, German speakers were constrained by their native phonology and demonstrated difficulty in perceiving the voiced nature of a word-final affricate.

It is therefore clear from the literature that there is some interaction between the perceptual cues that exist within native and non-native languages, as constrained by both the phonetics and the phonology. Native perceptual cues and the structure of native languages appear to have the potential to inhibit our sensitivity to non-native cues. Given that preceding vowel duration is such a salient cue for word-final voicing in English but not in German, the research within this thesis that focuses on deciphering the extent to which German speakers are sensitive to this cue when operating in L2 English will provide an interesting and relevant contribution to this body of research.

2.6 Summary

To conclude, the literature reviewed in this chapter reinforced the decision to investigate the effect of both vowel duration and the lexicon on the perception of word-final voicing in English. In addition, the literature addressing native and non-native speech in this area of research has introduced some interesting questions. Specifically, the extent to which L1 phonology may interact with L2 phonology when operating in a second language. The studies discussed in this chapter have also influenced the chosen methodology and the techniques utilised in carrying out the investigations for this thesis. It is hoped that the results from this thesis will contribute to this wider body of knowledge, enriching our understanding of some of the primary factors which contribute to language perception, and the ways in which this perception can be modelled.

CHAPTER THREE

Vowel duration in the context of *voiced* and *voiceless* obstruents in English: production studies for native and non-native speakers

3.1 Introduction

As demonstrated in the prior review of existing literature (cf. Chapter Two), the relationship between vowel duration and voicing in English has been well-documented; when preceding a voiced word-final obstruent, a vowel will be systematically longer in duration than when it precedes a voiceless word-final obstruent. This difference is most notable when two words ending in a minimal pair are being compared; [məʊd]~[məʊt] *mode~moat* for example. The central focus of this thesis is on the perceptual relevance of this phonological tendency. Specifically, to what extent is the perception of word-final voicing directly related to the duration of the preceding vowel? A second related question is the relevance of vowel length for second language learners of English. In terms of their competence in English, does the duration of the vowel play a significant role in determining the voicing of the word-final consonant? However, to investigate these perceptual consequences we must first examine the extent to which native and non-native English speakers demonstrate this phonological tendency when producing speech.

3.2 Experiment 1a: native speaker production of vowel length before [d]~[t]

Experiment 1a (hereafter, Exp1a) focused on three native speakers of Southern British English. This initial pilot study served the purpose of establishing first-hand that native speakers of English do indeed lengthen their vowel duration in relation to word-final voicing, and determining to what extent this appears to be the case. This production study was primarily

designed to form the basis for a second production study, involving both native and non-native English speakers, and to inform latter perception experiments.

3.2.1 Research questions

Exp1a is interested in determining whether all speakers lengthen the vowel in a CVC structure prior to a voiced word-final consonant, relative to a voiceless word-final consonant. Based on the findings of previous research, it was expected that native speakers will consistently demonstrate this contrast in vowel duration.

3.2.2 Methodology

Speakers: three male speakers between the ages of twenty-seven and sixty were recorded. They had no known hearing or language disorders. Male speakers were chosen for these recordings given that, on average, they tend to have a longer vocal tract than female speakers. This results in acoustically lower-pitched, more resonant speech that proves more suited to speech analysis. To minimise dialect variability, all three speakers had a Southern British English dialect. This is considered to be the standard dialect of English. The speakers were not reimbursed for their time.

Materials: six real words in English were incorporated into Exp1a. All of the words began with the same word-initial consonant, [m], and ended in an equal number of the voiced and voiceless stops, [d] and [t]. The vowel nuclei within the words also needed to be held consistent, with an equal number of words containing an [æ] (low), [eɪ] (mid), or [əʊ] (mid) vowel. The consideration in choosing these three vowels centred around quality and positioning, covering as much of the vowel space as possible whilst still being able to form real word CVC minimal

pairs. All word-tokens were monosyllabic. Word frequency was not controlled for throughout Experiment 1, as this thesis is only concerned with the potential of lexical bias caused by word frequency in the latter perceptual experiments and not regarding speech production. Ensuring the consistency of linguistic form was therefore prioritised over matching word frequencies. The word-tokens are listed in Table 1 (below).

Table 1: The list of word-tokens used for Exp1a[1]

English minimal pairs /d/~/t/	**IPA**	
cad~cat	*/kæd/*	*/kæt/*
mad~mat	/mæd/	/mæt/
mode~moat	/məʊd/	/məʊt/
made~mate	/meɪd/	/meɪt/

Recording procedure: the speakers were recorded in the Language and Brain Laboratory at the University of Oxford. Each speaker was recorded individually in a soundproof booth to reduce the likelihood of background noise being present on the recordings. The recordings were made using a Rode NT-USB microphone and a Macbook computer (OS X El Capitan version 10.11.6), running both *Audacity* (version 2.1.2) and *Praat* (version 6.0.21).[2] *Audacity* was used for capturing the voice recordings, using a mono 44.1khz sample rate. *Praat* was then utilised for the subsequent speech analysis.

Speakers were recorded sitting down, with the microphone positioned a comfortable distance of around 20cm away from their mouth. The distance between the microphone and the speaker's mouth was held as consistent as possible between speakers. The microphone was

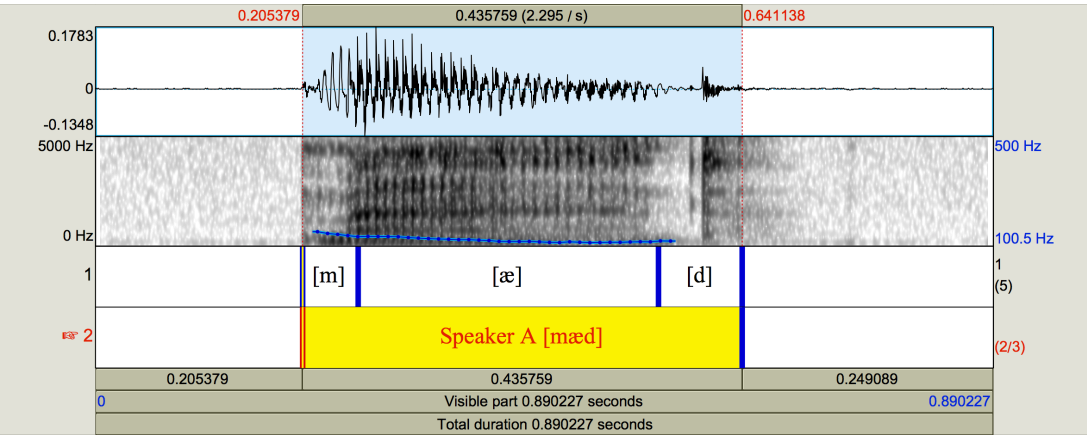
[1] Words found in italics are fillers
[2] All versions remain consistent throughout this thesis

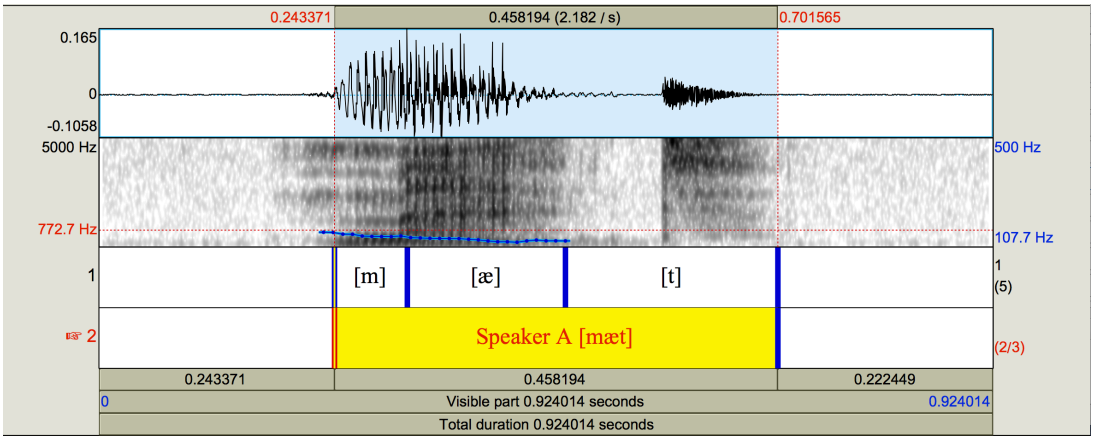
also placed on a foam mat to reduce table vibrations, which may have led to interference on the recordings.

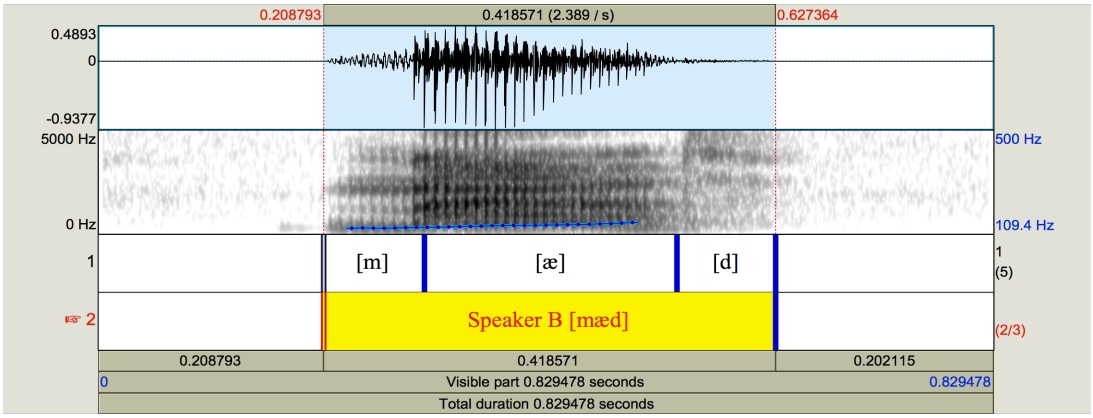
The speakers were asked to read out the words listed in Table 1 (above) in the following order: [kæt] *cat*, [mæt] *mat*, [mæd] *mad*, [meɪt] *mate*, [meɪd] *made*, [məʊt] *moat*, [məʊd] *mode,* [kæd] *cad,* working their way down the list from top to bottom. They repeated this process five times, thereby generating five instances of each word-token. They were instructed to read out the words in the list at a fixed speed and to keep their intonation as consistent as possible. To avoid list-intonation, or falling-pitch, the speakers were requested to record the words in this way, rather than simply repeating each word five times. One of the well-attested to limitations of production studies is that it is very difficult to obtain natural speech in an artificial environment. When speakers are being recorded in an unfamiliar, controlled environment this is likely to affect the way that they produce speech. To overcome this issue, the environment for Exp1a was held as consistent as possible between the speakers. Unfortunately, it is not possible to obtain the data necessary for this study in an organic way, and therefore the benefits of this methodology outweighed these limitations. Several precautions were taken to regulate the speech being recorded, for example the first and last words recorded were fillers, and gave the speaker the opportunity to normalise their intonation and volume at the beginning and end of the recording. The speakers were not made aware of what this study was investigating, or of which words were going to be included in the analysis. This participant naivety was necessary to avoid any effects caused by the observer's bias; the idea that participants will behave differently, either subconsciously or consciously, in an experimental setting if they know the specific aims of an investigation.

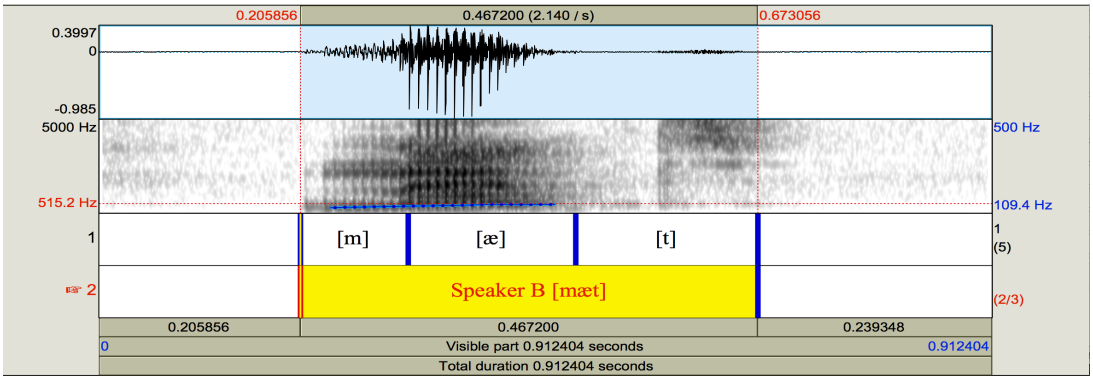
3.2.3 Measurements

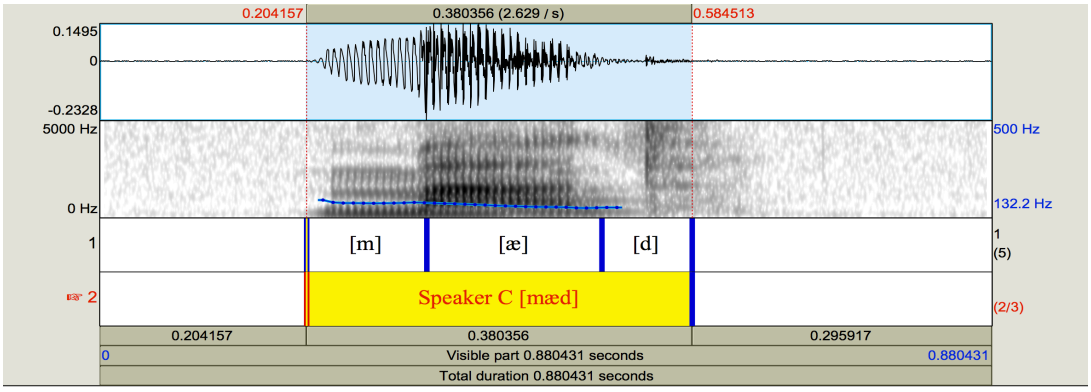
A total of ninety recordings were made; five repetitions of the six words found in Table 1 (above) spoken by each of the three speakers. These recordings were then analysed using *Praat.* The individual word-tokens were saved as separate *.wav* files and a *TextGrid* was created that corresponded to each of these files. Each individual phoneme was measured in the CVC sequence; the onset consonant, vowel, and coda consonant. Figure 1 (below) illustrates this process. All boundaries were placed at the nearest zero-crossing boundary, and measurements were taken on this basis. The boundaries were placed manually in positions where the spectrogram demonstrated both a visual and auditory change in articulation, and the positioning of these boundaries were kept as consistent as possible throughout the mark-up process. A *Pratt* script was then utilised to extract the durational measurements from each of these *TextGrids*. This created a *.txt* file for each of the three speakers, detailing the individual *.wav* files and each of the durational measurements in milliseconds for the three phonemes within these files.




0.243371
0.458194 (2.182 / s)
0.701565
0.165
0
-0.1058
5000 Hz
500 Hz
772.7 Hz
107.7 Hz
1
[m]
[æ]
[t]
1
(5)
☞ 2
Speaker A [mæt]
(2/3)
0.243371
0.458194
0.222449
0
Visible part 0.924014 seconds
0.924014
Total duration 0.924014 seconds


0.208793
0.418571 (2.389 / s)
0.627364
0.4893
0
-0.9377
5000 Hz
500 Hz
0 Hz
109.4 Hz
1
[m]
[æ]
[d]
1
(5)
☞ 2
Speaker B [mæd]
(2/3)
0.208793
0.418571
0.202115
0
Visible part 0.829478 seconds
0.829478
Total duration 0.829478 seconds


0.205856
0.467200 (2.140 / s)
0.673056
0.3997
0
-0.985
5000 Hz
500 Hz
515.2 Hz
109.4 Hz
1
[m]
[æ]
[t]
1
(5)
☞ 2
Speaker B [mæt]
(2/3)
0.205856
0.467200
0.239348
0
Visible part 0.912404 seconds
0.912404
Total duration 0.912404 seconds


0.204157
0.380356 (2.629 / s)
0.584513
0.1495
0
-0.2328
5000 Hz
500 Hz
0 Hz
132.2 Hz
1
[m]
[æ]
[d]
1
(5)
☞ 2
Speaker C [mæd]
(2/3)
0.204157
0.380356
0.295917
0
Visible part 0.880431 seconds
0.880431
Total duration 0.880431 seconds

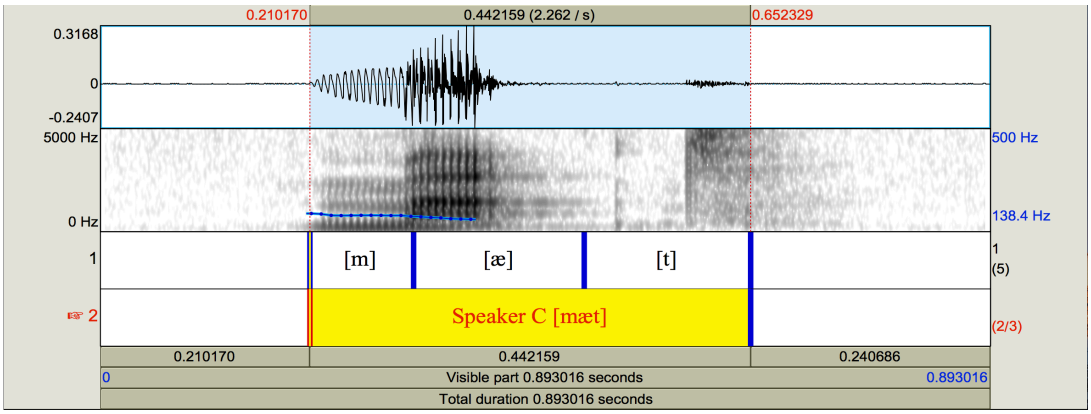


Figure 1: A demonstration of the use of *Praat* to segment the word pair [mæd]~[mæt] *mad~mat* across the three speakers for Exp1a

3.2.4 Analysis

Initially, the durational measurements were transferred into a *Microsoft Excel* spreadsheet for organisation, and then into *JMP Pro 14 (SAS)* (version 14.2) (hereafter, *JMP (SAS)*)[3]. *JMP (SAS)* is a statistical analysis program, developed by the SAS Institute, which allows the user to visualise data and analyse trends. For all three speakers, mean durations were calculated for each of the phonemes; the initial consonant, the vowel, and the word-final consonant. The mean vowel durations were then extracted and organised into word-final voiced and voiceless contexts.

Turning now to the analysis, let us first consider the overall results, before reviewing the individual word-token pairs. These results are illustrated using Figures 2-6 (below).

[3] The use of this version remained consistent throughout this thesis

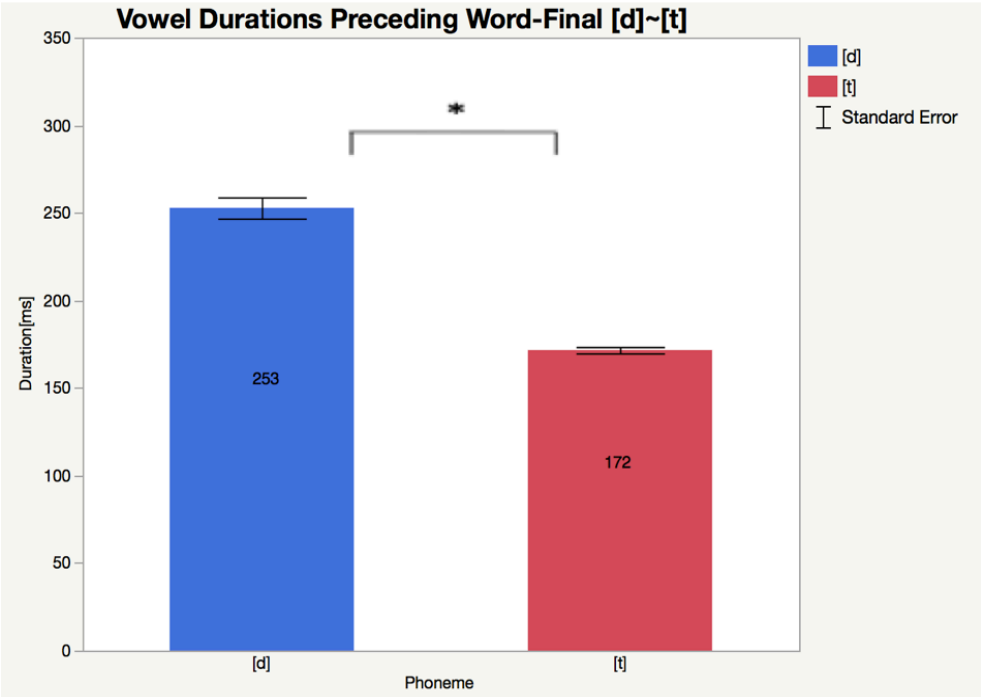


Figure 2: The difference in mean vowel duration preceding word-final [d]~[t] for Exp1a[4]

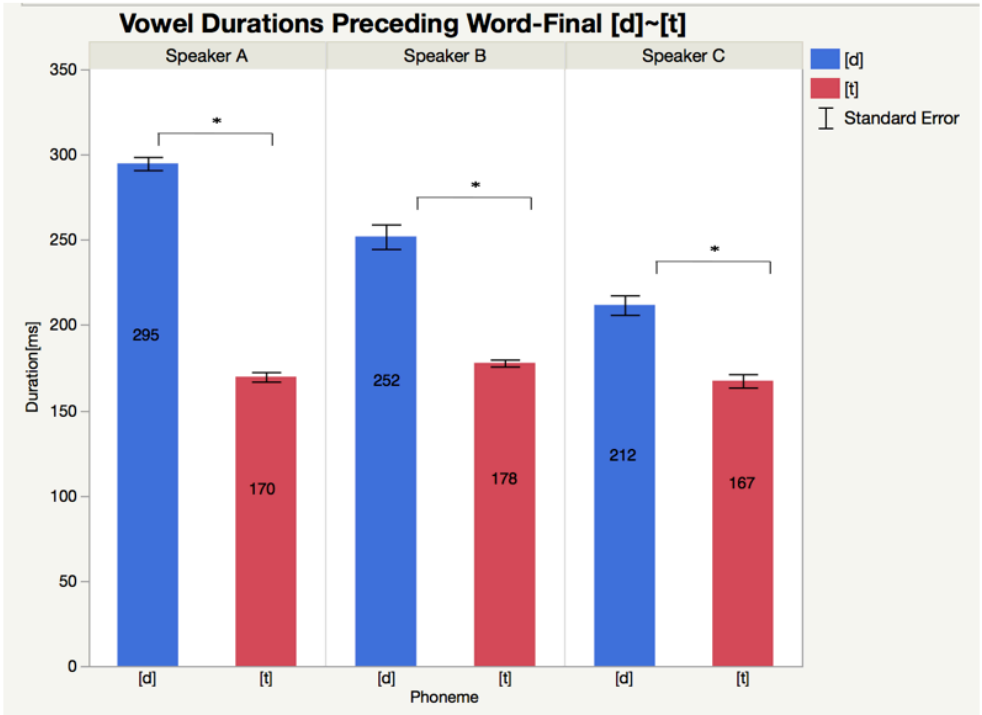


Figure 3: The difference in mean vowel duration preceding word-final [d]~[t] on a speaker-by-speaker basis for Exp1a

[4] SE is calculated as 1 standard error from the mean throughout Exp1a and Exp1b

A series of one-way ANOVAs (Analysis of Variance) were conducted to establish that the differences between the mean vowel durations for each speaker were significant. Running a one-way ANOVA was considered to be the most appropriate method to determine this as Exp1a examines one independent variable in relation to two levels; vowel duration in relation to voiced and voiceless word-final phonemes.

Across all items and speakers, there is a highly significant difference (F (1, 88) = 164.8134, $P < .0001$) in vowel duration, with longer vowel durations preceding the voiced [d] and shorter vowel durations preceding the voiceless [t] as expected. Indeed, Speaker A (F (1, 28) = 685.07, $P < .0001$), Speaker B (F (1, 28) = 98.67, $P < .0001$), and Speaker C (F (1, 28) = 40.93, $P < .0001$) all demonstrate this highly significant difference on an individual speaker basis.

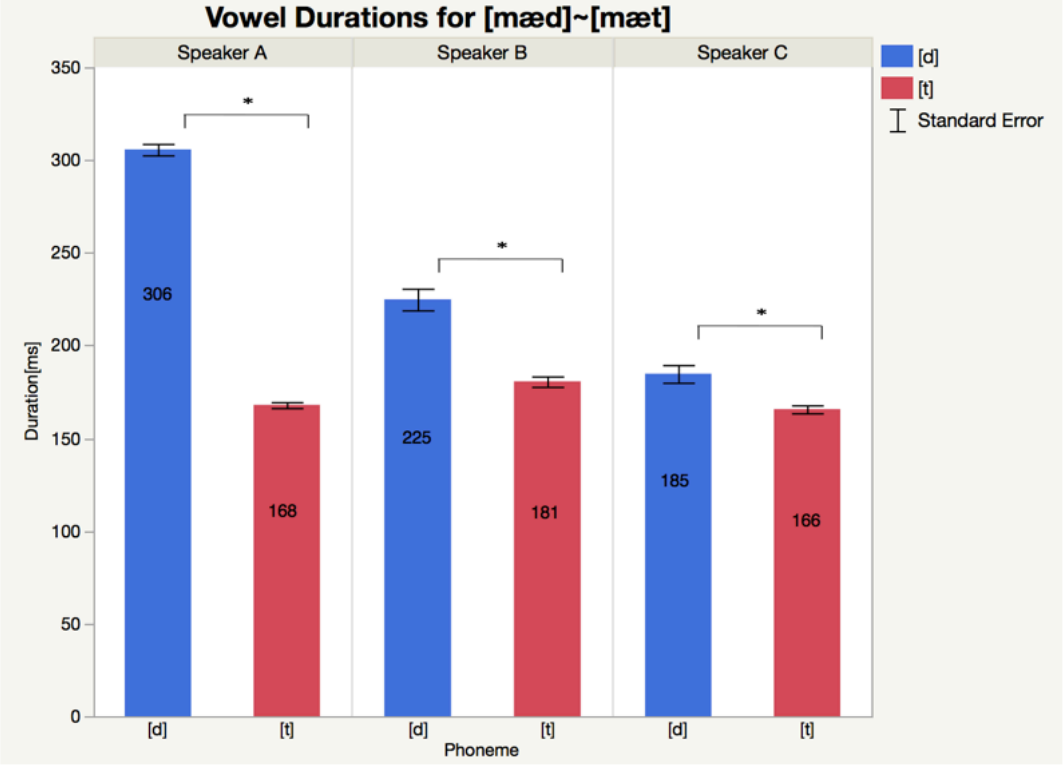


Figure 4: The difference in mean vowel duration for [mæd]~[mæt] *mad~mat* on a speaker-by-speaker basis for Exp1a

Turning now to the individual word pairs, in the case of [mæd]~[mæt] *mad~mat*, all speakers demonstrate significant differences in vowel duration prior to the voiced and voiceless word-final endings; Speaker A (F (1, 8) = 1559.45, P < .0001), Speaker B (F (1, 8) = 46.54, P < .0001), and Speaker C (F (1, 8) = 13.37, P = .0064). Once again, longer durations precede the voiced [d] and shorter durations precede the voiceless [t]. Speaker C shows a lower degree of differentiation in comparison to Speakers A and B who both exhibit a highly significant statistical difference in their responses.

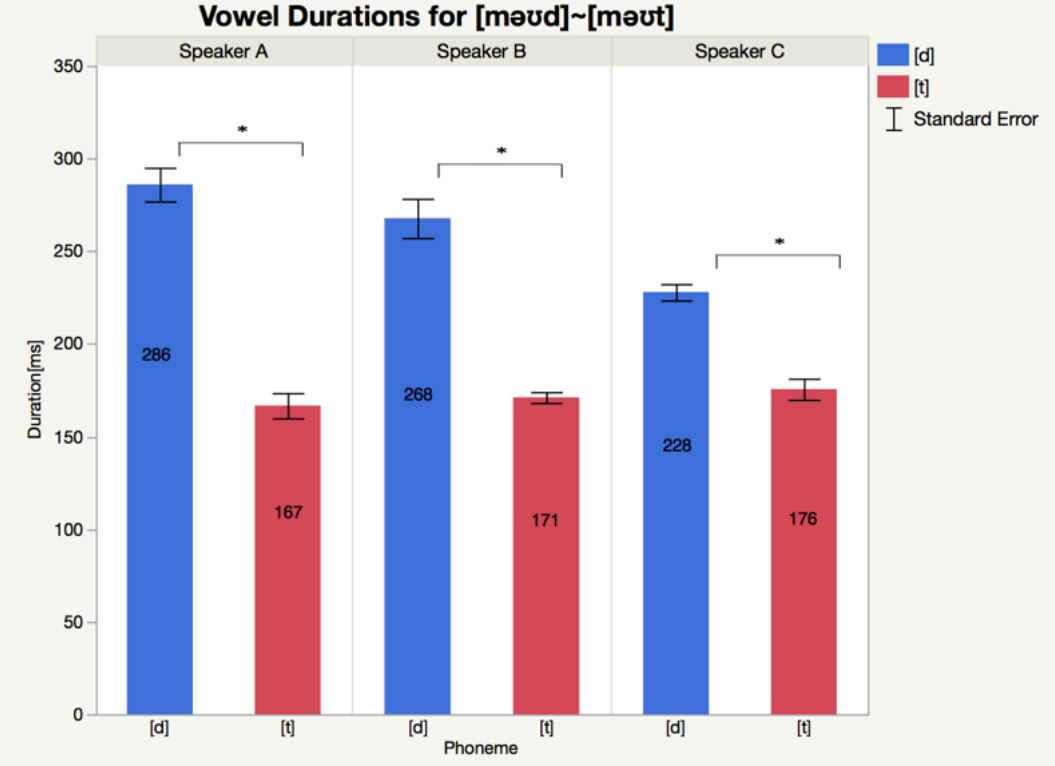


Figure 5: The difference in mean vowel duration for [məʊd]~[məʊt] *mode~moat* on a speaker-by-speaker basis for Exp1a

When producing [məʊd]~[məʊt] *mode~moat*, Speaker A (F (1, 8) = 110.59, P < .0001), Speaker B (F (1, 8) = 77.57, P < .0001), and Speaker C (F (1, 8) = 52.79, P < .0001) all establish highly significant differences in vowel duration preceding word-final [d] and [t]. Again, longer vowel durations precede voiced instances of word-final voicing.

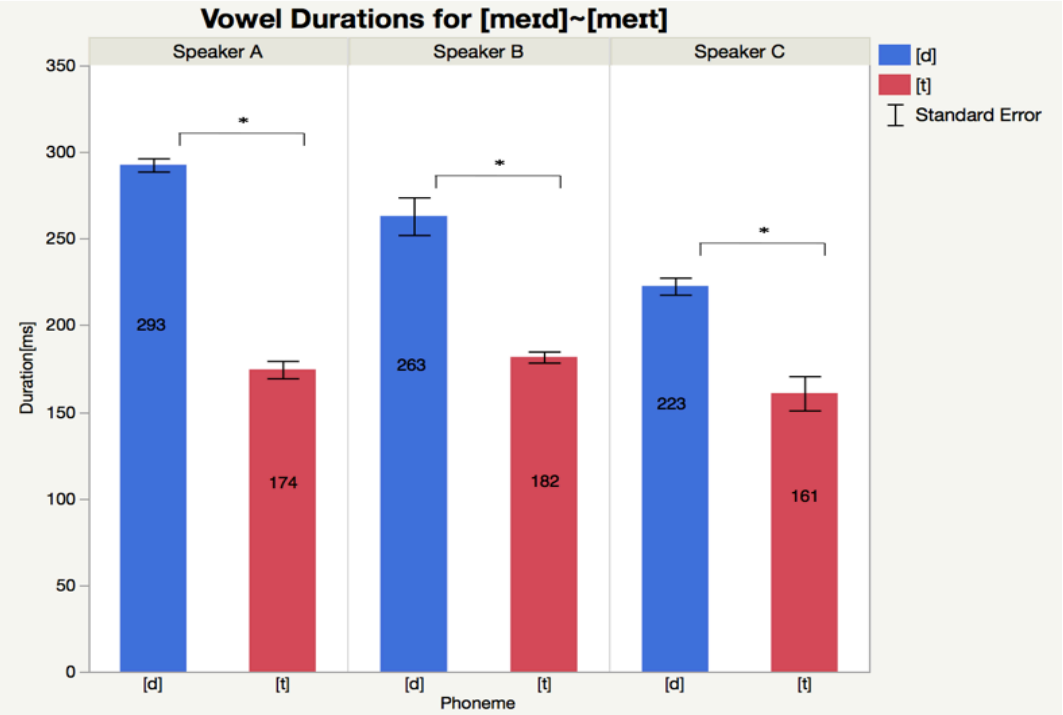


Figure 6: The difference in mean vowel duration for [meɪd]~[meɪt] *made*~mate on a speaker-by-speaker basis for Exp1a

Finally, for [meɪd]~[meɪt] *made~mate*, Speaker A (F (1, 8) = 350.12, P < .0001), Speaker B (F (1, 8) = 51.77, P < .0001), and Speaker C (F (1, 8) = 31.41, P = .0005) produce significant differences between the longer vowels preceding word-final [d] and the shorter vowels preceding word-final [t]. As with [mæd]~[mæt] *mad~mat*, Speaker C has produced a smaller difference overall in comparison with Speaker A and Speaker B who both demonstrate highly significant differences. A summary of the differences in duration for each of the above categories can be found in Table 2 (below).

Table 2: A summary of the difference in mean vowel duration for Exp1a

Speaker	**[d]~[t]**	**[mæd]~[mæt]** ***mad~mat***	**[məʊd]~[məʊt]** ***mode~moat***	**[meɪd]~[meɪt]** ***made~mate***
Speaker A	125ms	138ms	119ms	119ms
Speaker B	74ms	44ms	97ms	81ms
Speaker C	45ms	19ms	52ms	62ms

The primary finding from Exp1a is that all speakers do indeed produce comparatively longer vowels when preceding the word-final voiced [d] as opposed to the voiceless counterpart [t], with an average difference of 81ms across all speakers and items. This tendency is consistent across all results, demonstrated in the case of both monomoraic and bimoriac vowels, and this supports the prediction posited earlier in this chapter.

However, the extent to which this tendency proves true varies between speakers. Given the nature of speech, this is not surprising. Table 2 (above) illustrates that stimuli containing the [əʊ] vowel show the greatest average durational difference (89ms), followed by [eɪ] (87ms), and finally [æ] (67ms). For stimuli containing the vowel [æ], the speaker with the greatest variation in vowel length is Speaker A (138ms), followed by Speaker B (44ms), and finally Speaker C with the smallest durational difference (19ms). For stimuli containing the vowel [əʊ], Speaker A once again exhibits the greatest difference in vowel duration (119ms), followed by Speaker B (97ms), and finally Speaker C (52ms). Finally, for stimuli containing the vowel [eɪ], this ordering remains the same with Speaker A producing the largest durational difference (119ms), followed by Speaker B (81ms), and finally Speaker C (62ms).

It is important to ascertain whether these values differed because the three speakers did indeed lengthen and shorten the vowel duration to a greater or lesser extent, or whether these trends were actually due to speech rate. For instance, it could be that Speaker C simply speaks more quickly than Speakers A and B. As such, Table 3 (below) shows the average percentage of each token that consisted of the vowel.

Table 3: The average vocalic proportion of each token for Exp1a

Speaker A		
Token	**Percentage (%)**	**Difference (%)**
[mæd] *mad*	66	30
[mæt] *mat*	36	
[məʊd] *mode*	56	16
[məʊt] *moat*	40	
[meɪd] *made*	60	17
[meɪt] *mate*	43	
Speaker B		
Token	**Percentage (%)**	**Difference (%)**
[mæd] *mad*	53	15
[mæt] *mat*	38	
[məʊd] *mode*	54	22
[məʊt] *moat*	32	
[meɪd] *made*	54	17
[meɪt] *mate*	37	
Speaker C		
Token	**Percentage (%)**	**Difference (%)**
[mæd] *mad*	45	9
[mæt] *mat*	36	
[məʊd] *mode*	48	8
[məʊt] *moat*	40	
[meɪd] *made*	47	11
[meɪt] *mate*	36	

For each speaker, the relative difference between [d]~[t] is shown, and these values support the previous analysis. Speakers A and B did produce vowel proportions which were more differentiated on the basis on word-final voicing, whilst Speaker C adhered to this phonological tendency to a lesser extent. As such, these relative proportions do support the hypothesis that voiced word-final consonants are consistently preceded by a relatively longer vowel than voiceless word-final consonants; and this is production is happening regardless of speech rate. As illustrated by the results of the ANOVAs, all three speakers demonstrate highly significant differences across the two voicing conditions *(P* < .0001). The only exception to this was Speaker C who produced slightly less differentiated durations for [mæd]~[mæt] *mad~mat*

(P = .0064) and [meɪd]~[meɪt] *made~mate* (P = .0005). However, the *p*-values recorded here were still significant and therefore the overall findings hold true.

3.2.5 Discussion

The results from Exp1a support the predictions posited earlier in this chapter; that speakers will systematically lengthen their vowel duration prior to a voiced word-final consonant, relative to a voiceless word-final consonant. These results are therefore in support of the phonological tendency that, in English, voiced word-final consonants are preceded by comparatively longer vowels than their voiceless counterparts.

Though this is a small-scale pilot study, the results do indeed suggest that variations in vowel duration are a prominent feature of spoken English in relation to word-final voicing. Speaker C is an interesting case, as they consistently produced smaller durational differences than Speakers A and B. Therefore, despite evidence that Speaker C has conceptualised the difference between a /d/ and /t/ underlyingly, there is not a great deal of difference in their production of the vowel durations. However, the durational differences produced by Speaker C do remain significantly different, and this behaviour can therefore most likely be put down to inter-speaker variation.

3.2.6 Conclusion

Exp1a was interested in determining whether all speakers lengthen the vowel in a CVC structure prior to a voiced word-final consonant, relative to a voiceless word-final consonant. It is clear from the data that speakers do consistently illustrate a significant difference in the production of vowel duration preceding voiced and voiceless word-final consonants. These results therefore suggest that even though duration is not contrastive in English, in so far as the

meaning of the words containing this minimal pair do not change depending on the length of the vowel, speakers do exhibit a perceptual differentiation according to duration. These durational differences may therefore, in turn, prove to be a primary and necessary cue for the perception of word-final voicing in English.

3.2.7 Limitations of Experiment 1a

Exp1a was intended as a pilot study. It was conducted with the aim of establishing on a first-hand basis that, in English, vowel lengthening does indeed occur before a voiced word-final sound in the manner suggested by the literature. However, as this was a pilot study, the tokens themselves only contained the word-final stops [d] and [t]. For a small-scale study, this was sufficient. However, the latter perceptual experiments of this thesis incorporate stimuli ending in a range of word-final sounds; including fricatives and affricates, as well as stops. Therefore, increasing the range of word-tokens in a production study would allow for a more nuanced analysis to be conducted ahead of devising the stimuli for these latter experiments. Additionally, the formation of the word-list may be considered too simple. Rather than having one filler at the beginning and end of the recording, it became apparent that it would be preferable to have fillers distributed throughout the recording. This would mean that the pattern of the words being investigated would not be as obvious to the speakers being recorded. Equally, the incorporation of nonwords would be beneficial. The experiments in the latter part of the thesis incorporate stimuli formed from both real words and nonwords in order to investigate the influence of the lexicon in speech perception. Being able to analyse any discrepancies in the production of these two categories would therefore provide a complementary and useful insight into the role of the lexicon in speech production. Finally, only English speakers were incorporated into Exp1a. As detailed in Chapter One, this thesis

aims to addresses potential differences between the role of vowel duration as a cue for voicing in native and non-native English. Therefore, it would prove useful to have some first-hand foundation knowledge of the way in which speech production is affected for these two types of speakers. This can be achieved by recording L2 English speakers for which the word-final voicing contrast is neutralised in the surface form of their native language.

To address each of these limitations, a second production study was devised incorporating a wider range of stimuli, alongside including native English speakers and native German speakers who were second language learners of English. This second production study forms the basis for Experiment 4; a forced choice identification task investigating perceptual differences in native and non-native English speakers.

3.3 Experiment 1b: native and non-native speaker production of vowel length before stops and affricates

Experiment 1b (hereafter, Exp1b) consisted of a further production study incorporating both native and non-native speakers of English. As previously discussed, the word-final voicing alternation is neutralised in German and there is also no alternation in terms of vowel duration. As such, native German speakers are not exposed to vowel duration as a cue for word-final voicing. It would therefore follow that native speakers of German who have good English proficiency may not be as sensitive to vowel duration as a phonological cue for voicing as native English speakers, despite their proficiency. Two groups of German speakers with varying English proficiency were recorded for Exp1b. One group of speakers were recorded in Germany, and one group were recorded in the United Kingdom. As such, the speakers had experienced varying exposure to spoken English. This allowed for inferences to be made regarding the relationship between an increased exposure to a second language, and a more native-like production of fine-grained phonetic cues such as vowel duration.

3.3.1 Research questions

As with Exp1a, Exp1b is interested in deciphering whether all speakers lengthen the vowel in an English CVC structure prior to a voiced word-final consonant, relative to a voiceless word-final consonant. It is additionally interested in whether there is a difference in the extent of this lengthening in accordance with whether a speaker is a native or non-native English speaker, and if there is a notable difference in the extent of this lengthening based on the amount of exposure to spoken English the non-native speakers have received. Furthermore, Exp1b incorporates both real and nonwords. As such, Exp1b aims to determine whether there

is a difference in the production of vowel duration in relation to word-final voicing based on the effect of the lexicon.

It was expected that non-native English speakers would demonstrate a greater extent of lengthening in accordance with an increased exposure to spoken English. It was also expected that native English speakers would show the greatest degree of lengthening overall.

3.3.2 Methodology

Speakers: Three pairs of speakers were recruited for Exp1b. The first pair consisted of male native Southern British English speakers aged between eighteen and twenty-five. The second pair comprised of male native German speakers aged between eighteen and twenty-five, who identified as having 'good' English proficiency. These speakers were recorded in Frankfurt, Germany at Goethe University. It was required that these speakers had never lived in the United Kingdom or any other English-speaking country, and that they were not studying for an English-based degree. Additionally, neither of their parents or caregivers were native English speakers. The final pair consisted of male native German speakers aged between eighteen and twenty-five who identified as being 'fluent' in English. These speakers were recorded in London, UK. The speakers were required to be fluent in English and to have lived in the United Kingdom or another English-speaking country for a minimum of three years.

Male speakers were chosen for the same methodological reasons as those detailed in Exp1a. None of the speakers had any known hearing or language disorders. Speakers were compensated £5/€5 for their time.

Materials: the word-lists incorporated into the recording for Exp1b were selected to form two blocks; one ending in a minimal pair of stops, [d] and [t], and the other ending in a minimal

pair of affricates, [ʤ] and [ʧ]. Fricatives were not included as part of this study as Experiment 4, for which Exp1b forms the basis, incorporates stops and affricates only. Fricatives are included as part of Experiment 5, a lexical decision task, however the stimuli and format of Experiment 5 were not directly influenced by Exp1b.

The word pairs employed in Exp1b can be found in Table 4 (below)

Table 4: The list of word-tokens used for Exp1b

English minimal pairs /d/~/t/	**Pair**	**IPA**	
bad~bat	word /d/~word /t/	/bæd/	/bæt/
dad~*dat	word /d/~nonword /t/	/dæd/	/dæt/
*jad~*jat	nonword /d/~nonword /t/	/ʤæd/	/ʤæt/
*vad~vat	nonword /d/~word /t/	/væd/	/væt/
English minimal pairs /ʤ/- /ʧ/	**Pair**	**IPA**	
badge~batch	word /ʤ/~word /ʧ/	/bædʒ/	/bæʧ/
*radge~*ratch	nonword /ʤ/~nonword /ʧ/	/rædʒ/	/ræʧ/

Formulating the word pairs in this way allows for not only the effect of vowel duration on the production of word-final voicing to be measured, but also enables the results to evaluate the role of the lexicon and of lexical status.

As with Exp1a, for a controlled study to take place, it was important that there was clear consistency in linguistic formation across all word-tokens. Therefore, all of the recorded pairs were monosyllabic, and the word-endings consisted of an equal number of voiced and voiceless obstruents. All tokens contained the same vowel nucleus; [æ]. The decision to employ this vowel as the nucleus was informed by the results of Exp1a. On average, the stimuli containing the [æ] nucleus demonstrated the smallest durational differences in the production of the vowel. It was therefore decided that the word-tokens recorded for Exp1b would contain this monomoraic vowel, as any differences found between the measurements produced may be

considered more noteworthy. It was also necessary to use the [æ] vowel to formulate the tokens as there were few options within the English Language that would generate the necessary word and nonword pairs.

Additional materials were identical to those found in Exp1a.

Recording procedure: as with Exp1a, English speakers based in Oxford were recorded at the Language and Brain Laboratory in a soundproof booth. German speakers based in Frankfurt were recorded at Goethe University in a quiet office in the Phonetics Department. German speakers based in London were recorded in a quiet residential office. The recording set-up was identical to Exp1a, and the procedure was recreated to the greatest degree of accuracy in each of the above locations. Each of the speakers were asked to read out the word-list found in Table 5 (below), working their way from the top to the bottom of the columns.

Each of the word-tokens were repeated three times within the list, and were separated by filler-words which were also repeated three times. It was explained to the speakers that the word-list consisted of both real words and made-up words, and they were instructed to read the list out at a fixed speed, whilst controlling their intonation as much as possible. The limitations of production studies, as discussed earlier, were mitigated as much as possible in Exp1b using several strategies. Firstly, to avoid list-intonation, speakers read from a long list as opposed to simply repeating each word three times in isolation. Reading the word-tokens as part of a list in which there were fillers interspersed also aimed to encourage more regulated speech, as the speakers were not alerted to what this study may be investigating through the pattern of word-endings. As with Exp1a, speakers were unaware of which words were going to be included in the analysis, and the first and last words recorded were fillers. Filler words were generated using a random word generator.

Table 5: The list of word-tokens used for Exp1b[5]

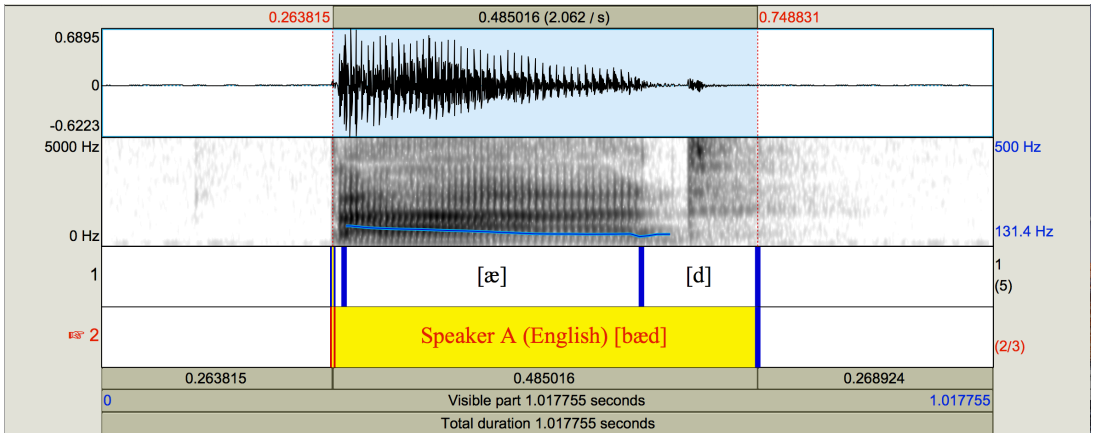
Word-token	IPA	Word-token	IPA	Word-token	IPA
drum	*/drʌm/*	*tool*	*/tuːl/*	bat	/bæt/
sour	*/saʊə/*	*dat	/dæt/	badge	/bædʒ/
*jat	/dʒæt/	bat	/bæt/	*sour*	*/saʊə/*
hip	*/hɪp/*	*jat	/dʒæt/	*jat	/dʒæt/
tool	*/tuːl/*	*deal*	*/diːl/*	*deal*	*/diːl/*
dad	/dæd/	*radge	/rædʒ/	*sour*	*/saʊə/*
hip	*/hɪp/*	*line*	*/laɪn/*	*peak*	*/piːk/*
bend	*/bɛnd/*	*peak*	*/piːk/*	vat	/væt/
launch	*/lɔːnʧ/*	*tight*	*/taɪt/*	*jad	/dʒæd/
drum	*/drʌm/*	*drum*	*/drʌm/*	bad	/bæd/
gear	*/gɪə/*	*jad	/dʒæd/	batch	/bæʧ/
*ratch	/ræʧ/	*radge	/rædʒ/	*vad	/væd/
house	*/haʊs/*	*tight*	*/taɪt/*	*dat	/dæt/
dad	/dæd/	*line*	*/laɪn/*	*launch*	*/lɔːnʧ/*
batch	/bæʧ/	badge	/bædʒ/	*jad	/dʒæd/
tool	*/tuːl/*	*bend*	*/bɛnd/*	*peak*	*/piːk/*
*ratch	/ræʧ/	*hip*	*/hɪp/*	*vad	/væd/
line	*/laɪn/*	bad	/bæd/	*launch*	*/lɔːnʧ/*
batch	/bæʧ/	dad	/dæd/	bad	/bæd/
*vad	/væd/	vat	/væt/	*gear*	*/gɪə/*
tight	*/taɪt/*	*ratch	/ræʧ/	*dat	/dæt/
bat	/bæt/	*house*	*/haʊs/*	badge	/bædʒ/
deal	*/diːl/*	vat	/væt/	*radge	/rædʒ/
house	*/haʊs/*	*bend*	*/bɛnd/*	*gear*	*/gɪə/*

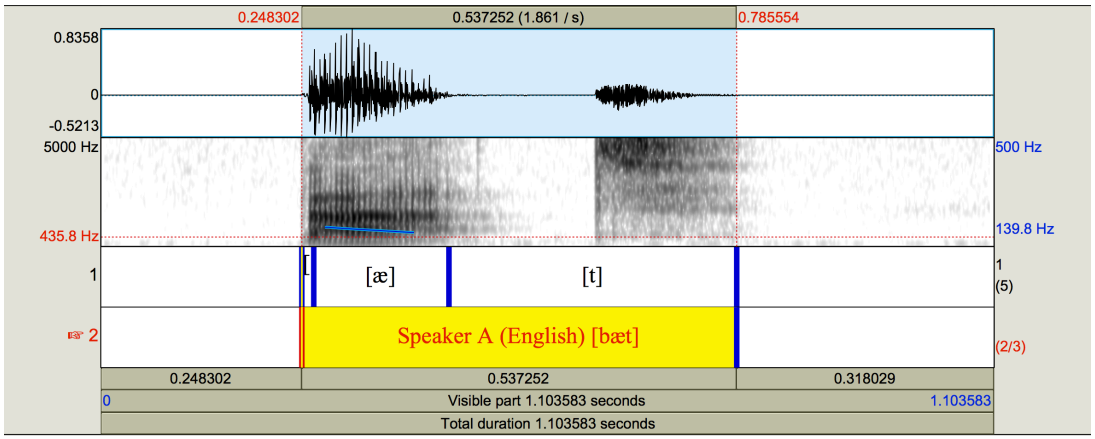
3.3.3 Measurements

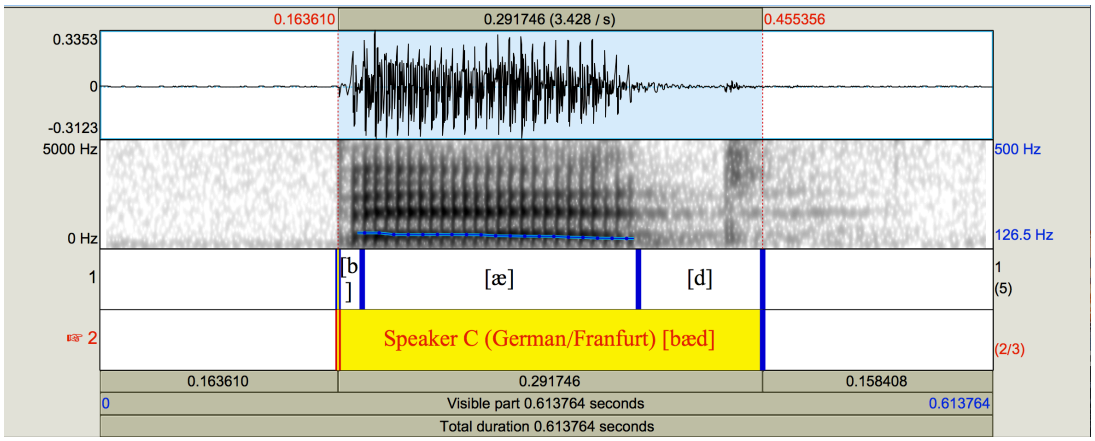
A total of two hundred and sixteen recordings were made; three repetitions of each of the twelve word-tokens as spoken by the six speakers. These recordings were then analysed using *Praat*. The format of this analysis was identical to Exp1a. This process is illustrated in Figure 7

[5] Words found in italics are fillers

(below). As with Exp1a, the measurements of each of the CVC phonemes were subsequently extracted using a *Praat* script and analysed using *JMP (SAS)*.







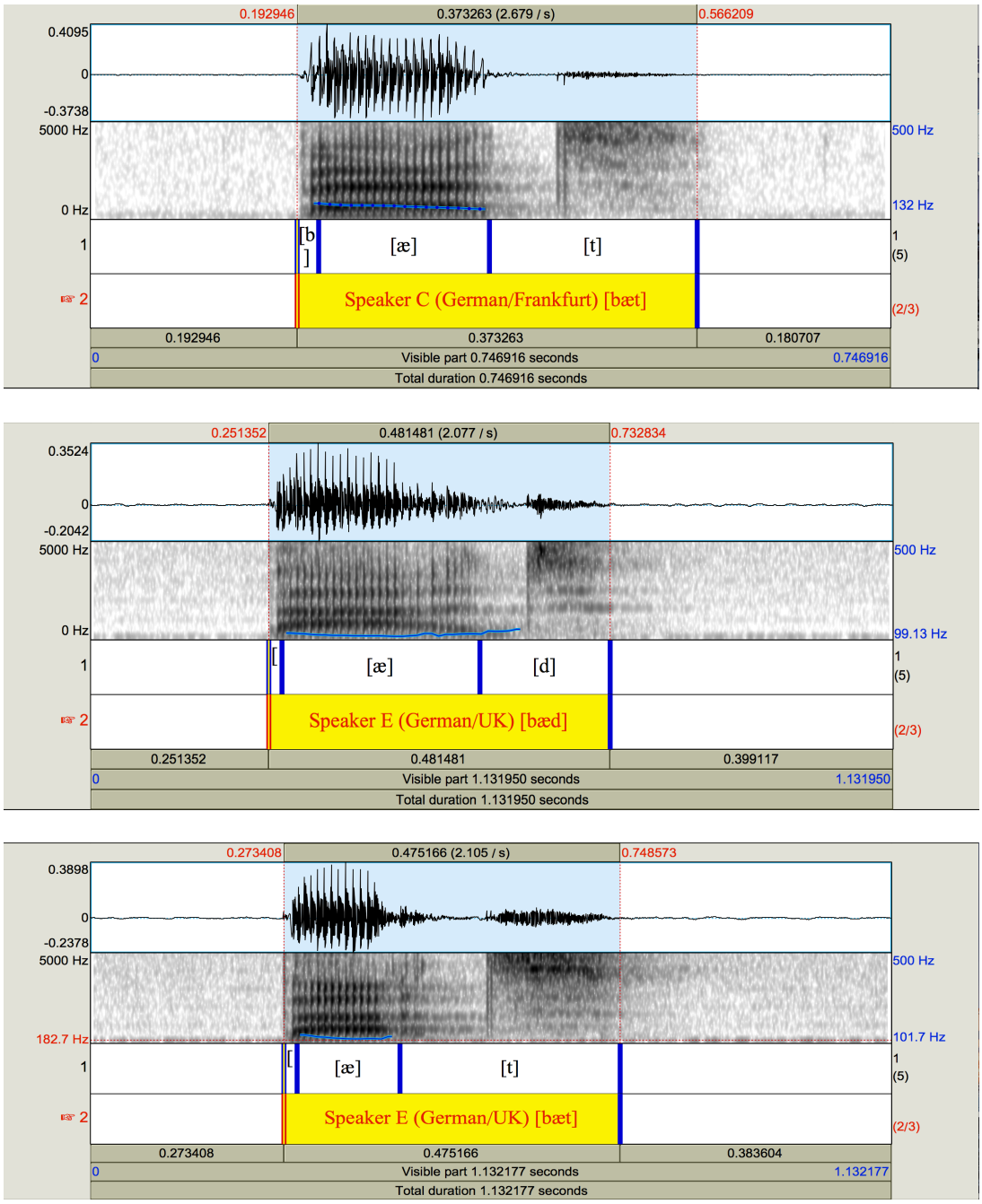


Figure 7: A demonstration of the use of *Praat* to segment the word pair [bæd]~[bæt] *bad~bat* across the three language groups for Exp1

3.3.4 Analysis

Let us first consider the results across all speakers and items for each language group, before moving on to analyse the findings from the stop and affricate pairs. Both the English (F (1, 70) = 307.29, P < .0001) and German (UK) (F (1, 70) = 119.9679, P <.0001) speakers demonstrated a highly significant difference in vowel duration, with longer vowels preceding voiced word-final consonants and shorter vowels preceding voiceless word-final consonants. German (Germany) speakers (F (1, 70) = 5.3631, P = .0235) also demonstrated a significant difference in vowel duration according to the same trend, albeit not as differentiated as the other two language groups. English and German (UK) speakers demonstrated an average difference of 128ms and 86ms

respectively, whereas German (Germany) speakers exhibited just a 25ms difference. These results are illustrated in Figures 8-10 (below).

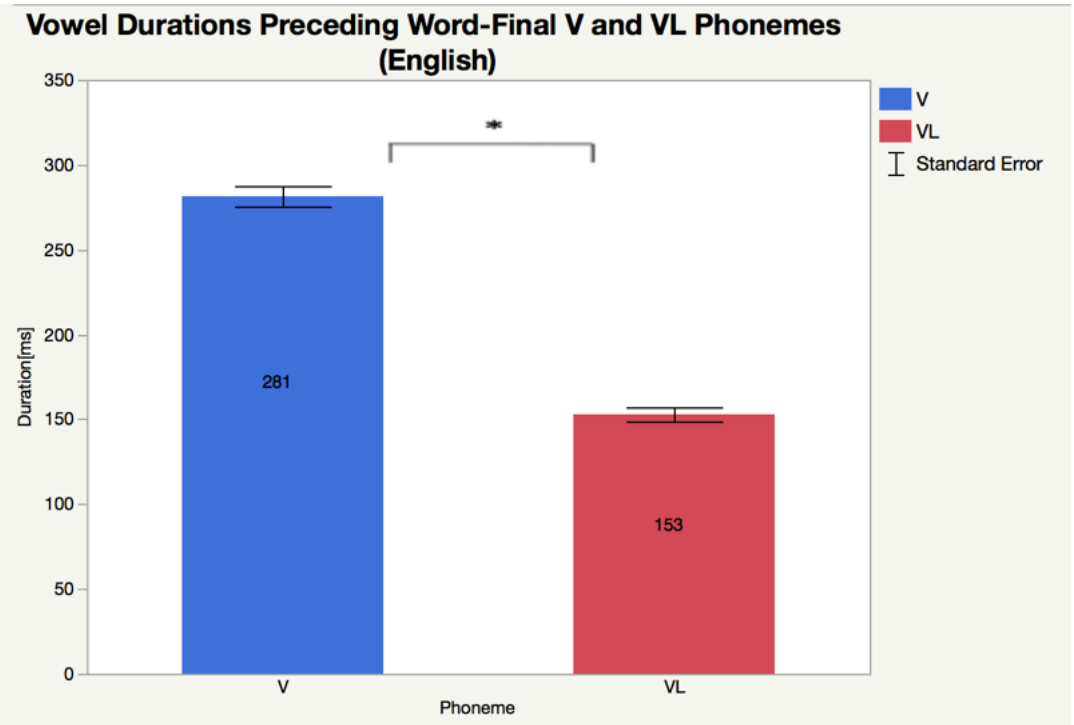


Figure 8: The difference in mean vowel duration for English speakers preceding voiced and voiceless word-final consonants for Exp1b

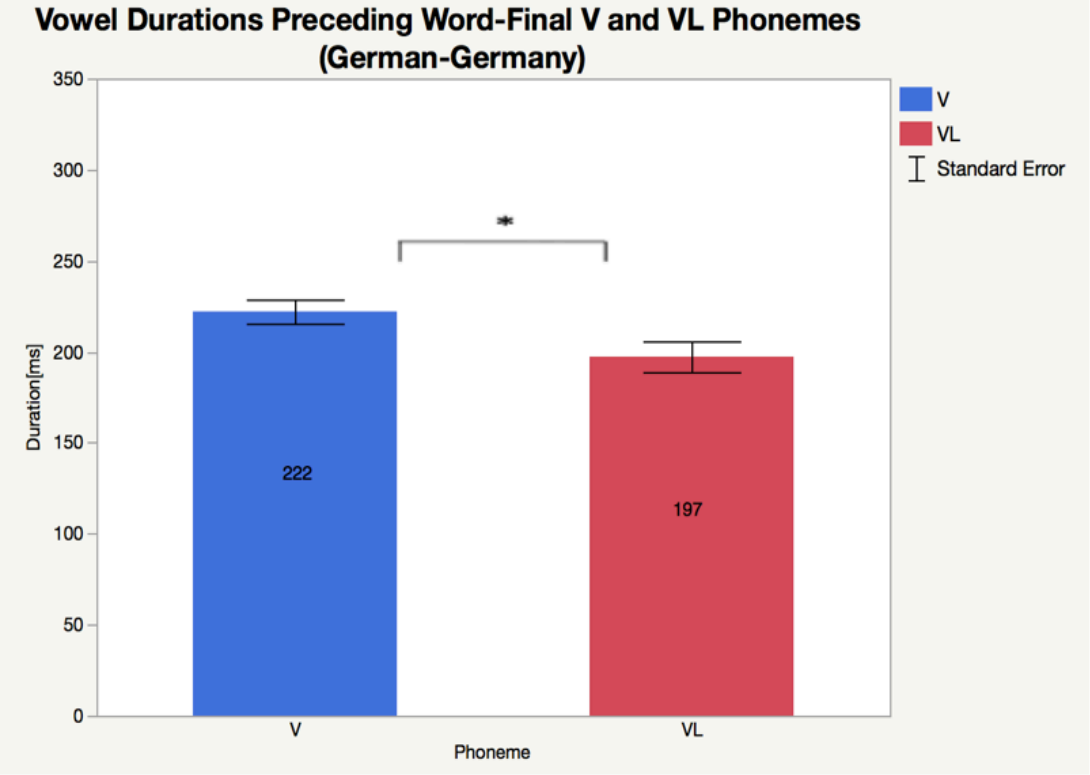


Figure 9: The difference in mean vowel duration for German (Germany) speakers preceding voiced and

voiceless word-final consonants for Exp1b

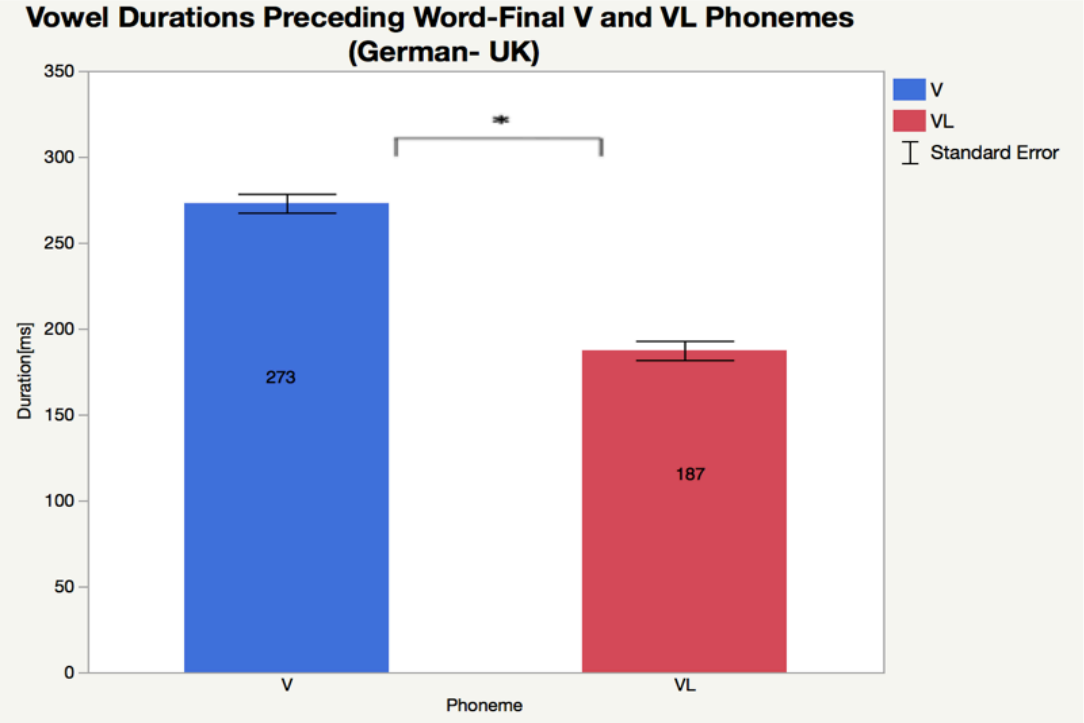


Figure 10: The difference in mean vowel duration for German (UK) speakers preceding voiced and voiceless word-final consonants for Exp1b

3.3.4.1 Stops

We will begin this analysis by characterising the results for word-final stops on a speaker-by-speaker basis. Figure 11 and Table 6 (below) illustrate the mean vowel duration that Speakers A-F exhibited in the word pairs preceding word-final [d] and [t].

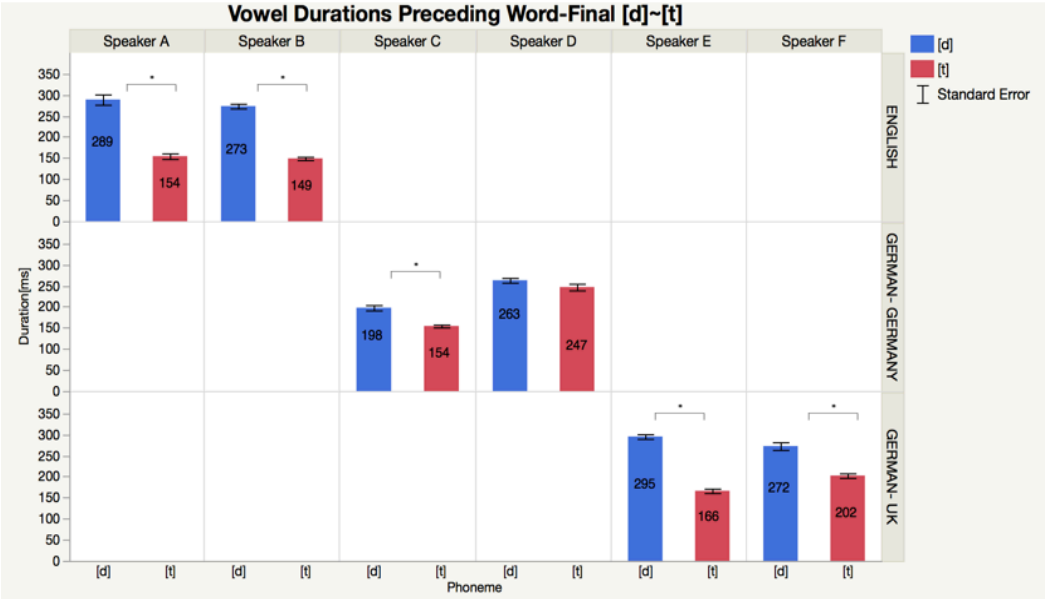


Figure 11: The difference in mean vowel duration preceding word-final [d]~[t] across the three language groups for Exp1b

Table 6: A summary of the difference in mean vowel duration preceding word-final [d]~[t] across the three language groups for Exp1b

Speaker	Language group	Difference in duration (ms)
Speaker A	English	135
Speaker B	English	124
Speaker C	German- Germany	44
Speaker D	German- Germany	16
Speaker E	German- UK	129
Speaker F	German- UK	70

Interestingly, all speakers lengthen the vowel prior to a voiced as opposed to a voiceless word-final stop. As expected, the greatest differentiation is demonstrated by the English speakers, with an average of 130ms difference. Both English speakers show a highly significant difference in vowel duration preceding word-final [d]~[t]; Speaker A (F (1, 22) = 92.84, $P < .0001$), Speaker B (F (1, 22) = 326.07, $P < .0001$). Conversely, the smallest difference originates from German speakers based in Germany, with an average difference of 30ms. Within this group, Speaker C (F (1, 22) =36.75, $P < .0001$) matches the results of the English speakers and shows a high degree of differentiation, whereas Speaker D (F (1, 22) = 2.78, $P = .1097$) does not demonstrate a significant difference overall.

Comparatively, both German speakers based in the UK, Speaker E (F (1, 22) = 289, $P < .0001$) and Speaker F (F (1, 22) = 43.75, $P < .0001$), produce vowels which show a highly significant difference preceding the voiced and voiceless word-final phonemes. As such, the German (UK) speakers look to be behaving more similarly to the group of English speakers than they are to the group of German (Germany) speakers, who show less consistency in their results. German (UK) speakers demonstrate an average difference in vowel duration of 100ms. All trends will be subsequently discussed in more detail (cf. Section 3.3.5 Discussion).

An analysis will now be presented according to the word/nonword tokens; as illustrated in Figures 12-15 and summarised by Table 7 (below):

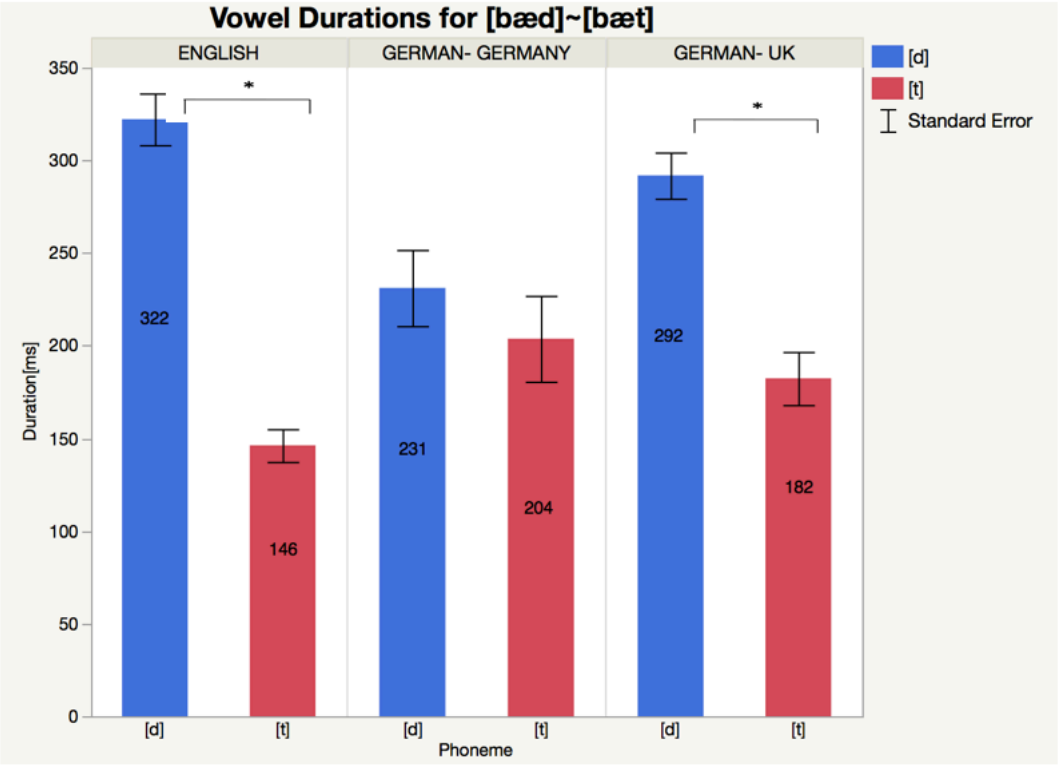


Figure 12: The difference in mean vowel duration for [bæd]~[bæt] *bad~bat* across the three language groups for Exp1b

For [bæd]~[bæt] *bad~bat*, the English (F (1, 10) = 113.65, P < .0001) speakers exhibit highly differentiated vowel durations while the German (UK) speakers (F (1, 10) = 33.31, P = .0002) demonstrate a significant difference. However, German (Germany) F (1, 10) = 0.8, P = .3968) speakers do not demonstrate a significant difference overall.

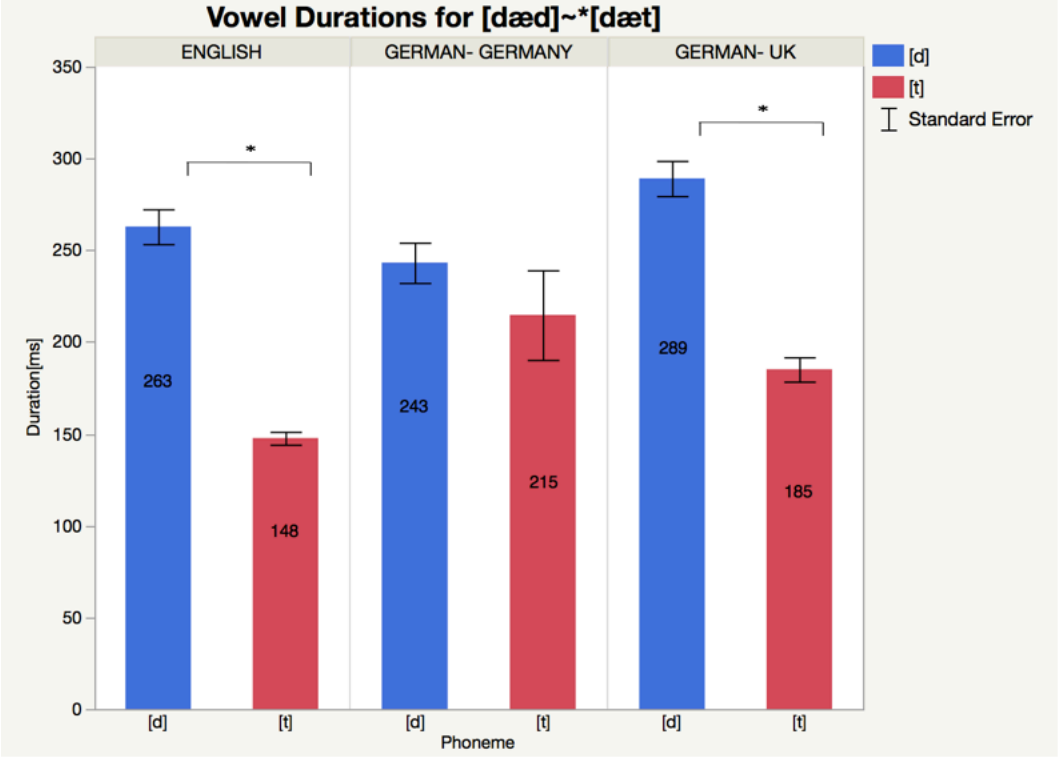


Figure 13: The difference in mean vowel duration for [dæd]~*[dæt] *dad~*dat* across the three language groups for Exp1b

Similarly, for [dæd]~*[dæt] *dad~*dat*, speakers within the English (F (1,10) = 129.5, P < .0001) and Germany (UK) (F (1, 10) = 79.8, P < .0001) groups show a high degree of differentiation between the mean vowel durations preceding the voiced and voiceless word-final endings. Again, speakers from the German (Germany) (F (1, 10) = 1.13, P = .3124) group do not exhibit a significant difference.

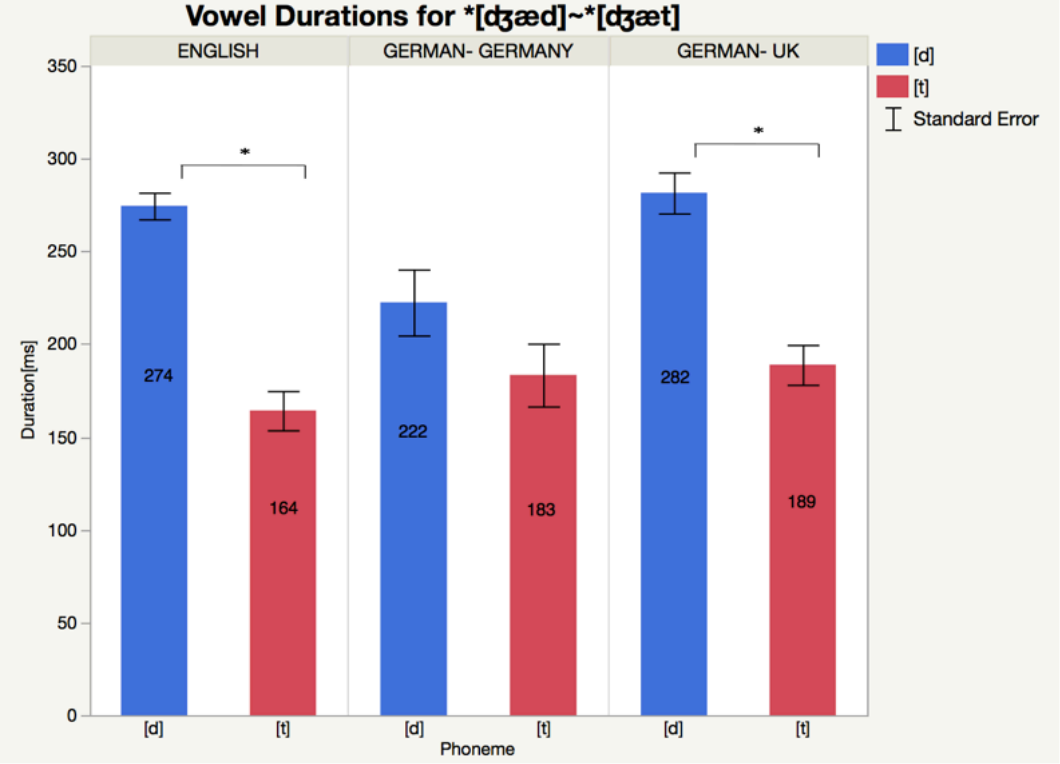


Figure 14: The difference in mean vowel duration for *[ʤæd]~*[ʤæt] **jad*~**jat* across the three language groups for Exp1b

The results for *[ʤæd]~*[ʤæt] **jad*~**jat* continue to follow this trend, with English (F (1, 10) = 74.5, P < .0001) and German (UK) (F (1, 10) = 36.16, P < .0001) speakers producing highly differentiated mean vowel durations prior to word-final [d]~[t]. Once again, German (Germany) (F (1, 10) = 2.5, P = .1419) speakers do not demonstrate a significant difference.

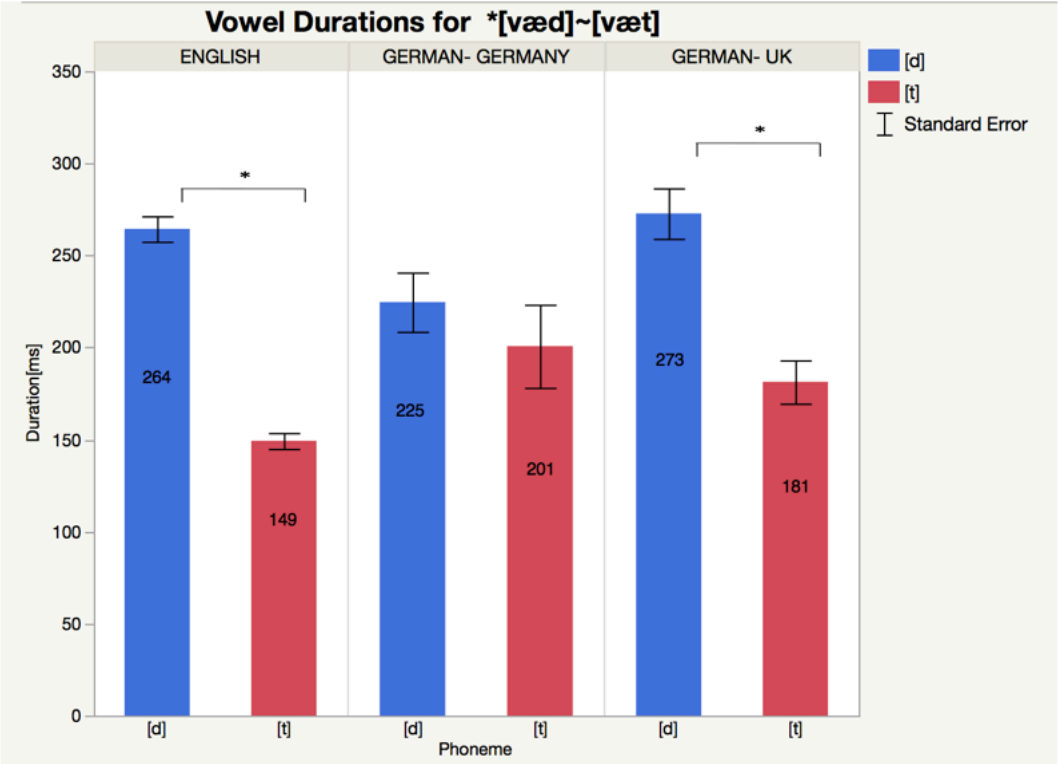


Figure 15: The difference in mean vowel duration for *[væd]~[væt] **vad~vat* across the three language groups for Exp1b

Finally, for *[væd]~[væt] **vad~vat*, English (F (1, 10) = 198.2, P < .0001) speakers show a statistically high degree of differentiation and German (UK) (F (1, 10) = 26, P = .0005) speakers show a significant difference in mean vowel duration. Comparatively, German (Germany) (F (1, 10) = 0.75, P = .4079) speakers demonstrate an insignificant difference.

Table 7: A summary of the difference in mean vowel duration preceding word-final [d]~[t] across the three language groups for Exp1b

Language group	**[bæd]~[bæt]** ***bad~bat***	**[dæd]~*[dæt]** ***dad~*dat***	***[ʤæd]~*[ʤæt]** ***jad~*jat***	***[væd]~[væt]** ***vad~vat***
English	176	115	110	115
German-Germany	27	28	39	24
German- UK	110	104	93	92

These results remain consistent in that longer vowels precede voiced word-final stops, and conversely shorter vowels precede voiceless word-final stops. However, the significance of this difference only proves relevant for the English and German (UK) speakers. Some lexical effects can be inferred from this data. Specifically, for both English and German (UK) speakers, the token-pair containing two real words yields the greatest mean durational difference between the vowels. This may be because the speakers were able to conceptualise the target words within their lexicon, and this resulted in the difference in duration between the two vowels becoming more pronounced.

Let us now turn to the results originating from stimuli ending in word-final affricates.

3.3.4.2 Affricates

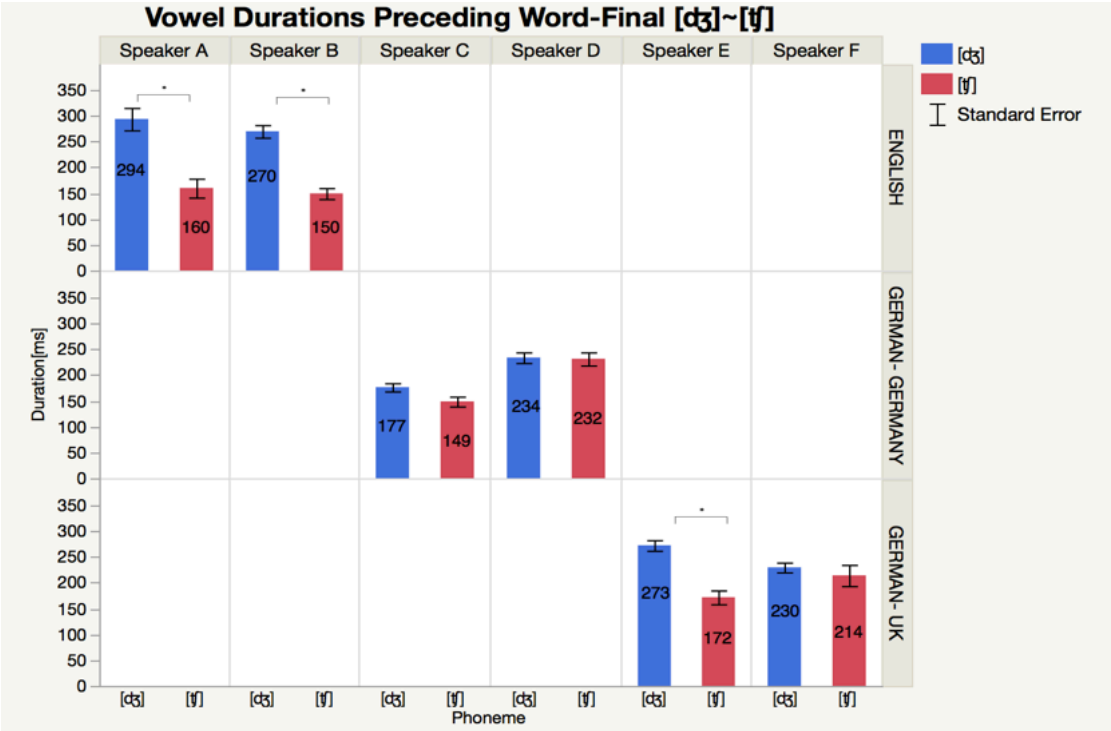


Figure 16: The difference in mean vowel duration preceding word-final [ʤ]~[ʧ] across the three language groups for Exp1b

Table 8: A summary of the difference in mean vowel duration preceding word-final [ʤ]~[ʧ] across the three language groups for Exp1b

Speaker	Language group	Difference in duration (ms)
Speaker A	English	134
Speaker B	English	120
Speaker C	German- Germany	28
Speaker D	German- Germany	2
Speaker E	German- UK	101
Speaker F	German- UK	16

Once again, CVC sequences ending in the voiced word-final affricate have been consistently preceded by a longer vowel compared with those ending in the voiceless word-final affricate, as demonstrated in Figure 16 and Table 8 (above). The difference in mean vowel duration preceding [ʤ]~[ʧ] was statistically differentiated for both English speakers; significant for Speaker A (F (1, 10) = 22.15, P = .0008) and highly significant for Speaker B (F (1, 10) = 55.44, P < .0001). However, the extent to which this is true proves negligible in the case of both German (Germany) speakers, with average differences of 28ms and 2ms. Here, neither speaker produced mean vowel durations which were significantly different; Speaker C (F (1, 10) = 4.93, P = .0507), Speaker D (F (1, 10) = 0.02, P = .8846). Interestingly, German speakers based in the UK are more divided. Speaker E (F (1, 10) = 35.55, P = .0001) produced a statistically significant difference, but Speaker F (F (1, 10) = 0.48, P = .5052) did not. Speaker F demonstrated a mean durational difference of 16ms, in comparison with 101ms from Speaker E. These results may suggest that for non-native English speakers, the production of vowel duration as a cue for word-final voicing is much less consistent for affricates than for stops. These results showed a greater variability between speakers of German who are based in the UK, as Speaker E's production was much closer in nature to that of the English speakers and

Speaker F was much closer in production to the German speakers based in Germany. This further illustrates that there can be a great deal of variation in production on a speaker-by-speaker basis, but overall it would appear that the ease of acquiring a native-like production of vowel duration varies depending on the manner of production of word-final phoneme, as well as on a speaker's own individual performance. Let us now consider these results on the basis of the word-token pairs, as demonstrated by Figures 17-18 and Table 9 (below).

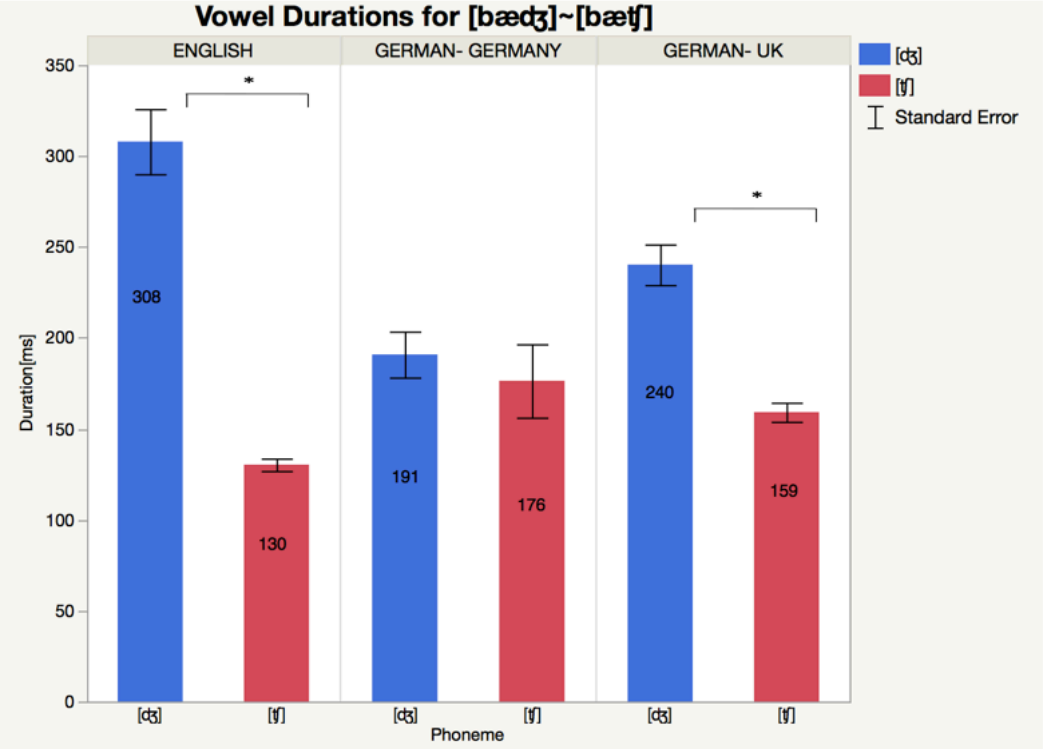


Figure 17: The difference in mean vowel duration for [bædʒ]~[bætʃ] *badge~batch* across the three language groups for Exp1b

In the case of [bædʒ]~[bætʃ] *badge~batch*, English (F (1, 10) = 95.2, P < .0001) and German (UK) (F (1, 10) = 43.14, P < .0001) speakers demonstrate highly significant differences between the vowel durations. Conversely, speakers from the German (Germany) (F (1, 10) = 0.37, P = .5569) do not demonstrate a significant difference.

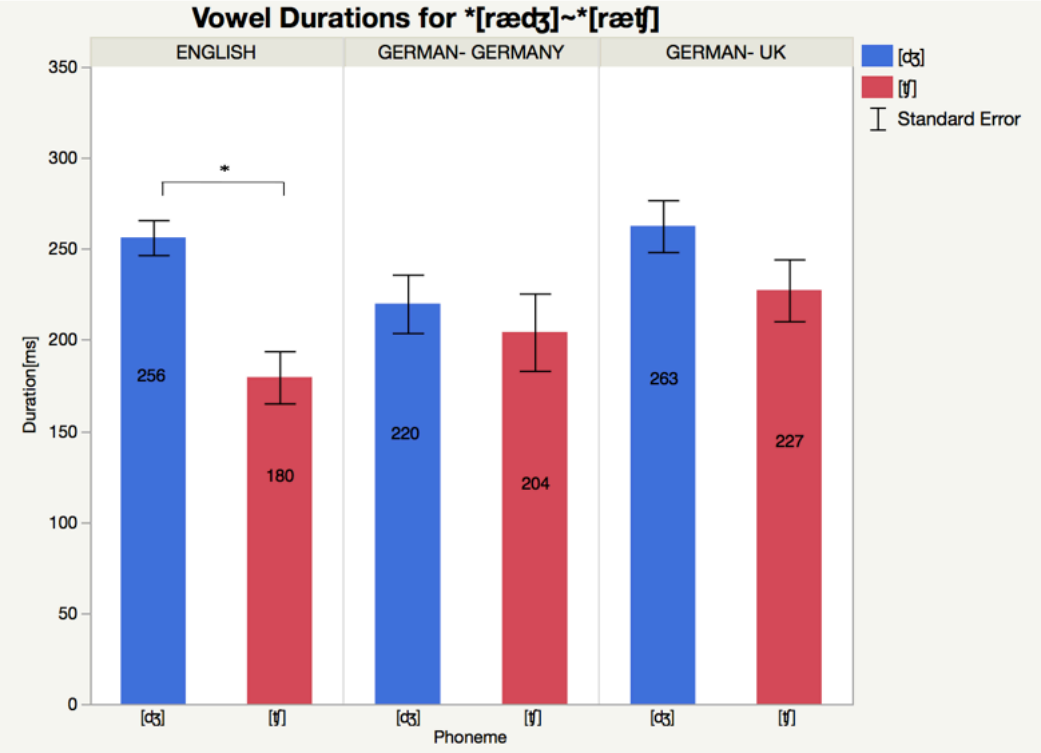


Figure 18: The difference in mean vowel duration for *[rædʒ]~*[rætʃ] **radge~*ratch* across the three language groups for Exp1b

In the case of *[rædʒ]~*[rætʃ] **radge~*ratch*, the English speakers (F (1, 10) = 19.80, P = .0012) show a significant differentiation. However, both German (UK) (F (1, 10) = 2.52, P = .1433). and German (Germany) (F (1, 10) = 0.34, P = .5708) speakers do not.

Table 9: A summary of the difference in mean vowel duration preceding word-final [dʒ]~ [tʃ] across the three language groups for Exp1b

Language group	**[bædʒ]~[bætʃ]** ***badge~batch***	***[rædʒ]~*[rætʃ]** ****radge~*ratch***
English	178	76
German- Germany	15	16
German- UK	81	36

For the group of English speakers, there was a difference of 102ms between the vowels produced for the two real words and the two nonwords. This may be related to an absence of representation in the lexicon for the two nonwords. Conversely, within the German (Germany) group, there was a negligible difference between these two groups of just 1ms. This is likely to

be because they did not have the same degree of activation in their non-native lexicon. The German (UK) speakers demonstrated a greater overall difference in the real word pairs (81ms) than the nonword pairs (36ms), similar to, but not as exaggerated as the English speakers. As with Exp1a, the individual measurements taken for vocalic duration on a speaker-by-speaker basis need to be compared in a relative manner in order for speech rate to be accounted for in the analysis. Table 10 (below) lists each speaker according to their language group, and demonstrates the average percentage of each word-token that consists of the vowel.

Table 10: The average vocalic proportion of each token for Exp1b

Speaker A- English			
Language group	**Token**	**Percentage (%)**	**Difference (%)**
English	[bædʒ] *badge*	50	30
English	[bætʃ] *batch*	20	
English	*[rædʒ] **radge*	34	8
English	*[rætʃ] **ratch*	26	
English	[bæd] *bad*	69	43
English	[bæt] *bat*	26	
English	[dæd] *dad*	51	24
English	*[dæt] **dat*	27	
English	*[dʒæd] **jad*	51	18
English	*[dʒæt] **jat*	33	
English	*[væd] **vad*	42	19
English	[væt] *vat*	23	
Speaker B- English			
Language group	**Token**	**Percentage (%)**	**Difference (%)**
English	[bædʒ] *badge*	47	23
English	[bætʃ] *batch*	24	
English	*[rædʒ] **radge*	40	14
English	*[rætʃ] **ratch*	26	
English	[bæd] *bad*	62	27
English	[bæt] *bat*	35	
English	[dæd] *dad*	58	25
English	*[dæt] **dat*	33	
English	*[dʒæd] **jad*	54	23
English	*[dʒæt] **jat*	31	
English	*[væd] **vad*	46	21
English	[væt] *vat*	25	

Speaker C- German (Germany)			
Language group	**Token**	**Percentage (%)**	**Difference (%)**
German- Germany	[bædʒ] *badge*	39	9
German- Germany	[bætʃ] *batch*	30	
German- Germany	*[rædʒ] **radge*	41	6
German- Germany	*[rætʃ] **ratch*	35	
German- Germany	[bæd] *bad*	66	25
German- Germany	[bæt] *bat*	41	
German- Germany	[dæd] *dad*	64	20
German- Germany	*[dæt] **dat*	44	
German- Germany	*[dʒæd] **jad*	54	15
German- Germany	*[dʒæt] **jat*	39	
German- Germany	*[væd] **vad*	49	12
German- Germany	[væt] *vat*	37	
Speaker D- German (Germany)			
Language group	**Token**	**Percentage (%)**	**Difference (%)**
German- Germany	[bædʒ] *badge*	41	0
German- Germany	[bætʃ] *batch*	41	
German- Germany	*[rædʒ] **radge*	43	-1
German- Germany	*[rætʃ] **ratch*	44	
German- Germany	[bæd] *bad*	56	2
German- Germany	[bæt] *bat*	54	
German- Germany	[dæd] *dad*	57	5
German- Germany	*[dæt] **dat*	52	
German- Germany	*[dʒæd] **jad*	51	9
German- Germany	*[dʒæt] **jat*	42	
German- Germany	*[væd] **vad*	45	3
German- Germany	[væt] *vat*	42	
Speaker E- German (UK)			
Language group	**Token**	**Percentage (%)**	**Difference (%)**
German- UK	[bædʒ] *badge*	36	13
German- UK	[bætʃ] *batch*	23	
German- UK	*[rædʒ] **radge*	41	14
German- UK	*[rætʃ] **ratch*	27	
German- UK	[bæd] *bad*	57	27
German- UK	[bæt] *bat*	30	
German- UK	[dæd] *dad*	53	21
German- UK	*[dæt] **dat*	32	
German- UK	*[dʒæd] **jad*	47	20
German- UK	*[dʒæt] **jat*	27	
German- UK	*[væd] **vad*	48	23
German- UK	[væt] *vat*	25	

Speaker F- German (UK)			
Language group	**Token**	**Percentage (%)**	**Difference (%)**
German- UK	[bædʒ] *badge*	39	11
German- UK	[bætʃ] *batch*	28	
German- UK	*[rædʒ] **radge*	34	-1
German- UK	*[rætʃ] **ratch*	35	
German- UK	[bæd] *bad*	58	18
German- UK	[bæt] *bat*	40	
German- UK	[dæd] *dad*	54	17
German- UK	*[dæt] **dat*	37	
German- UK	*[dʒæd] **jad*	53	16
German- UK	*[dʒæt] **jat*	37	
German- UK	*[væd] **vad*	50	16
German- UK	[væt] *vat*	34	

For the English speakers, the average percentages were similar for Speakers A and B. Speaker A demonstrated a 24% average difference for the proportion of each token which consists of the vowel between the voiced and voiceless word-final phonemes. Similarly, Speaker B demonstrates a 22% average difference. For native German speakers based in Germany, these averages were expectantly lower with Speaker C demonstrating a 15% difference in vocalic proportion and Speaker D demonstrating just a 3% difference. Consistent with the results reported above, German speakers based in the UK showed more variation. Speaker E behaved similarly to the English speakers with a 20% difference in vocalic proportion across all tokens, whilst Speaker F behaved at a more intermediate level between the two language groups, demonstrating a 13% difference. For all except Speaker D (German-Germany), the real word stop-pair consistently showed the largest difference in vocalic proportion when compared with the other word/nonword combinations. For all but Speaker E (German- UK), this was also the case for the affricate-pairs; with the real word pair demonstrating a larger difference than the nonword pair. In all instances, except for Speaker E, the nonword affricate-pair exhibited the smallest difference, overall. Interestingly, in the case of Speakers D (German- Germany) and F (German- UK), no real difference was found between

word-final [ʤ]~[ʧ] for the nonword affricate-pair, and the vocalic portion in the voiceless case is actually slightly longer than that in the voiced case. Speech rate is therefore not responsible for the durational differences in the vowel as if it were, we would expect to see similar percentages for the vocalic portions across the speakers.

3.3.5 Discussion

In the case of word-final stops, all speakers systematically lengthened the vowel duration prior to a voiced word-final consonant, relative to a voiceless word-final consonant. Looking at the results overall, although this tendency is consistent across all speakers, English and German (UK) speakers demonstrated a consistently high significant difference in mean vowel duration prior to a [d] and [t] (P <.0001). However, for the German (Germany) speakers, this trend was less consistent (P <.0001, P = .1097).

For the individual word and nonword stop-pairs, English speakers continued to consistently produce highly differentiated vowel durations (P <.0001), and the German (UK) speakers also demonstrated a significant difference in vowel duration, high in some cases, though lower in others, such as [bæd]~[bæt] *bad~bat* (P = .0002) and *[væd]~[væt] **vad~vat* (P = .0005). German (Germany) speakers showed insignificant differences in their production of vowel duration across the word and nonword pairs.

For word-final affricates, English speakers demonstrated a consistently significant difference in vowel duration (P = .0008, P <.0001). German speakers based in the UK demonstrated a significant difference in the case of one speaker but not the other (P = .0001, P = .5052), and for German speakers based in Germany, neither speaker demonstrated a significant difference (P = .0507, P = .8846).

Let us now consider these results by way of providing some explanation. It would appear that German speakers based in Germany were being primarily guided by their L1 phonology. Though the voicing contrast is neutralised word-finally in the surface form, German does contain both underlyingly voiced and voiceless stops. So, given that English demonstrates vowel lengthening in relation to surface-level voiced and voicelessness, it is likely that when operating in L2 English, native speakers of German who have had a greater exposure to English can produce more native-like vocalic durations. This is due to the underlying representation of the distinction between voiced and voiceless word-final stops.

Conversely, German does not contain underlyingly voiced affricates word-finally, unless borrowed into the language. This is particularly reflected in the individual speaker data for affricates, with only one German speaker, based in the UK, demonstrating a significant difference in vowel lengthening. As such, it would appear that in both the case of stops and affricates, an increased exposure to spoken English leads to a greater degree of vowel lengthening in accordance with the pattern of English and a more native-like production overall. However, this is speaker-dependent as Speaker E demonstrated this tendency to a greater extent than Speaker F, overall. Of course, given that this is a small-scale study, it is difficult to make inferences that could account for widespread trends in second-language English speakers. An increased immersion in a second-language may well lead to more native-like pronunciations in some speakers. Importantly, this set of small-scale results suggest that even fine-grained phonetic cues such as differences in vowel duration appear to be a learned phenomenon.

With regard to the lexicon, the greatest degree of lengthening was present in the case of the real word pairs for both the English and German (UK) speakers, with a lesser extent of lengthening exhibited in pairs containing nonwords. This was also demonstrated when looking

at the vocalic proportions of the tokens, with the real word pairs exhibiting the greatest degree of difference overall; with exceptions from the stop-pair for Speaker D (German- Germany) and the affricate-pair for Speaker E (German- UK). The nonword affricate pair also exhibited the smallest difference overall, except for Speaker E (German- UK). This alludes to the role of the lexicon and the relationship between a strong mental representation of a word and a more native-like production of phonology.

Deciphering whether these trends in production can translate into the perception of L2 languages, along with a more nuanced exploration of the role of the lexicon, will be further tested in the subsequent perceptual experiments encompassed within this thesis.

3.3.6 Conclusion

The aim of Exp1b was to build on the findings of Exp1a and undertake a broader production study in three key ways. Firstly, it incorporated two distinct groups of word-final consonants; stops and affricates. It also explored the notion of non-native speech production by measuring the differences between native English and native German speakers, along with investigating whether an increased degree of exposure to spoken English may lead L2 speakers to acquire more native-like production. Finally, it considered the role of the lexicon by including both real word pairs an nonword pairs.

As expected, English speakers consistently demonstrated vowel lengthening prior to a voiced word-final consonant. Additionally, German speakers based in the UK also consistently demonstrated this tendency, albeit to a lesser degree than the English speakers overall. Conversely, speakers of German who were based in Germany did not consistently demonstrate significant vowel lengthening. Native English speakers consistently produced the overall greatest degree of lengthening, in comparison with non-native English speakers. This indicates

that although an L2 English speaker may have a good grasp of English production, their production of fine-grained phonetic cues such as vowel duration are likely to be less pronounced than native speakers. Particularly, when these cues do not exist in their L1 phonology.

Interestingly, the results of Exp1b do seem to suggest that there is a positive correlation between the amount of exposure to spoken English that the German speakers have received and the extent to which they lengthen their vowels prior to a voiced word-final consonant. German speakers based in the UK demonstrate vowel lengthening across all word pairs, and their results were more similar to the English speakers than they were German speakers based in Germany who had good, but limited, English proficiency.

In addition, the largest durational differences between voiced and voiceless pairs were overall found within the real word pairs as opposed to pairs containing nonwords. For English speakers, this certainly the case, with the nonword/nonword pairs for both stops and affricates resulting in the lowest difference overall. As previously discussed, this is likely to be due to a stronger mental representation and activation in the case of real words, and an absence of this representation in the case of nonwords.

3.4 Overall conclusion

The results of both production studies support previous literature which states that in English, voiced word-final phonemes are preceded by a notably longer vowel than voiceless word-final phonemes (House, 1961; Raphael, 1975; Klatt, 1976; de Jong, 1991). Thereby, the production of preceding vowel duration is dependent on the nature of the [VOICE] feature word-finally. The results also support literature addressing non-native speech production which suggest that L1 phonology can influence L2 production (Edge, 1991), as well as literature which implies that the production of vowel duration is a learned aspect of language (House, 1961). Primarily, these findings suggest that physiological theories for vowel lengthening, such as those detailed in Beguš (2017), must frame themselves alongside the phonological attributes of a language in guiding the production of vowel duration in relation to word-final voicing.

One of the most significant outcomes of this research is the demonstration that an increased immersion within an L2 language seems to have the potential to lead to more native-like production. The concept that the pronunciation of a second language improves with exposure to that L2 language is not surprising. However, what is interesting is that this translates to even the fine-grained phonetic differences found within a second-language, even when they are absent from a first-language.

Each of these conclusions must be interpreted within the confines of this small-scale study. Speaker variability is a primary limitation of production studies, and consistency within a production study is crucial in order for viable conclusions to be drawn. Nine speakers in total were incorporated into Experiment 1, and these findings are therefore based on this limited sum of averages. A much larger-scale study would be needed in order to develop a more nuanced understanding of these results; however, the primary purpose of this thesis is to

investigate speech perception. These production studies were therefore designed simply to provide a first-hand and theoretical introduction to this research.

CHAPTER FOUR

The effect of vowel duration on the perception of English word-final voicing: identification tasks for English native speakers

This chapter encompasses two forced choice identification tasks concerned with determining the extent to which vowel duration acts as a primary perceptual cue for word-final voicing in English. In Chapter Three, Experiment 1 established that the production of vowel duration is closely linked to the nature of the word-final [VOICE] feature; with longer vowel durations preceding a voiced word-final consonant and comparatively shorter vowel durations preceding a voiceless word-final consonant. Chapter Four therefore seeks to address how this tendency translates from speech production into speech perception.

Experiment 2 contained stimuli consisting of CV:?/CV? structures in which the word-final sound was an ambiguous consonant. Participants were forced to choose between two sounds as to which one they perceived hearing word-finally. This experiment also considered the role of the lexicon. Conversely, in Experiment 3, participants were presented with a CV:/CV structure and asked to decide between two word-final sounds as to which they thought *should* finish the word. Both Experiments 2 and 3 utilised recorded speech to form the experimental stimuli.

The findings from each experiment will be presented individually, referring to literature which both supports and refutes their findings, before generalised conclusions are considered in an overall summary of the results.

4.1 Experiment 2: identifying [d]~[t] word-finally

In the first of the three identification tasks, Experiment 2 formed a preliminary investigation into the perceptual links between vowel duration and word-final voicing in English. Experiment 2 also considered the effect that the lexicon and lexical bias may have on perception, and the extent to which this may interact with or surpass the acoustic information made available to listeners through the duration of the vowel.

4.1.1 Research questions

Experiment 2 aimed to determine how manipulating vowel duration may affect participants' perception of word-final voicing. Additionally, it explored whether participants would behave in such a way that would suggest creating a real word is considered more influential on decision-making than the acoustic information made available to them through gated vowel durations.

Based on previous literature such as Ganong (1980), it was expected that a longer vowel gate would naturally elicit an underlying /d/ response; however, where this would produce a nonword participants would be more inclined to perceive an underlying /t/ response. For example, if a participant was expected to hear the word *[heɪd] **hade* due to a longer vowel gate, they may be more likely to perceive [heɪt] *hate*. Similarly, shorter gates were expected to elicit more underlying /t/ responses, except in the cases where this would create a nonword. In these instances, participants were expected to select a [d] response. For example, participants who would be expected to hear *[dʒeɪt] **jate*, as primed by a shorter vowel gate, may be more inclined to select a [d] response, due to being biased by the real word [dʒeɪd] *jade*. Results reflecting these patterns would suggest that lexical bias has a more significant effect on perception than the vowel duration. Conversely, in cases where there were either two word

pairs, for example [feɪd]~[feɪt] *fade~fate* or two nonword pairs, for example *[zeɪd]~*[zeɪt] **zade~*zate*, it was expected that the results would reflect a perceptual trend based on the vowel duration. Here the shift from a [d] response to a [t] response was expected to happen in a gradual manner due to the absence of lexical bias influencing the participants' perception. As such, it was expected that participants would respond along a continuum and that we would see steadily more [t] responses as the vowel gates got systematically shorter in duration. However, because it was expected that participants were likely to be influenced not only by vowel duration, but also by lexical status, participants were expected to shift from a [d] response to a [t] response in a categorical manner, and in the direction of a word as opposed to a nonword, where lexical bias was applicable.

4.1.2 Methodology

The recording: a thirty-year-old male speaker with a Southern British English dialect was recorded in a soundproof booth in the Language and Brain Laboratory at the University of Oxford using a Rode NT-USB microphone. *Audacity* was used to capture the recordings, using a mono 44.1khz sampling rate on a Macbook computer. The word and nonword tokens that were recorded for Experiment 2 can be found in Table 11 (below). The speaker was instructed to read the list out at a fixed speed, trying to avoid list-intonation by keeping their pitch, volume and intonation as consistent as possible. They repeated each item on the list five times. Two recordings of the list were made using this procedure, and the second recording was used to make the stimuli as the speaker had had time to familiarise themselves with the procedure to a greater extent, and this had produced an improved recording. Word frequencies were obtained using the CELEX database, and the word/word pair was controlled for word frequency.

Table 11: The list of word-tokens used for Experiment 2[6]

Token	IPA	Word frequency
fade	/feɪd/	643
fate	/feɪt/	615
*zade	*/zeɪd/	*
*zate	*/zeɪt/	*
*hade	*/heɪd/	*
hate	/heɪt/	1915
jade	/dʒeɪd/	31
*jate	*/dʒeɪt/	*
aid	*/eɪd/*	*
ate	*/eɪt/*	*

The stimuli: *Praat* was used to cut the recording into individual word-tokens. To do this, a *TextGrid* was positioned around the third example of each of the five recordings for each token, and labelled with the corresponding word or nonword. The third recordings were selected as they were the middle token of the five recordings for each example, and therefore proved to be the most consistent in terms of pitch, intonation, and volume.

Figure 19 (below) demonstrates the way in which each token was selected, including approximately 0.2s of silence on either side of the word boundary to ensure that no relevant acoustic information was omitted from the tokens. Here, as in all instances, the recordings were cut at a zero-crossing boundary to avoid any unwanted clicks and to ensure that the files sounded as natural as possible. Each of the sound files were then extracted from the recording using the *Extract Non-Empty Intervals* command in *Praat*, and saved as separate files.

[6] Words found in italics were used to synthetically manipulate the stimuli, and were not themselves an experimental word pair

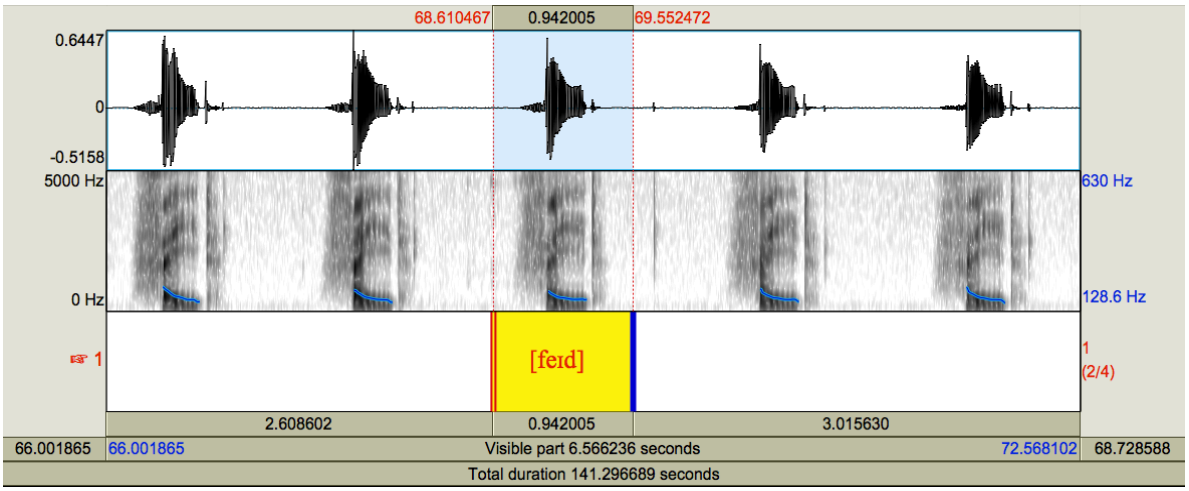


Figure 19: The use of *Praat* to segment the recordings, here [feɪd] *fade,* for Experiment 2

The next stage of forming the stimuli involved creating a word-final ambiguous obstruent that measured acoustically between a [d] and a [t] phoneme. This consonant would then form the word-final phoneme for each of the stimuli.

To achieve this, [eɪd] and [eɪt] files were opened in *Praat*. The release burst of each word-final consonant, [d] and [t] respectively, was selected and then extracted to form two new files. Here, the zero-crossing boundary was placed at end of the silent closure duration for both [d] and [t] where a clear change in the energy, signalling the end of the period of silence and the beginning of the release burst, was both visually and audibly clear. This segmentation is demonstrated in Figure 20 (below). These two files were then converted into one stereo file which consisted of two channels, the release burst of the [d] and the [t], before being converted again into a mono file using the *convert to stereo* and *convert to mono* commands in *Praat.* In this way, the two channels were acoustically overlaid and combined to form one ambiguous release burst measuring half way between that of the original [d] and [t] channels. This process can be seen in Figures 21-22 (below). The most important aspect to note about this release burst is that the difference in voicing between the [d] and [t] phonemes was neutralised.

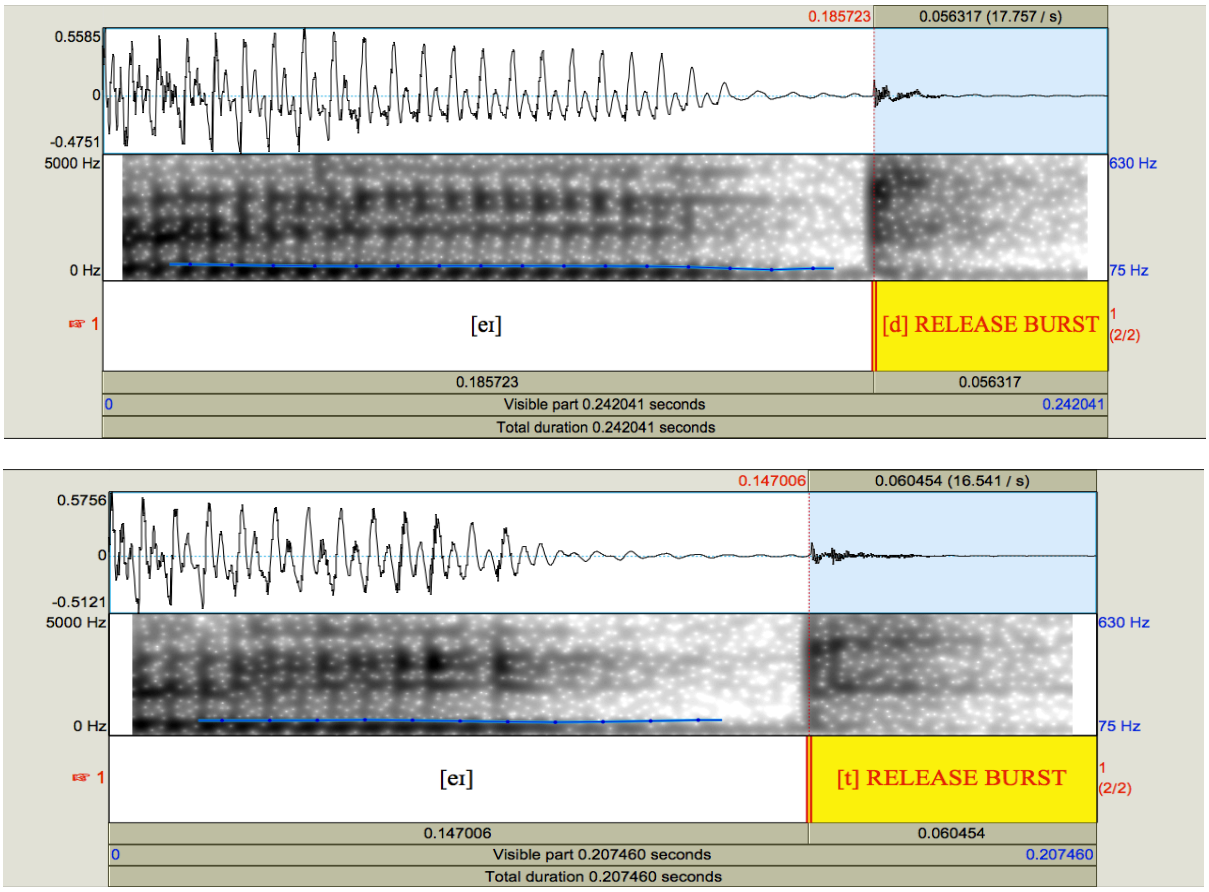


Figure 20: The use of *Praat* to segment the release burst for the recordings for Experiment 2

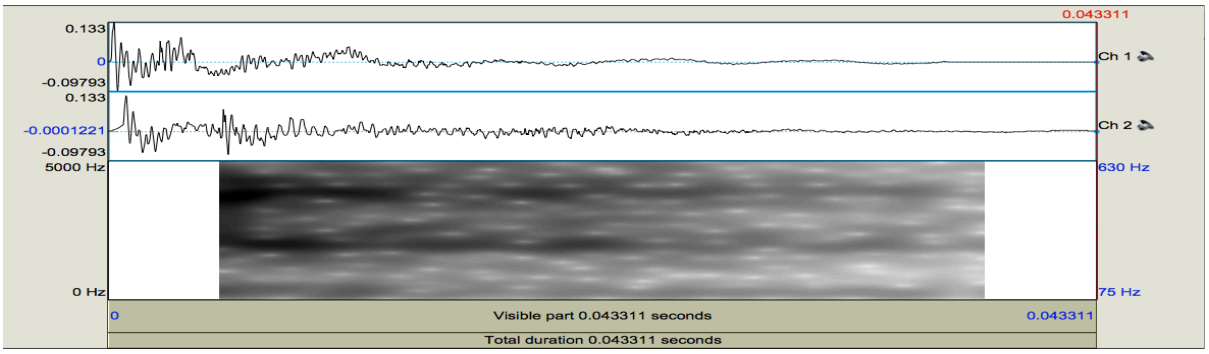


Figure 21: The use of *Praat* to convert the [d] (channel 1) and [t] (channel 2) release bursts into two stereo files for Experiment 2

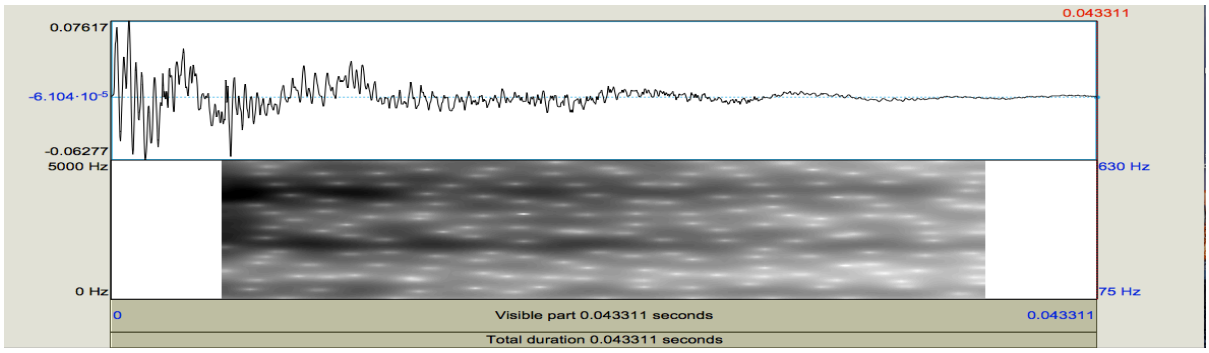


Figure 22: The use of *Praat* to create a single mono file by overlaying channel 1 and channel 2 (Figure 21) for Experiment 2

The next stage of the stimuli creation process was then to paste the ambiguous release burst onto the [eɪd] file. The [eɪd] file was chosen to form the basis of the stimuli as the vowel contained within this file was the longer of the two. This proved to be the most effective base, as making a longer vowel shorter in duration proves more methodologically efficient than making a shorter vowel longer in duration. The release burst of the original word-final [d] consonant was omitted from the [eɪd] file and the ambiguous release burst was pasted in its place at the end of the original closure duration. Retaining the original closure duration of the voiced phoneme overcame the issue of the durational differences between the original [d] and [t] phonemes, such that the closure of one did not overlap with the release of the other. Here, it must be acknowledged that some acoustic cues for voicing may be present in the retained closure duration. However, the most crucial aspect of the stimuli formation was that the nature of voicing in the release burst had been neutralised. The benefit of retaining a consistent closure duration across the stimuli in order to overcome issues with the overlapping of the original [d] and [t] phonemes was judged to be of greater importance than the possibility of retained cues in the closure duration.

This process is illustrated in Figure 23 (below). The ambiguous release burst was pasted into the spectrogram in the same position on the spectrogram as the original word-final [d] release burst had been excised. This file was then saved as [eɪX], and subsequently formed the basis for the varying vowel durations.

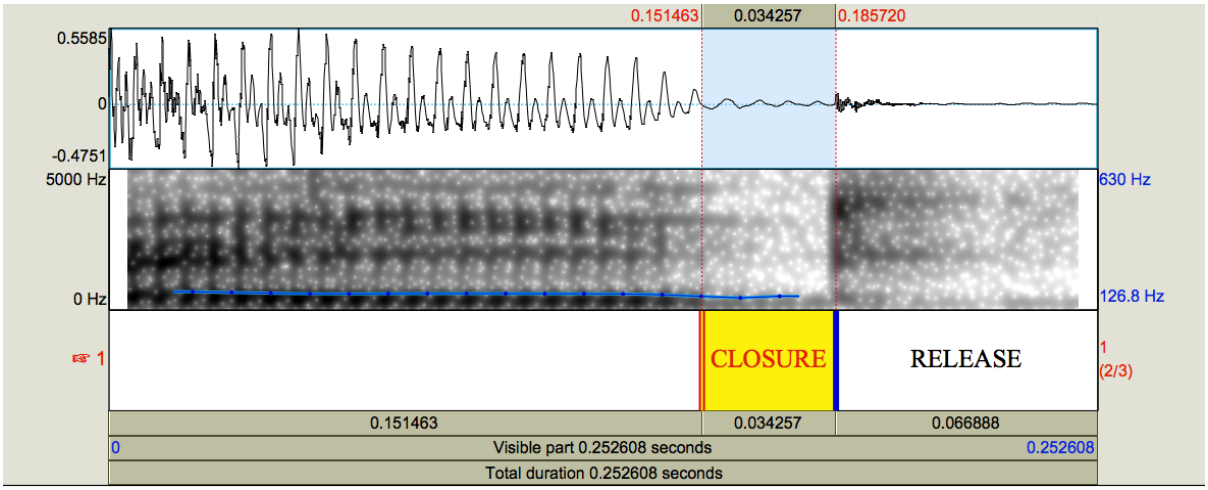


Figure 23: The use of *Praat* to replace the word-final [d] consonant in the [eɪd] file with the ambiguous consonant to create the [eɪX] file for Experiment 2

The vowel gates were created by systematically removing one glottal pulse at a time from the [eɪX] file. This process involved cutting out the middle glottal pulse before moving outwards from the centre of the vowel, cutting one to the left, one to the right and so on (Figure 24, below). This method was chosen as it retained the most natural volume, pitch, and intonation across the vowel, thereby avoiding synthetic sounding speech as much as possible. Once again, all cuts were made at the zero-crossing boundary. Again, it is possible that the end of the vowel retained cues for the original voiced nature of the word-final consonant. However, as with overcoming potential cues retained in the closure duration, consistency was key in controlling for any obvious bias caused by these cues in the analysis phase of Experiment 2. Had shorter vowel durations consistently been perceived as voiced, one would have been able to identify these cues as a potential reason for this bias. Maintaining the natural shape of the vowel was judged to be more important than cutting the glottal pulses from the end of the vowel, despite the potential for spectral cues to be retained in the formant frequency transitions.

Each time a glottal pulse was removed from the [eɪX] file, a new file was created and saved containing the systematically shorter vowel durations. This process resulted in ten vowel gates, each one consisting of a vowel duration that was one glottal pulse shorter than the previous.

Gate 10 was 65ms shorter than gate 1. The specific durations for each of the vowel gates are depicted in Table 12 (below):

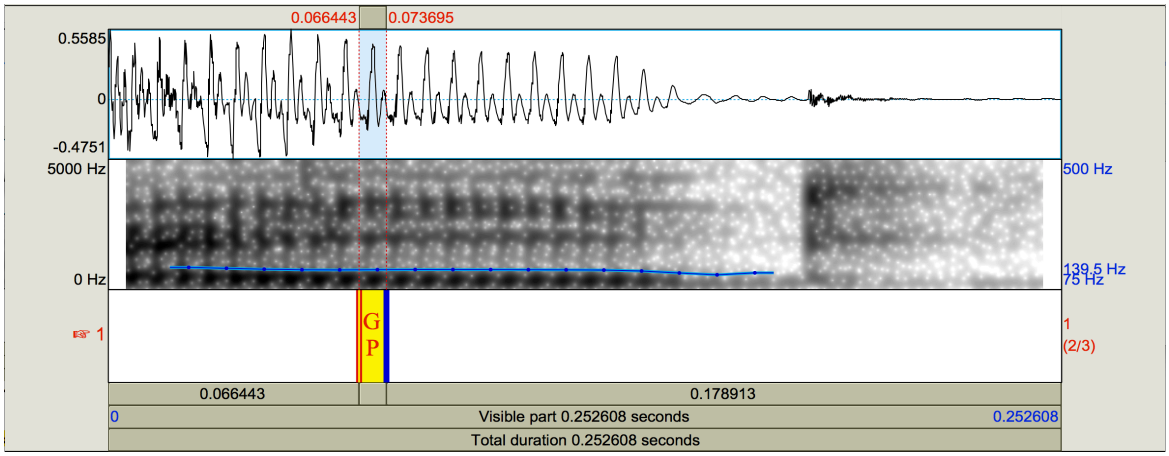


Figure 24: The use of *Praat* to remove glottal pulses systematically from the centre of the [eɪX] file for Experiment 2

Table 12: Measurements of vocalic duration across the ten vowel gates for Experiment 2

Gate	Duration (ms)
1	253
2	245
3	238
4	231
5	224
6	217
7	209
8	203
9	195
10	188

The final stage of creating the stimuli was to paste the word-initial consonants from the four word/nonword pairs onto each of the ten vowel gates. The third repetition of each of the four consonants [f], [z], [h], and [dʒ] was cut from the voiced word-final recordings. Again, the third repetition was chosen as it was the middle recording of the five repetitions for each

stimulus, and was therefore judged to be the most regulated. The voiced base was also once again used for consistency. Both visual and auditory judgments were made in order to select and extract the relevant phonetic information from the spectrogram needed to identify the initial consonant. Here, as with all other instances, the zero-crossing boundaries were placed where there was a clear change in the energy and shape of the spectrogram. Where the word-initial consonant moved into the vowel, a boundary was placed prior to the beginning of the first vocalic glottal pulse. As referenced earlier (Coleman, 2003), it is possible that word-final voicing cues may be retained within these word-initial consonants. As with the previous instances, it must be acknowledged that these cues may have the potential to bias the listeners. However, as the same instance of the four word-initial consonants was used for all stimuli to which each consonant pertained, any cues for word-final voicing which were biasing the listener would be made apparent. For example, if the stimuli containing shorter vowels were perceived more often as voiced. As with all other aspects of the stimuli formation, ensuring this level of consistency across the stimuli acted as a control which enabled any additional cues, such as those possibly retained within these word-initial consonants, to be identified. The four consonants were saved individually, and then systematically added to the respective vowel gates, word-initially, to create a total of forty stimuli, with four initial consonants and ten different vowel gates (Figures 25-26, below). The word-initial consonants were pasted onto the vowel gates prior to the beginning of the first vocalic glottal pulse.

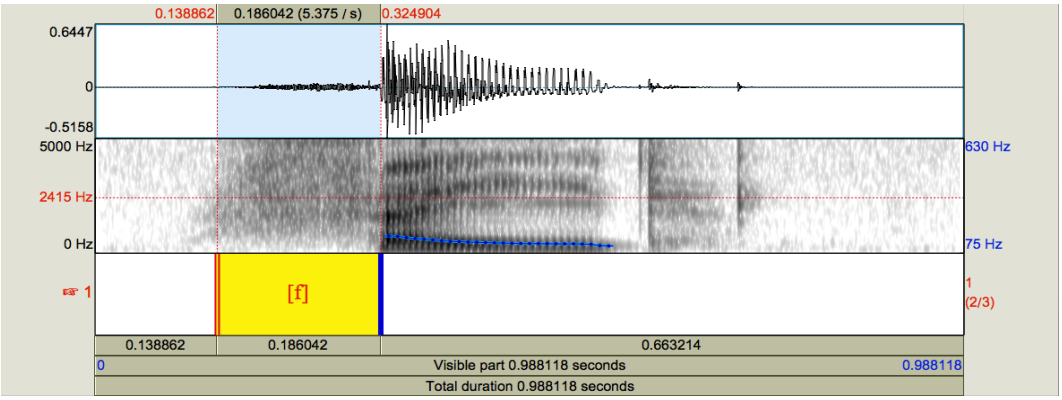


Figure 25: The use of *Praat* to cut the word-initial consonants, here [f]. from the original sound recordings for Experiment 2[7]

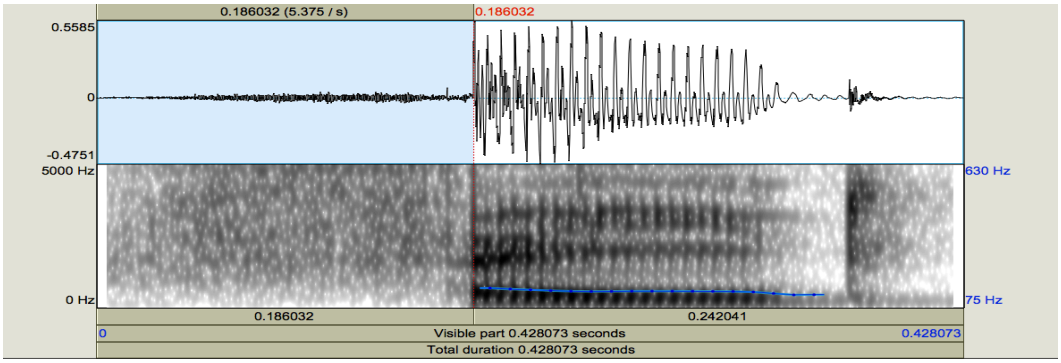


Figure 26: The use of *Praat* to paste the word-initial consonants, here [f], onto each of the ten vowel gates for Experiment 2[8]

The result of this process meant that each individual element of the final experimental stimuli had been synthetically altered and pasted into place, leaving no aspect of the recordings remaining in their original position. The level of consistency across the stimuli formation ensured that the only acoustic cue which was systematically varied was the vowel duration. It was highly important that each stimulus sounded natural, however it was equally important that no naturalness in the sound files themselves remained. This allowed for a representative analysis to take place.

[7] The word-initial [f] extracted from the word [feɪd]
[8] The word-initial [f] pasted onto the each of the 10 vowel gates

The experimental audio file: the forty stimuli were transferred into a single sound file that would present the stimuli during the experimental phase of Experiment 2, allowing participants to respond to and categorise the word-final consonants that they perceived hearing.

To create this file, all tokens were randomised and repeated three times. *Microsoft Excel* was used to randomise the stimuli and this *Excel* spreadsheet was then transferred to *SPLICE*[9] (version 060.pl), a *Perl script* used to code a continuous audio file. The resulting audio file was coded such that participants heard a *BLEEP* sound, followed by a 500ms *PAUSE*, after which the auditory stimulus was presented binaurally at a comfortable loudness. The visual presentation of the two sound choices appeared simultaneously with the auditory stimuli. Participants then had a response-window of 2500ms, after which there would be another 2500ms *PAUSE* before hearing the beginning of next trial, as signalled by another *BLEEP*.

The use of *SPLICE* ensured that the participants heard the audio files at a fixed speed with no variable breaks. Participants were unable to control the timing of the stimuli presentation, and this ensured quick and consistent processing and responses time across the participants' results. Randomising the stimuli three times also ensured that any ordering bias was avoided as much as possible. In total, one hundred and twenty stimuli were presented to each participant. These stimuli consisted of the four series of word and nonword options, differing according to the ten vowel gates, each randomised three times.

The *SPLICE* program then automatically recorded the participants' responses and reaction times and converted this data into a results file that could be copied and pasted into *JMP (SAS)* for subsequent analysis.

[9] This version of *SPLICE* was used throughout the thesis

Participants: twenty-six participants took part in Experiment 2. All participants were native speakers of English. Thirteen participants were female, and thirteen participants were male. Participants were between the ages of eighteen and sixty-seven. They had no known hearing or language disorders, and all had normal-to-corrected vision. One participant was left-handed. Participants could participate regardless of their spoken dialect of English, however the results from the participants were coded as either *Northern* or *Southern* in variety so that any differences in perception between these two dialect categories could be accounted for in the analysis.

Participants were recruited from the University of Oxford and did not receive any compensation for their time.

Procedure: as previously stated, Experiment 2 consisted of a forced choice identification task. A forced choice identification task was chosen as it was regarded to be the most appropriate method to efficiently measure the perception and categorisation of a single phoneme in the manner required for this study.

An information sheet was compiled specifying the full details of the study and what participation would involve. A consent form was also formulated, and written consent demonstrating a willingness to participate in this study was obtained from each of the subjects before they took part in the experiment. One consideration when drawing up these documents was the importance of not being overly explicit about the exact nature of the study. To avoid observer's bias, the notion that participants may behave differently in light of known experimental aims, only the generalised aims of this experiment were stated in all documents. These documents therefore indicated that this study was generally interested in speech

perception, and avoided mentioning anything more explicit regarding sound categorisation or the relationship between vowel duration, voicing, and lexical status[10].

Following ethics approval from Central University Research Ethics Committee (Ref No: R50069/RE001)[11], participants were recruited via email before being invited to attend a session of their choice in the Language and Brain Laboratory. Sessions lasted approximately thirty minutes and consisted of an initial briefing, the task itself, and then a short debrief after the experiment was over. During the initial briefing, participants were presented with the information sheet which, as previously stated, detailed the task itself and how the data that the participant provided would be anonymised before its use in this thesis and in any subsequent research. Participants were then given the opportunity to ask any questions that they had about the task before being asked to sign the written consent form. Within the consent form, participants were made aware of their right to withdraw from the experiment, and that they were able to stop the experiment at any time and choose to leave with no consequences or need for explanation.

Experiment 2 took place in a quiet room. Following the initial briefing session, participants were invited to sit in front of a 17" CRT monitor, attached to which was a button box labelled *D* and *T* and a pair of headphones. They were instructed to put on their headphones and hold the button box such that each of their thumbs aligned with one of the two buttons. Using both thumbs to make their selection increased the speed with which participants could react. Up to eight participants could participate at any one time, and partitions were positioned such that the participants could not see the responses of those around them. Participants were once again informed that a series of words and nonwords (pre-defined on the information sheet as *made up words*) would be played binaurally at a comfortable volume through the set of noise-

[10] A copy of the information sheet and consent form for Experiment 2 can be found in Appendix B
[11] A copy of the CUREC approval document can be found in Appendix A

cancelling headphones. They were instructed that following each word or nonword presented to them, they would have approximately two and a half seconds to choose between [d] or [t] according to which of the two sounds they perceived the word or nonword ending with. Participants were instructed that they must submit their choice by pressing the corresponding button on the box in front of them. The choices were visually presented on the screen in front of them. Left-handed participants held the box the opposite way up so that their dominant hand remained in control. Participants were asked to make a choice even in cases when they might not have been sure of their answer, and to respond as quickly as possible. All participants heard the stimuli in the same order, as they were each presented with the same *SPLICE* file. The experimental recording lasted approximately seven minutes. None of the participants taking part in Experiment 2 had been exposed to synthetically altered speech prior to this study, therefore they were each played a short example of the stimuli to get them familiarised with the format of the sound files that they were about to hear prior to the experiment beginning.

Following the completion of the task, participants were invited to attend a short debriefing session in which they could ask more detailed questions about the nature of the experiment and what this research aimed to achieve.

4.1.3 Analysis

The analysis of this experiment was conducted using *JMP (SAS)*. This analysis primarily considers the percentage of mean voiced responses recorded as the vowel gates get shorter, and whether a lexical effect is evident, i.e. participants' results had shifted in the direction of a word versus a nonword regardless of the duration of the vowel.

In total, three thousand, one hundred and twenty responses were recorded from the twenty-six participants. The data was cleaned such that all null responses (i.e. the participant had

selected neither the [d] nor [t] option) were excluded from the analysis. All reaction times that were more than two standard deviations away from the mean reaction time were also excluded. This analysis was conducted on a participant by participant basis. Additionally, any instances in which both the [d] and [t] options had been pressed simultaneously were excluded.

On the basis of the above criteria, one participant and six trials were excluded due to demonstrating less than 90% accuracy. Here, 90% was considered to be the appropriate level of accuracy as all participants were native speakers of English, and were provided with the full CV:?/CV? structure of the word-tokens. As such, a total of 13% of responses were excluded.

4.1.3.1 Acoustic input: the effect of vowel duration

Before analysing the stimuli grouped by individual word-initial consonants, let us first consider the overall responses across the vowel gates. Figure 27 (below) illustrates a steady shift from a [d] response to an [t] response as the vowel gates get consistently shorter. Recall that there is a 65ms durational difference across the ten vowel gates, with the longest vowel gate (gate 1) measuring 253ms, and the shortest vowel gate (gate 10) measuring 188ms. Therefore, in line with the predictions of this experiment, there is an overall steady decline in the percentage of mean [d] responses recorded as the vowel gates shorten.

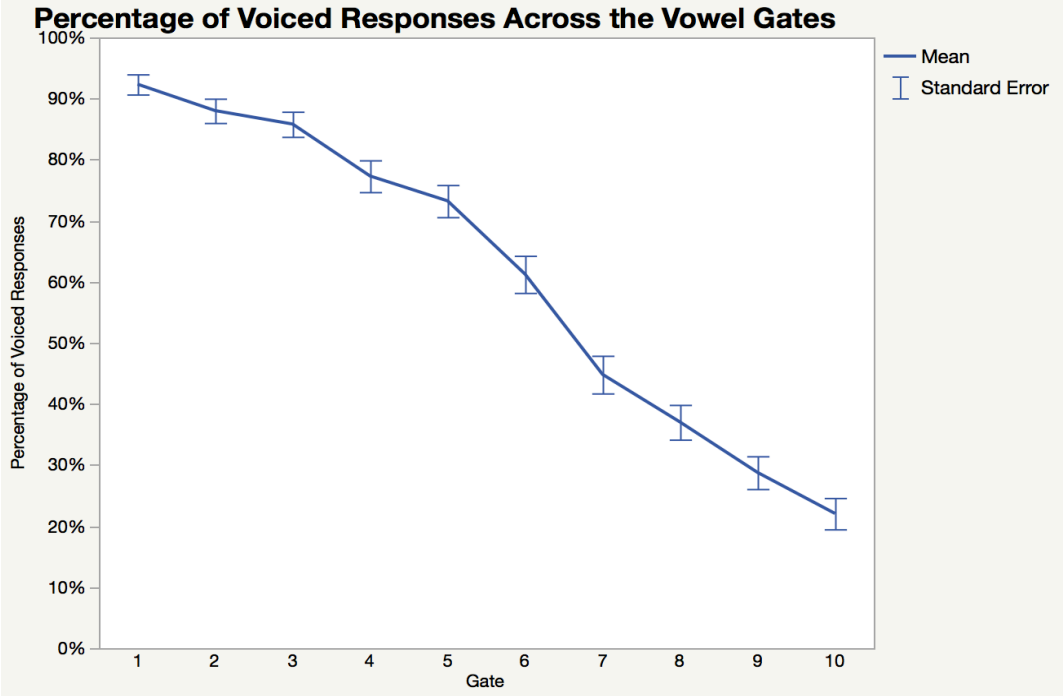


Figure 27: The percentage of voiced responses across the ten vowel gates for Experiment 2[12]

To further illustrate this shift from a [d] to a [t] response as the vowel gates get shorter, the percentage of [d] responses at each vowel gate can be found in Table 13 (below).

Table 13: The percentage of mean [d] responses across the ten vowel gates for Experiment 2

Gate	Percentage (%) of mean [d] responses
1	92
2	88
3	86
4	77
5	73
6	61
7	45
8	37
9	29
10	22

[12] SE is calculated as 1 standard error from the mean throughout Experiment 2

Figure 28 (below) illustrates this distribution on a histogram which makes it clear to see that participant responses are trending in the expected direction. These results are in support of the notion that voiced phonemes are consistently perceptually associated with longer vowel gates and conversely, voiceless phonemes are consistently associated with shorter vowel gates. Let us consider the end-points of the data. A one-way ANOVA (F (9, 2707) = 104.68, P < .0001) was conducted to measure the difference between the mean percentage of [d] responses across the vowel gates. This difference proved to be of high statistical significance.

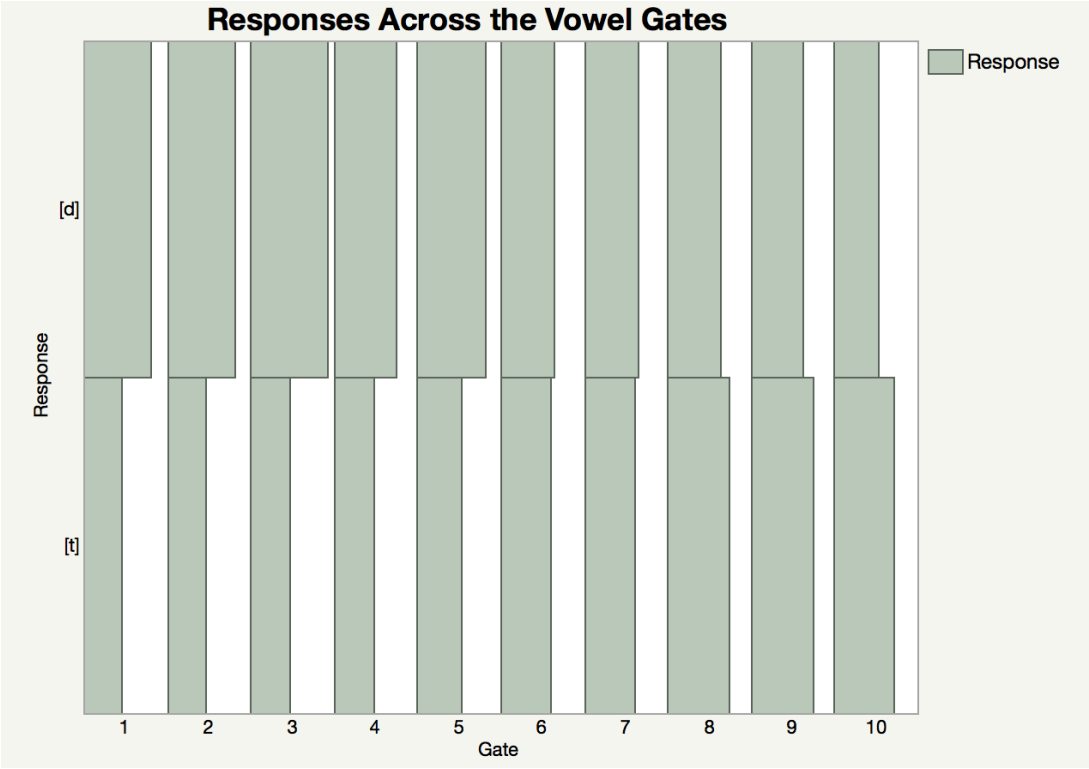


Figure 28: The [d]~[t] response distribution across the ten vowel gates for Experiment 2

These results illustrate the important finding that based on a durational difference in the vowel of 65ms; participants consistently perceive an ambiguous word-final stop consonant as voiced or voiceless according to the rules of English. This finding is significant in supporting the extent to which English speakers are sensitive to vocalic duration. These results provide

evidence in support of the remarkable effect that small durational differences in a preceding vowel segment can have on the perception of a neighbouring word-final stop.

4.1.3.2 The effect of the lexicon

Let us now consider these same results based on the information provided to the listener through the four word-initial consonants; [f], [z], [h], and [dʒ] to decipher how this may alter the pattern of results.

Table 14: The percentage mean [d] responses across the ten vowel gates based on word-initial consonant for Experiment 2[13]

Gate	[d] responses for word-initial consonant [f] *fade~fate*	[d] responses for word-initial consonant [z] **zade~*zate*	[d] responses for word- initial consonant [h] **hade~hate*	[d] responses for word-initial consonant [dʒ] *jade~*jate*
1	96	93	93	89
2	90	85	92	83
3	99	81	89	74
4	81	71	90	72
5	85	74	76	57
6	49	57	79	59
7	55	42	44	37
8	38	38	43	30
9	22	34	29	31
10	24	22	15	28

[13] Shaded gates denote largest shift in mean [d] response

Utilising the information encompassed in Table 14 (above), it is clear to see that for the word/word case, word-initial consonant [f], the largest shift in mean [d] responses happens between gates 5 and 6 (36%). This can be compared with the nonword/nonword case, word-initial consonant [z], in which the shift from a mean [d] response to a mean [t] response is again largest between gates 5 and 6 (17%). Looking now to the word/nonword case; word-initial [h], the largest shift occurs between gates 6-7 (35%) and in the nonword/word case, word-initial [dʒ], the largest shift is again demonstrated between the same gates (22%).

It was expected that in the case of word-initial consonants [f] and [z], a mean [d] response would shift to a mean [t] response towards the centre of the vowel gates. This is because both in the case of a [t] response and in the case of a [d] response, two words or two nonwords would be generated, [feɪd]~[feɪt] *fade~fate* or *[zeɪd]~*[zeɪt] **zade~*zate* respectively. A categorical shift in the middle of the vowel gates in which vowel duration was providing the participants with the primary voicing cue would therefore be expected. Conversely, in the case of word-initial consonant [h], participants were expected to shift to a mean [t] response significantly earlier in the vowel gate spectrum, due to the status of [heɪt] *hate* as a word as opposed to the nonword *[heɪd] **hade*. Conversely, in the case of [dʒ] it was expected that participant responses were likely to shift to a [t] response later, due to [dʒeɪd] *jade* being a real word as opposed to the nonword *[dʒeɪt] **jate*. However, according to Table 14 (above) and Figure 29 (below), the largest shift across all four consonants appears to cluster around the centre of the vowel gates. This supports the predictions outlined for word-initial [f] and [z], for which the largest shift from a mean [d] response occurs between gates 5 and 6. Equally, the indication that participants are shifting slightly later for word-initial [dʒ] is also supported. However, it would be expected that if the lexical bias were having a primary effect, participants would have shifted their response much later in the sequence of vowel gates. In addition, there is no

evidence that supports participants shifting to a mean [t] response earlier for word-initial [h]. Despite this, Figure 29 (below) illustrates that there are still clear end-points in which the longer vowel gates cue a higher percentage of [d] responses, and the shorter vowel gates cue a higher percentage of [t] responses, regardless of the word-initial consonants. These results therefore support the notion that the perception of word-final voicing is sensitive to vowel duration more-so than the influence of lexical bias.

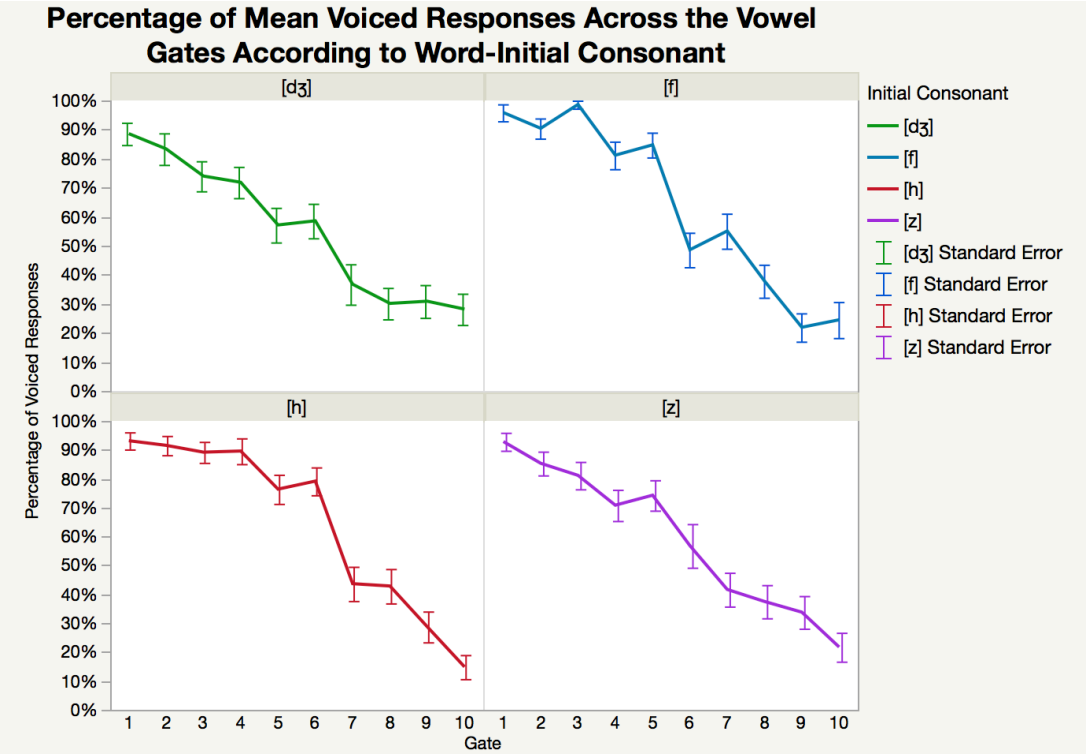


Figure 29: Cumulative frequency curves corresponding to the four word-initial consonants across the ten vowel gates for Experiment 2

In order to gain a fuller understanding of the reasons behind this perceptual behaviour, it is important to consider more carefully the relationship between the consonant pairs that were expected to behave similarly. The observations made in this section of the analysis, will be further explored in the subsequent discussion of Experiment 2. Let us first consider the case of word-initial consonants [f] and [z], the results from which can be found in Figure 30 (below).

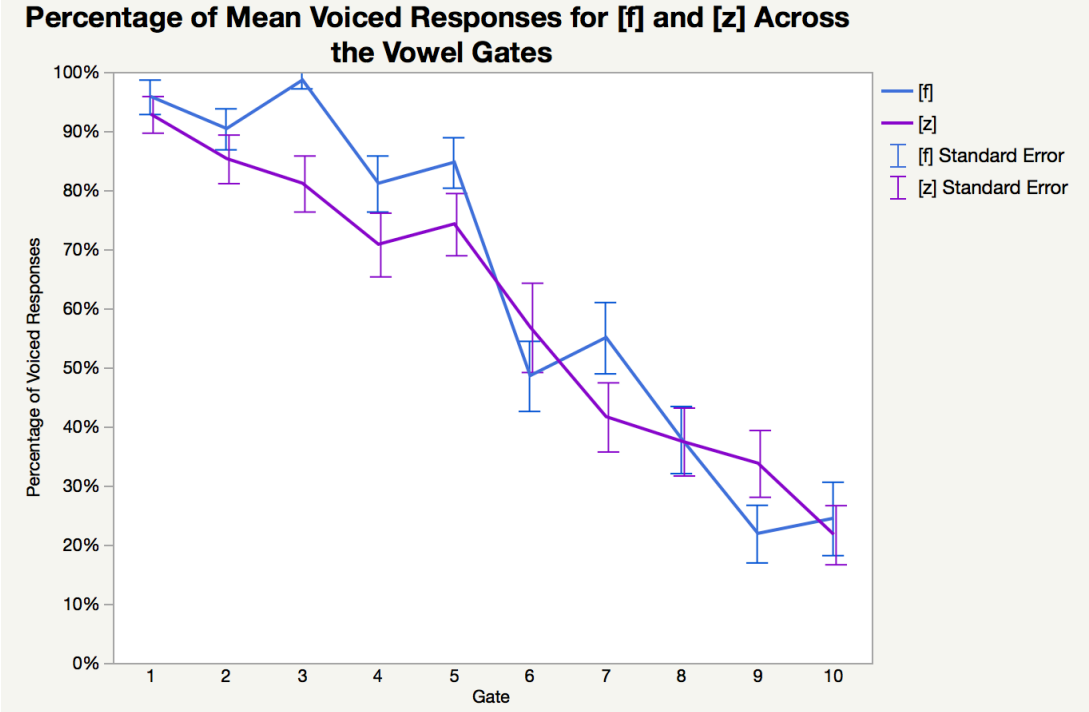


Figure 30: The mean [d] responses for the word- initial consonants [f] and [z] across the ten vowel gates for Experiment 2

The pattern of responses for word-initial [f] indicates a starker categorical shift between gates 5 and 6. This trend is in support of the predictions, in that participants are shifting their response around the midpoint of the vowel gates in the absence of lexical bias, and this perception is following a categorical rather than a continuous trend. A one-way ANOVA (F (9, 662) = 3.97, $P < .0001$) was conducted which demonstrated a highly significant difference between the mean [d] responses recorded across the gates. Conversely, the pattern of results for word-initial consonant [z] appear to indicate a more continuous pattern of perception, and despite the largest shift in responses also occurring between gates 5-6, the overall shift in responses is more gradual. Despite this, the differentiation of the mean [d] responses remain highly significant as the vowel gates get shorter (F (9, 674) = 23.14, $P < .0001$).

Secondly, the results comparing word-initial [dʒ] and [h] are depicted in Figure 31 (below). In the case of word-initial consonant [h], participant responses appear to support a categorical manner of perception, and the participants' responses appear to shift from a mean [d] response

to a mean [t] response between gates 6 and 7. However, though a categorical shift is present, the positioning of this shift is not as expected. Regardless of the lexical status of [heɪt] *hate* as opposed to the pseudoword *[heɪd] **hade*, it would appear participants are shifting to a [t] response later in the vowel gate spectrum than expected. The end-points of the stimuli beginning with word-initial [h] remain differentiated to a highly significant degree (F (9, 688) = 39.96, $P < .0001$).

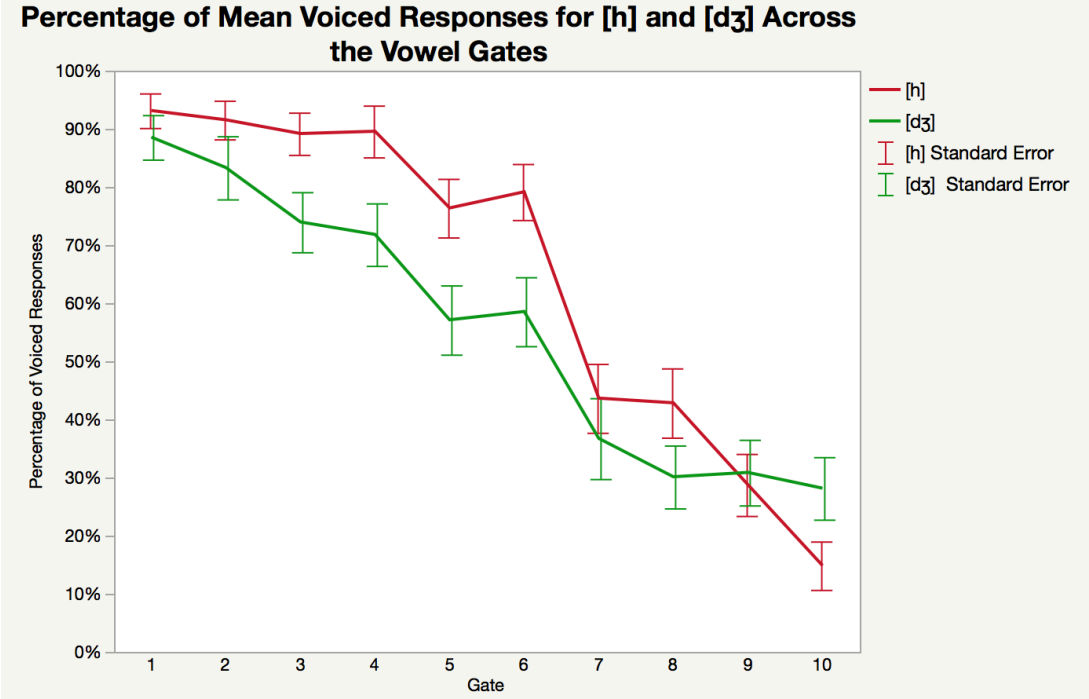


Figure 31: The mean [d] responses for the word- initial consonants [h] and [dʒ] across the ten vowel gates for Experiment 2

The participants' responses for stimuli beginning with word-initial consonant [dʒ] are reminiscent of the results obtained from word-initial consonant [z], outlined earlier. Table 14 (above) indicates that the greatest difference between the percentage of [d] responses occurs between gates 6-7. We would have expected this shift to have occurred later due to the lexical status of the word [dʒeɪd] *jade* as opposed to the pseudoword *[dʒeɪt] **jate*, and for this shift to demonstrate a categorical trend as opposed to the more continuous pattern illustrated in

Figure 31 (above). Despite this, and as with all other word-initial consonants, the mean [d] responses across the data are still differentiated to a highly significant degree (F (9,653) = 17.43, $P < .0001$).

In light of these results, it proves useful to compare word-initial consonants [f] and [h], along with [z] and [dʒ] respectively, as these are the pairs of consonants that have produced the most similar perceptual trends. Figure 32 (below) illustrates the responses acquired from the stimuli beginning with [f] and [h]. Both word-initial consonants appear to be patterning categorically, with the shift from a mean [d] to a mean [t] response occurring approximately one-gate apart from one another.

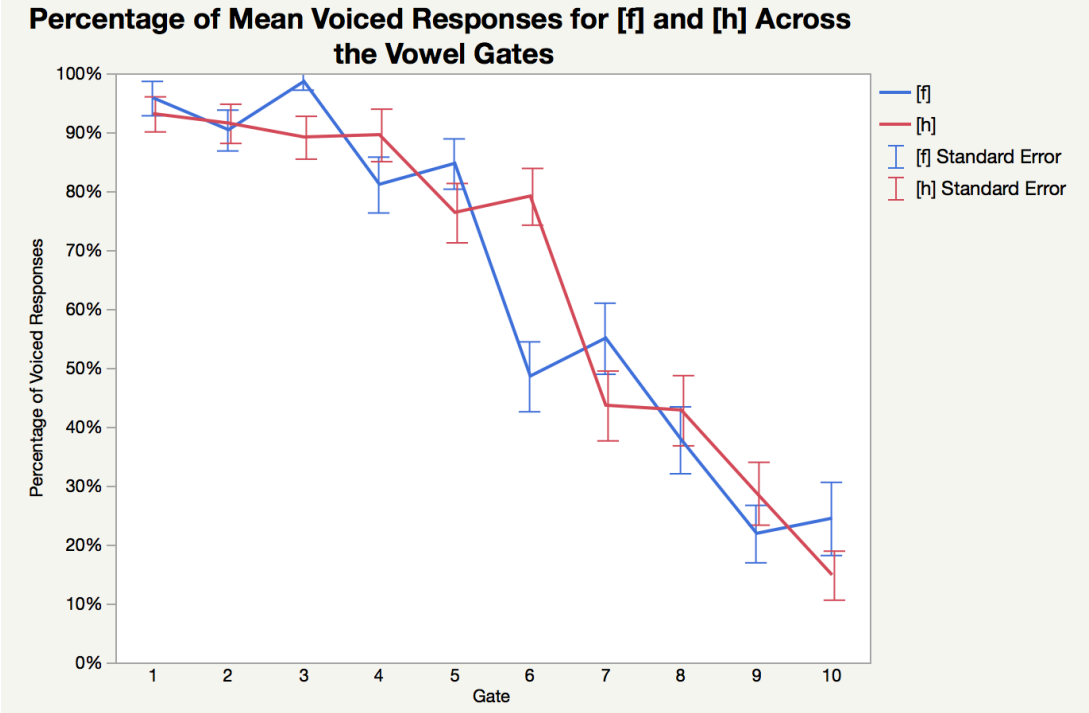


Figure 32: The mean [d] responses for the word- initial consonants [f] and [h] across the ten vowel gates for Experiment 2

Conversely, Figure 33 (below) illustrates the relationship between the results obtained from word-initial consonants [dʒ] and [z]. In contrast with word-initial consonants [f] and [h], the

word-initial consonants [dʒ] and [z] demonstrate a gradual perceptual shift from mean [d] to mean [t] responses.

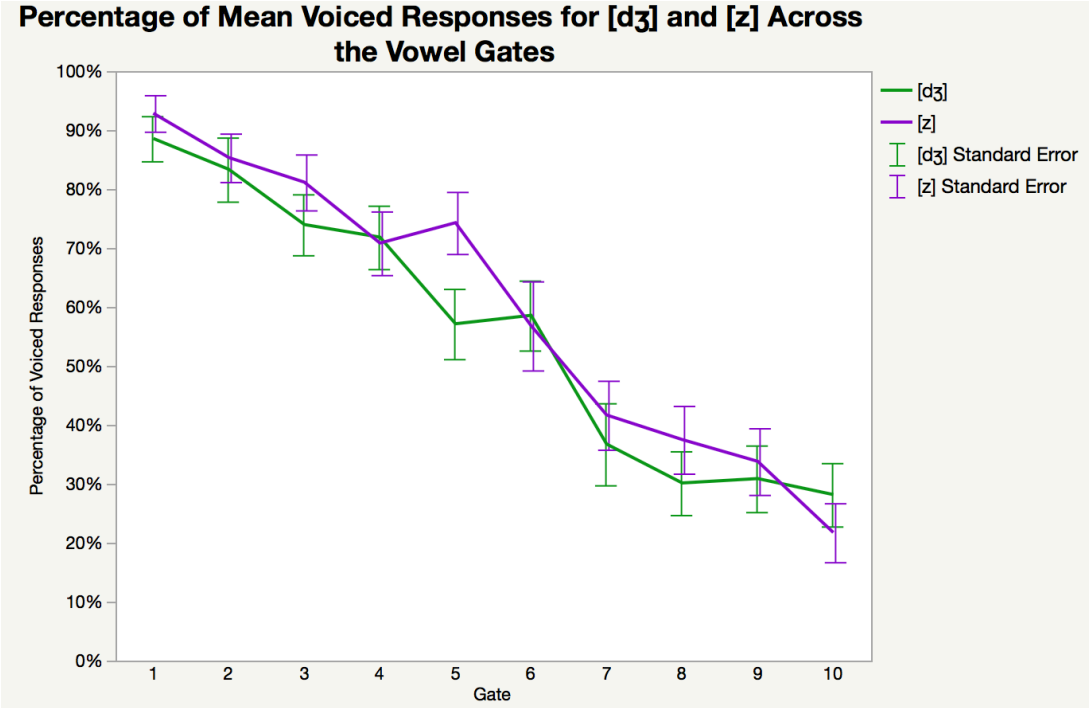


Figure 33: The mean [d] responses for the word- initial consonants [dʒ] and [z] across the ten vowel gates for Experiment 2

Looking at the data overall it would appear that vowel duration has had a primary effect on the responses given by participants, and lexical status appears to have had a secondary effect. The cumulative frequency curves suggest that vowel duration is extremely sensitive in cueing a word-final voicing distinction. Despite the data patterning according to categorical or continuous shifts in perception, it is not possible to say based upon this data whether a strict categorical or continuous model of perception is adhered to. This is because the identification task in Experiment 2 has not been paired with a discrimination task. In addition, the effect of the lexicon on the data appears variable. This study has produced some interesting observations and significant results regarding the perception of native English speakers, and their subsequent

ability to identify and categorise word-final stop consonants. The trends identified in this analysis will be discussed in (cf. Section 4.1.5 Discussion).

4.1.3.3 A note on dialect

Before advancing to further discuss the results presented in this section, it proves interesting to additionally present the data according to participant dialect. It is worth noting that participants could take part in this experiment regardless of their dialect, providing that they were a native speaker of British English. However, as previously mentioned in the methodology, the participants' results were also coded according to whether they had a Northern or Southern dialect. As dialect was not a controlled variable, there is a significant imbalance in the number of participant responses with a Northern versus a Southern dialect, with eight hundred and ninety-two Southern responses recorded, and only two hundred and eighty Northern responses recorded. However, in order to make some preliminary inferences about dialect, Figure 34 (below) depicts these results in order to identify any obvious differences in perception according to the dialect of English spoken by the participants.

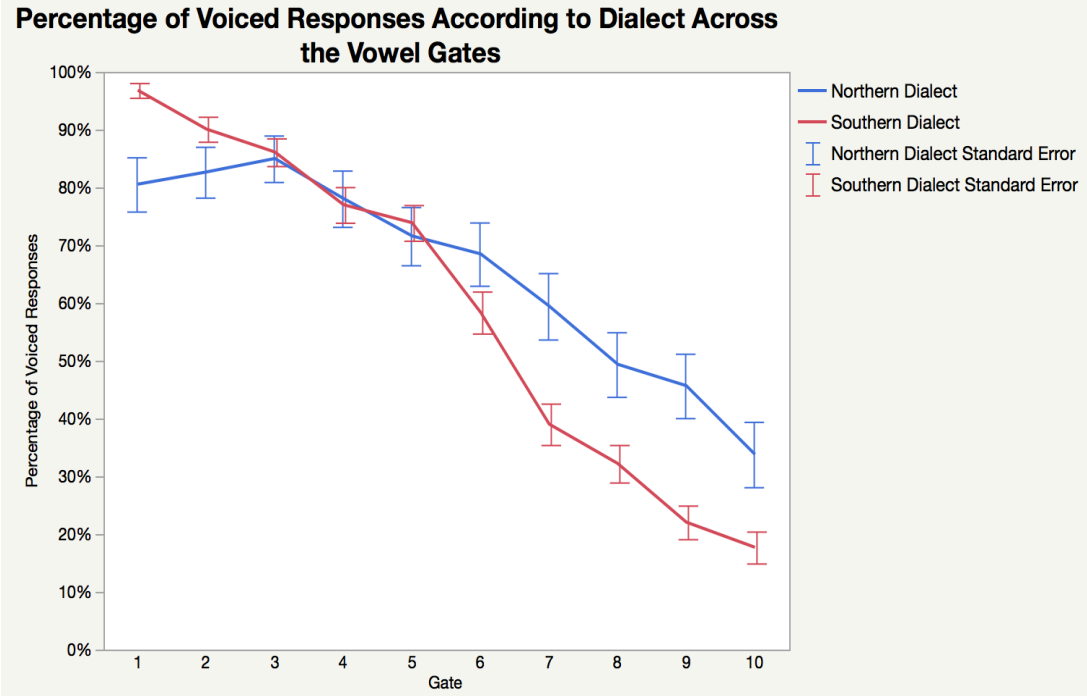


Figure 34: The mean [d] responses for the participants' dialects across the ten vowel gates for Experiment 2

From this very preliminary data, it would appear that participants who have a Southern dialect of English behave more categorically with regard to their perception than those who speak with a Northern dialect of English. However, the difference in mean [d] responses between gates 1 and 10 are differentiated to high significance for both the Southern participants (F (9, 1946) = 105.03, P < .0001) and the Northern Participants (F (9, 751) = 11.66, P < .0001). Therefore, the data presented in Figure 34 (above) is likely related to there being fewer Northern speakers overall. However, participants with a Southern dialect may have responded more categorically to the stimuli due to the speaker whose recordings were used to form the basis of the stimuli also being a speaker of Southern British English.

The pronunciation of the [eɪ] vowel nuclei also varies across different varieties of English, for example between Standard British English and Yorkshire dialects. Therefore, the participants' sensitivity to vowel lengthening may vary depending the dialect that they speak. Though this should be unlikely due to the underlying representation of vowel duration in

English phonology, these differences are interesting to note. Future studies may benefit from controlling dialect and measuring perception across several different dialects, or varieties of English. All subsequent experiments within this thesis therefore controlled for participant dialect.

4.1.4 Discussion

At the end-points of the data, there is consistently a statistically significant difference between the mean number of [d] responses recorded at the longest and shortest gates. Therefore, results of this study indicate that manipulating vowel duration has the capacity to cue a native English speaker to perceive voiced and voiceless word-final sounds accordingly. This observation is supported by several previous findings from the literature. Recall that Klatt (1976) stated that, in English, vowel duration is considered to be one of the primary variables utilised by listeners when interpreting auditory signals.

The findings from Experiment 2 are particularly remarkable as they signify that in the absence of any acoustic information provided by the word-final consonant, and when the closure duration, VOT, and all aspects of the vowel's production are held constant, vowel duration has the capacity to influence listener's perception in identifying and categorising an ambiguous word-final sound as correspondingly voiced or voiceless. This shift in perception proves significant as, within this experiment, it occurs across a 65ms durational difference, thereby highlighting the salience of this perceptual cue in English.

Turning now to the role of the lexicon, let us first consider the results obtained from the word-initial consonants that were not intended to cue a lexical bias. Stimuli beginning with word-initial consonants [f] and [z] were designed to create two words or two nonwords respectively. In the case of stimuli beginning with word-initial consonant [f] ([feɪd]~[feɪt]

fade~fate), the results patterned most similarly to the predictions laid out earlier with a shift occurring categorically around the centre of the vowel gates. However, this categorical shift was less prevalent in the case of stimuli beginning with word-initial consonant [z], from which two nonwords (*[zeɪd]~*[zeɪt] **zade~*zate*) were created. Here, a more continuous shift from a [d] to a [t] response was demonstrated. However, this more gradual shift in the perception of voicing did occur towards the centre of the vowel gates as expected. The more gradual nature of this shift is likely to be due to the lack of activation in the mental lexicon. As such, it seems that the participants were relying primarily on the acoustic information provided by the varying vowel durations. This would account for the central positioning of the shift from a mean [d] response to a mean [t] response, and the gradual lengthening of the vowel gates resulting in the continuous pattern of responses.

Let us now consider the results for word-initial consonants [h] and [dʒ]. In both cases, a lexical effect was expected to occur in the direction of the word options [heɪt] *hate* and [dʒeɪd] *jade*, as opposed to the nonword options *[heɪd] **hade* and *[dʒeɪt] **jate*. However, instead of the results indicating a categorical shift in perception based on these word boundaries, the results continued to cluster around the midpoint of the vowel duration spectrum with the largest shifts in responses being recorded between gates 6-7 in both instances. The responses for word-initial [h] behaved in a categorical manner, whereas the shift in responses for word-initial [dʒ] was more continuous in nature.

Interestingly, the [h] responses demonstrated a trend in the opposite direction of that which was expected, with a shift occurring later in the spectrum rather than earlier. One possible reason for this could be that the acoustic cue provided by the original [d] base of the vowel gates remained underlyingly prominent, and therefore the participants are not trending towards *[heɪt] **hate* as clearly as one would expect. An important consideration for further research

therefore would be to conduct a subsequent experiment using a base-line [t] phoneme to formulate the vowel durations. As previously mentioned, this method proves more complex in creating naturally sounding stimuli, as adding glottal pulses to an existing vowel length is more intricate than removing existing glottal pulses. However, to achieve this in future studies would provide this investigation with an interesting comparison.

Regarding word-initial consonant [dʒ], it would appear that participants are not being influenced strongly enough by the word-initial consonant to cue a categorical shift at all. It seems that participants treated the [dʒeɪd] *jade* option as a nonword, and this is likely to be due to its lower frequency of use within the English language.

Ganong's (1980) work on categorical perception and the role of the lexicon is among the most significant conducted in this area of study. The results of Experiment 2 do not support the findings made by Ganong (1980) to the extent that was expected earlier, in that the lexicon does not appear to be overriding the acoustic information provided to the listener by means of the vowel duration. Though a strict model of categorical nor continuous perception can be supported by the findings from Experiment 2, due to no discrimination data, a shift in the direction of a word as opposed to a nonword has not been realised to a great extent within the results. However, this absence of a Ganong Effect could be due to some methodological considerations. Specifically, the participants were asked to select between two phonemes, [d] and [t], as opposed to two whole word or nonword options. Ganong's (1980) study also dealt with word-initial as opposed to word-final sounds, thereby enabling the participants in his study to access their lexicon at the beginning as opposed to at the end of the stimuli. These methodological differences are likely to have affected the probability of obtaining a clear Ganong Effect within the results of this study; though the data does not demonstrate a strong

effect of lexical bias regardless. Instead, the role of the lexicon in Experiment 2 appears to be dependent largely upon factors such as word frequency and methodological considerations.

Most notably, the results from Experiment 2 demonstrate that listeners are able to use the information from vowel duration to categorise an ambiguous word-final phoneme, perceiving it as voiced or voiceless according to the rules of English phonology.

4.1.5 Conclusion

Based on the results of this study, it would appear that across a durational difference of 65ms, native speakers of English are able to categorise word-final sounds as being voiced or voiceless according to the rules of their language. In manipulating the vowel duration, the perception of the ambiguous word-final stop was altered, with longer vowel durations cuing listeners to perceive a voiced [d] phoneme, and shorter vowel durations cuing listeners to perceive a voiceless [t] phoneme. This clear shift in perception is particularly remarkable considering that the vowel duration changes between the gates were minimal.

Incorporating the role of the lexicon into this experiment has also provided evidence that the perception of voicing can be offset somewhat according to word and nonword boundaries. However, the results of Experiment 2 suggest that the lexicon does not have a consistent ability to override the acoustic information provided by the vowel duration. The extent to which lexical bias appears able to affect perception seems to be dependent on factors such as word frequency and the methodology used to form the stimuli. Conducting further investigations which control for word-frequency or consider the classification of whole words and nonwords as opposed to individual word-final phonemes may have the ability to further our understanding of this variable.

The results from Experiment 2 suggest that in the absence of clear lexical representation, the perception of word-final voicing occurs according to a continuous pattern of perception. This was the case for both the pair of nonwords (*[zeɪd]~*[zeɪt] **zade*~**zate*) and also ([dʒeɪd]~ *[dʒeɪt] *jade*~**jate*), in which [dʒeɪd] *jade* appears to have been treated as a nonword. In cases where two words are activated, or there is a clear word/nonword distinction, as in the cases of [feɪd] *fade* and [feɪt] *fate* and also *[heɪd] **hade* and [heɪt] *hate*, a more categorical trend of perception is demonstrated which does appear to move somewhat in the direction of a word versus a nonword. This is likely to be due to a more prominent lexical representation.

Experiment 2 overall supports the notion that vowel duration acts as a primary cue for word-final voicing in English. Experiment 3 aims to build upon this study and provide more information about the extent to which this finding proves true in the absence of word-final acoustic information.

4.2 Experiment 3: predicting an original [d]~[t] word-finally

Experiment 3 consisted of a forced choice identification task. It did not consider the role of the lexicon, and in doing so focused purely on the extent to which vowel duration can be demonstrated to cue the perception of word-final voicing in English. As opposed to forming stimuli which incorporate a word-final ambiguous consonant, as in Experiment 2, Experiment 3 omitted the word-final consonant altogether. This study therefore provided participants with the CV:/CV fragment of a CVC word, requesting that they indicate between a minimal-pair which of the two sounds the stimuli *should have* ended in. By comparing participants' responses with the voicing on the original recording, the extent to which participants can accurately predict the voicing of a word-final consonant on the basis of just the CV:/CV fragment is considered.

4.2.1 Research questions

Experiment 3 is interested in determining how vowel duration influences participants' prediction of word-final voicing when the word-final consonant is omitted from a CV:C/CVC sequence.

Based on the results of the Experiment 2, it was expected that stimuli containing comparatively longer vowel durations would cue listeners to predict a voiced word-final consonant. Conversely, stimuli containing comparatively shorter vowel durations were expected to cue listeners to predict a voiceless word-final consonant.

4.2.2 Methodology

The stimuli: Experiment 3 utilised the same recordings made for Exp1a, and this list of minimal pairs can be found in Table 15 (below). Word frequency was not controlled for as Experiment 3 was not concerned explicitly with the additional influence of the lexicon on speech perception.

Table 15: The list of word-tokens used for Experiment 3[14]

English minimal pairs /d/~/t/	**IPA**	
cad~cat	*/kæd/*	*/kæt/*
mad~mat	/mæd/	/mæt/
mode~moat	/məʊd/	/məʊt/
made~mate	/meɪd/	/meɪt/

Recall that these tokens were selected as they were consistent; each began with the same word-initial consonant, [m], and the tokens ended in an equal number of voiced and voiceless word-final consonants [d] and [t]. The vowel nuclei were also employed because they accounted for a maximal amount of the vowel space, whilst still allowing a series of minimal pairs to be generated in English. All word-tokens were monosyllabic.

Recall also that the three male speakers were recorded at the Language and Brain Laboratory at the University of Oxford. Each of the speakers read out the words in Table 15 (above), working their way down the list from top to bottom, and then repeating this process five times. This generated five instances of each of the word-tokens. The speakers were requested to regulate the speed of their speech, volume, and intonation as much as possible. The fillers gave the speakers the opportunity to regulate their intonation and volume at the beginning and end

[14] Words found in italics are fillers

of the recording. The speakers were not made aware of what this study was investigating, or of which words were going to be included or excluded from the analysis.

Aside from the word-list, the recording materials and computational programs used to capture and edit the recordings remain identical to those used in Exp1a.

The recordings from Exp1a were edited using *Praat*. The word-final consonants were omitted from each word-token. Here, both the closure duration and release burst were identified using visual and auditory changes in the spectrograms, with the movement of energy transitioning from the vowel and into the period of silence marking the beginning of the closure duration. This process is illustrated in Figures 35-37 (below). Ninety stimuli were created in this way; six words, repeated five times, by three speakers. Each of these stimuli contained a naturally occurring vowel duration as elicited in Exp1a. Using this method, the full vowel was retained in the CV:/CV portion of each stimulus. As with Experiment 2, it is important to acknowledge the possibility that perceptual cues for voicing may have been retained within the spectral cues in the latter part of the vowel, as well as in the word-initial consonant. Here, it is important to reiterate that this thesis does not deny the existence or potential bias provided by these cues. However, despite this complexity, Experiment 3 endeavoured to present participants with a range of naturally occurring vowel durations as opposed to the gated studies conducted in Experiments 2 and 4. As such, retaining the full vowel was crucial.

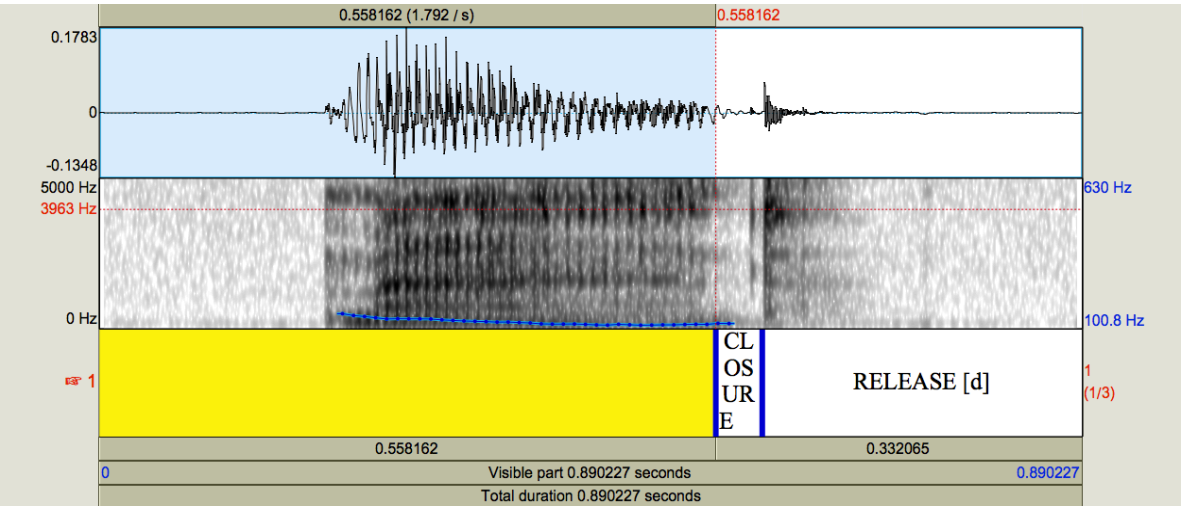


Figure 35: The use of *Praat* to place zero-crossing boundaries marking the closure duration and release burst of the word-final consonant for Experiment 3[15]

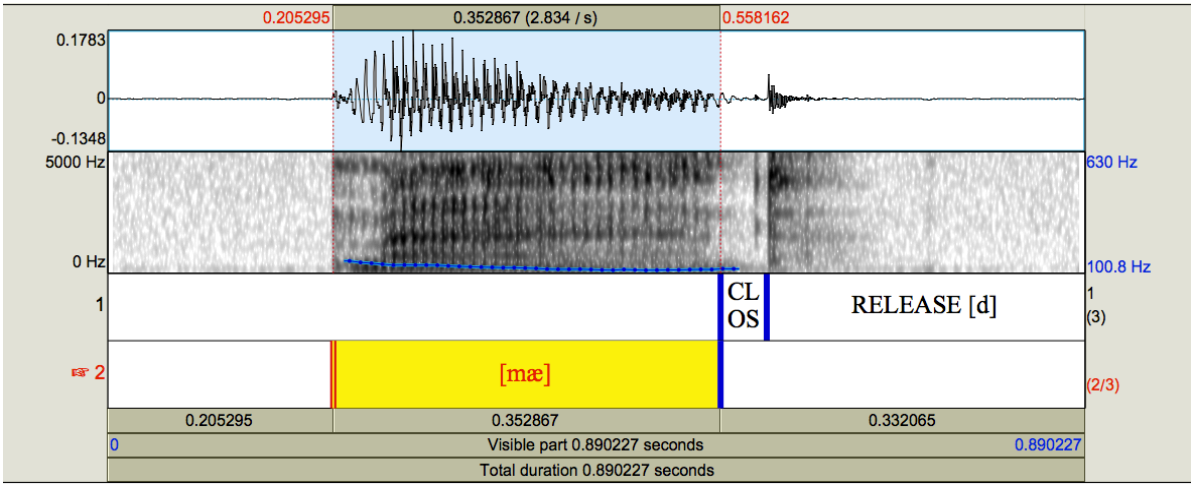


Figure 36: The use of *Praat* to place zero-crossing boundaries marking the CV portion of the recording for Experiment 3[16]

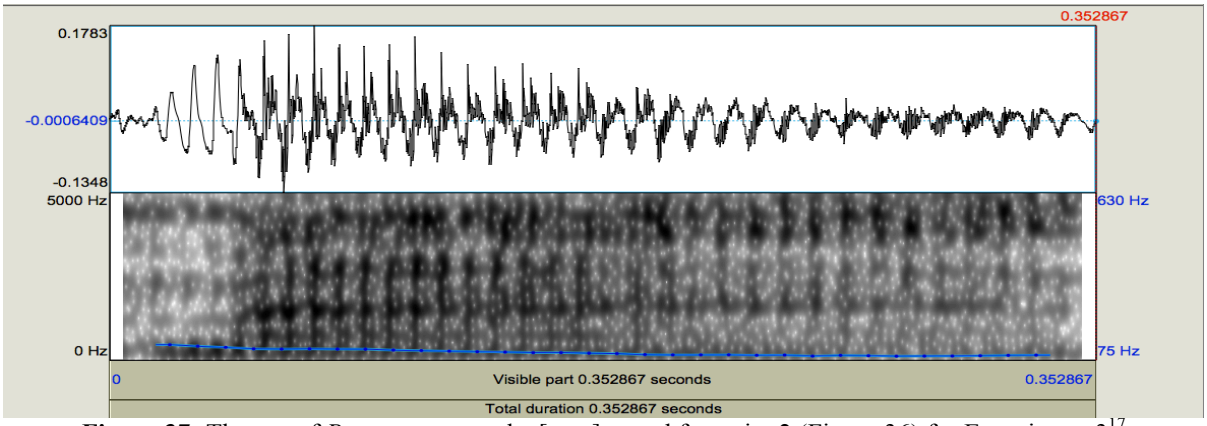


Figure 37: The use of *Praat* to extract the [mae] sound from tier 2 (Figure 36) for Experiment 3[17]

[15] Word-token: *mad* [mæd], spoken by Speaker A
[16] Word-token: *mad* [mæd], spoken by Speaker A ([mae] was then labelled in Tier 1)
[17] Word-token: *mad* [mæd], spoken by Speaker A

The experimental audio file: *Microsoft Excel* was used to order and randomise the stimuli. This process created a total of two hundred and seventy stimuli; six words repeated five times by three speakers, then randomised three times. Randomising the stimuli ensured that any potential ordering bias was avoided as much as possible. As with Experiment 2, *SPLICE* was utilised to generate an audio file that allowed the stimuli to be presented to participants at a fixed speed with consistent ordering, and recorded the participants' responses automatically. Again, participants were unable to control the timing of the stimuli presentation, and this ensured quick and consistent processing and accurate reaction time measurements across the participants' responses. The experimental audio file began with a *BLEEP* sound, followed by a 500ms *PAUSE*, after which the auditory stimulus was presented. The visual presentation of the two sound choices appeared simultaneously with the auditory stimuli. Participants had a window of 2500ms to respond, after which they would hear another 2500ms *PAUSE* before the beginning of the next trial, as signalled by another *BLEEP*.

Participants: twenty-one participants took part in Experiment 3. All participants were native speakers of Southern British English. Fifteen participants were female, and six participants were male. Participants were between the ages of eighteen and sixty. They had no known hearing or language disorders, and all had normal-to-corrected vision. None of the participants were left-handed. Participants were recruited via email from the University of Oxford and did not receive any compensation for their time.

Procedure: the experimental procedure for Experiment 3 was largely based on the procedure for Experiment 2. Participants were invited to attend an experimental session of their choice at the Language and Brain Laboratory. Each session lasted for around thirty minutes,

inclusive of the time taken to fill in paperwork and ask questions both before and after the study. Upon arrival, participants were provided with an information sheet, detailing the study[18]. Written consent was also obtained from all participants.

As with Experiment 2, participants were requested to sit in front of a 17" CRT monitor. They were required put on a pair of headphones through which the auditory stimuli would be played binaurally at a comfortable volume. Participants were asked to hold a two-button response box, the left button corresponding to a [t] response and the right button corresponding to a [d] response. This choice was indicated orthographically on the monitor in front of them as *T or D*. Participants were asked to select the button corresponding to the sound that they felt the stimuli *should end in*. They were advised that they would have approximately two and a half seconds to respond. Up to eight participants could take part in any one session, and partitions were positioned such that participants could not see the responses of those around them. Participants were instructed to respond as quickly as possible, and that once their choice had been made they could not change their decision. Each participant heard the stimuli in the same order. The experimental recording lasted for sixteen minutes, after which participants were given the opportunity to ask questions. As the experiment had been completed at this stage, participants were given permission to ask more specific questions about the nature of the experiment.

4.2.3 Analysis

The analysis of this experiment was conducted using *JMP (SAS)*. Given that participants had only heard the CV:/CV components of a CVC word, this analysis was interested in determining whether the acoustic information provided by the initial consonant and subsequent

[18] A copy of the information sheet and consent form for Experiment 3 can be found in Appendix C

vowel would provide enough information to elicit a *correctly matched* word-final response in line with the nature of voicing present on the original recording.

In total, five thousand, six hundred and seventy responses were recorded from the twenty-one participants. The first stage of the analysis was therefore to clean the results such that all statistical outliers and null responses were excluded. This was conducted in the same manner as Experiment 2.

As such, one participant was excluded due to demonstrating less than 90% accuracy. Eight trials were also excluded for failing to demonstrate 80% accuracy overall. Here, 90% was considered to be the appropriate level of accuracy for the participants as all participants were native speakers of English. However, 80% was considered to be the most appropriate level of accuracy for the trials as participants were only presented with the CV:/CV fragment of a full CVC word. As such, a total of 12% of responses were excluded. This was in line with the percentage of excluded responses for Experiment 2.

The results from Experiment 3 will now be presented, looking firstly at the overall results before turning to analyse the findings from the different vowel nuclei. Given that there were an unequal number of male and female participants, an analysis will also be conducted according to a subset of participants who were controlled for sex. This will allow any inferences that can be attributed to this sociolinguistic factor to be identified.

4.2.3.1 Overall results

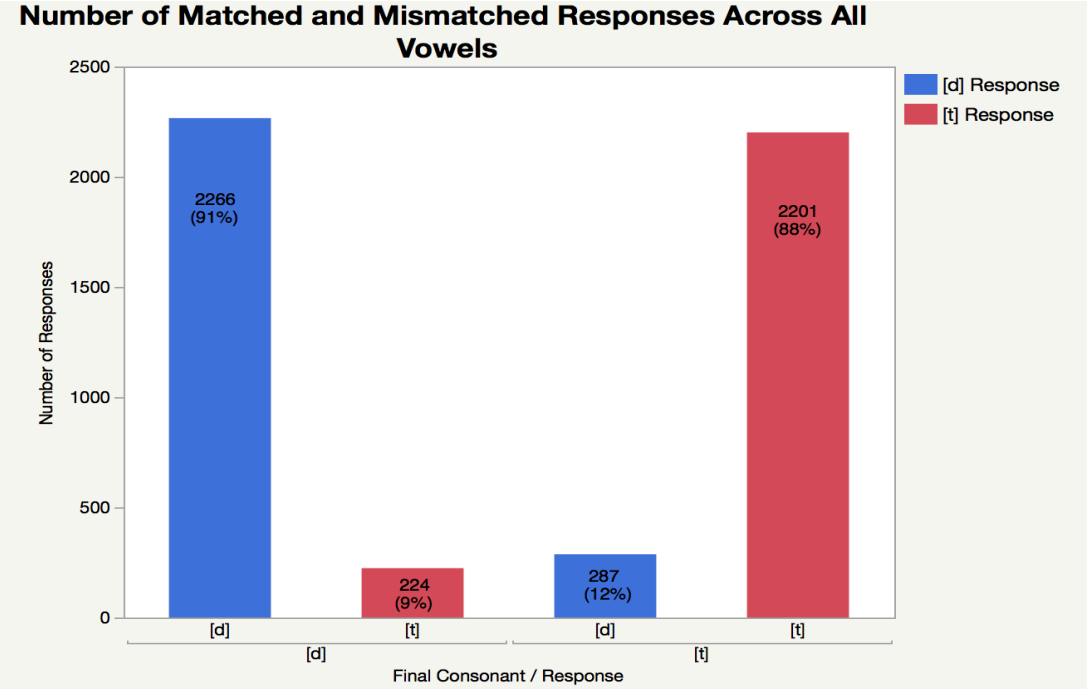


Figure 38: The number of correctly matched and incorrectly mismatched responses across all word-tokens for Experiment 3

Before we divide the results according to the three different vowel nuclei, let us first consider the overall results across all word-tokens (Figure 38, above). In the case of the word-final voiced [d], participants correctly matched their responses in 91% of cases. Likewise, for the voiceless word-final stop [t], participants correctly matched their responses in 88% of cases, using purely the CV:/CV fragment of the sequence.

To determine whether this relationship between the original *final consonant* present on the recording and the participants' *response* was significant, a contingency table was generated in *JMP*, and a *Fisher's Exact Test* was conducted. This was judged to be the most appropriate measure to determine the probability that the association between the two nominal variables *final consonant* and *response* was not random. The results can be found in Table 16 (below).

Table 16: Fisher's Exact Test Results (Final Consonant and Response) for Experiment 3

Fisher's Exact Test	Probability	Significance	Alternative hypothesis
Left	1.0000		Prob(Final Consonant=[t]) is greater for Response=[d] than [t]
Right	<.0001	*	Prob(Final Consonant=[t]) is greater for Response=[t] than [d]
2-Tail	<.0001	*	Prob(Final Consonant=[t]) is different across Response

Looking at Table 16 (above) we see the *P*-values listed for the null hypothesis, along with the alternative hypotheses. In this case, the null hypothesis is that there is no correlation between the original *final consonant* and the *response*. The left-handed hypothesis demonstrates that the probability of participants' complete inability to predict a [d] response when the final consonant was a [t] is $P = 1$. This is not intended to indicate certainty, but rather that the probability is extremely low. Conversely, the right-hand hypothesis demonstrates a highly significant probability ($P < .0001$) that participants are guided to perceive a [t] response when the original final consonant was a [t]. This is supported by the two-tail hypothesis which also shows a highly significant probability ($P < .0001$) that the difference between the [d] and [t] responses is dependent on the category of the original final consonant. The *p*-value given by the *Chi-Squared* test is also highly significant ($P < .0001$) further indicating that the *final consonant* and *response* are variables that are dependent on one another.

4.2.3.2 Results according to vowel nuclei

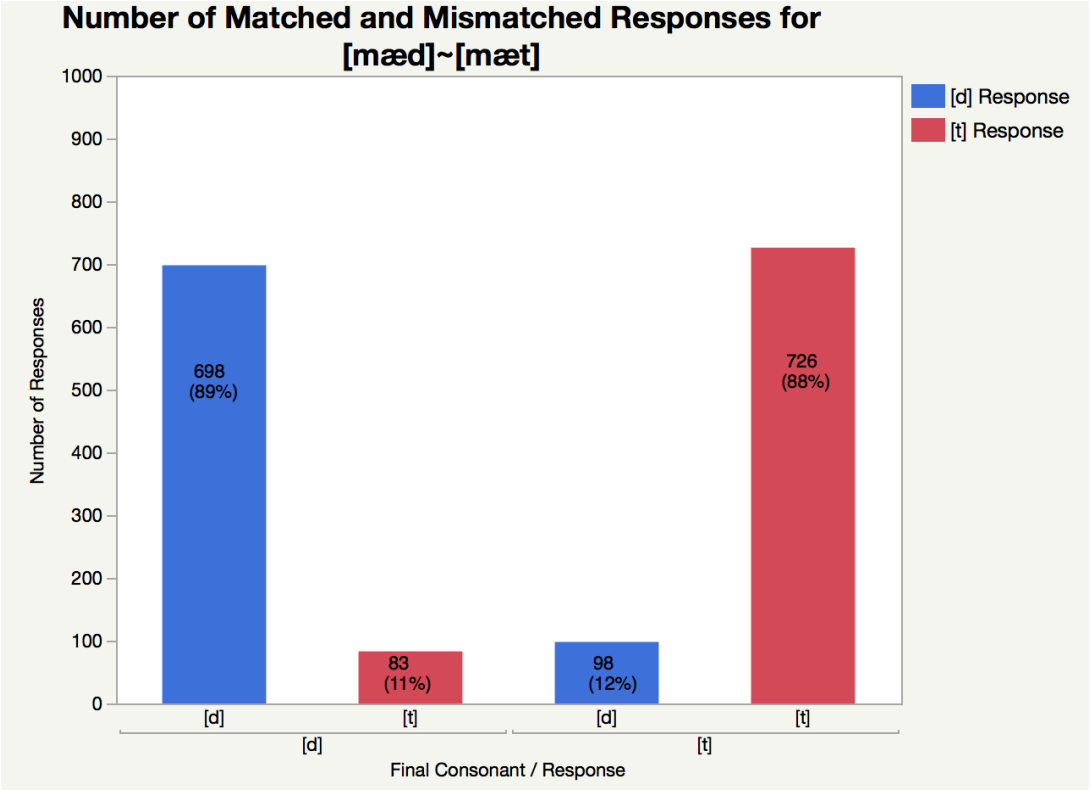


Figure 39: The number of correctly matched and incorrectly mismatched responses for [mæd]~[mæt] *mad~mat* for Experiment 3

Let us now consider participant responses for the word-tokens [mæd]~[mæt] *mad~mat* (Figure 39, above). Here, participants correctly identified the correct response in 89% of cases for the voiced [d] word-final ending, and 88% for the voiceless [t] word-final ending.

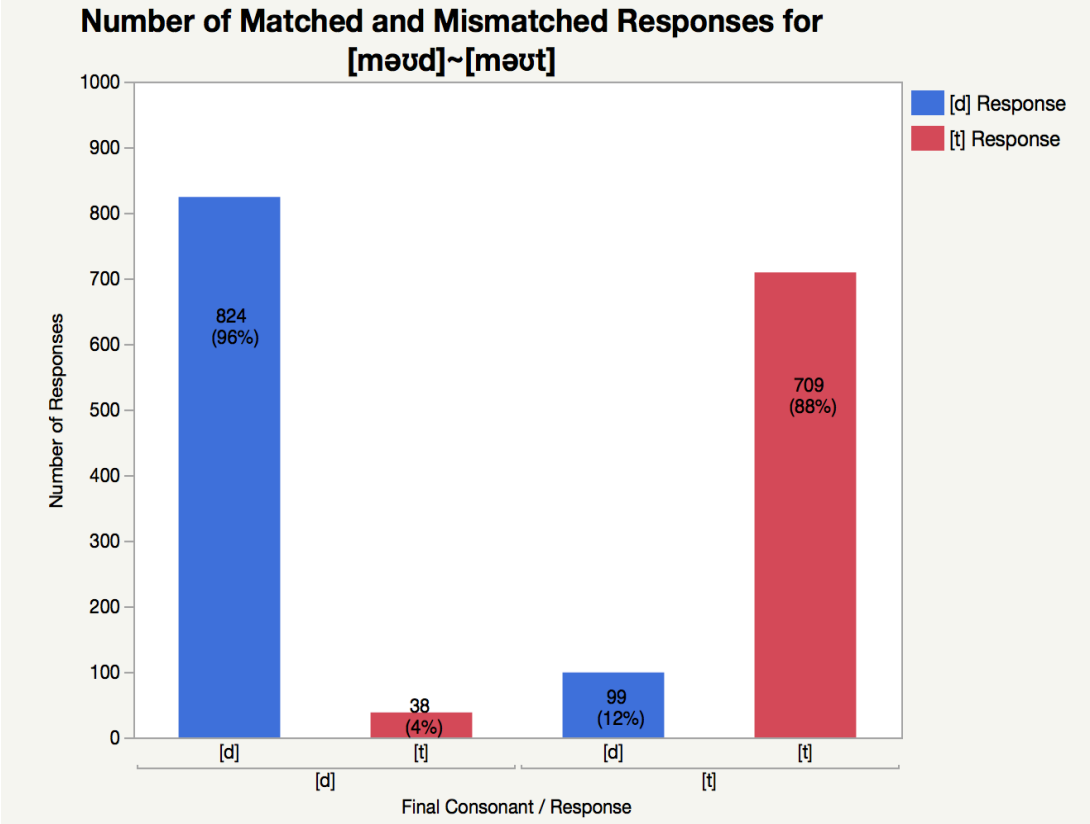


Figure 40: The number of correctly matched and incorrectly mismatched responses for [məʊd]~[məʊt] *mode~moat* for Experiment 3

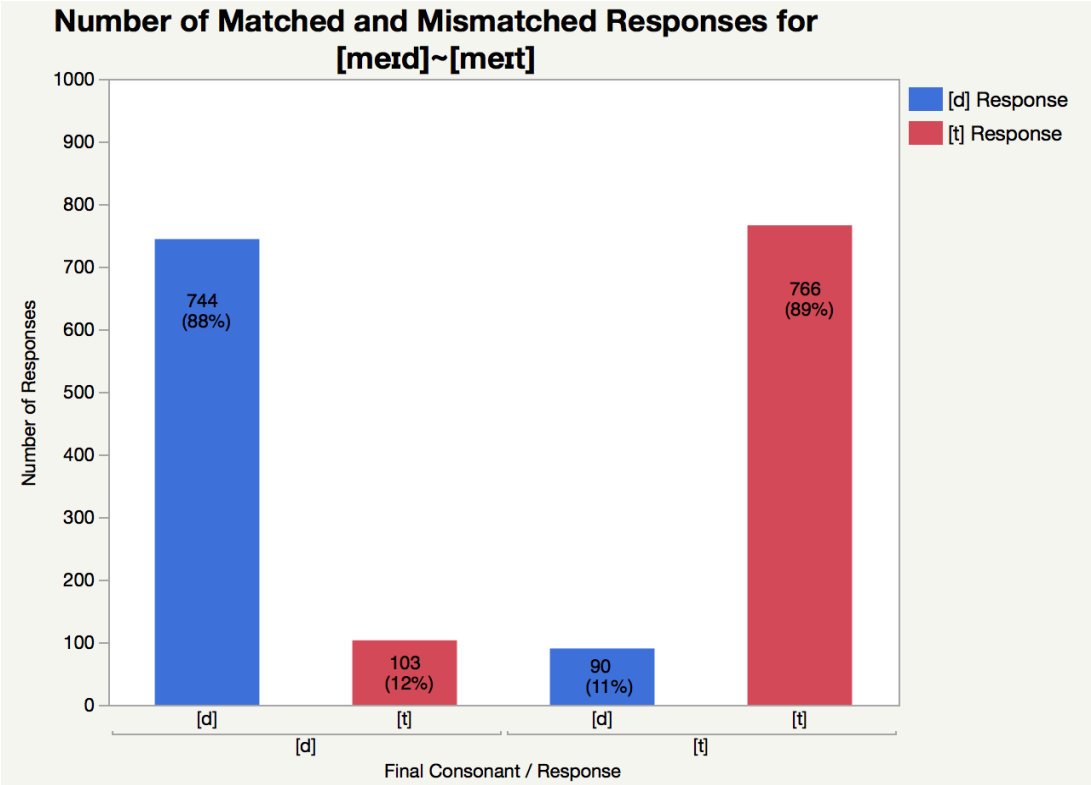


Figure 41: The number of correctly matched and incorrectly mismatched responses for [meɪd]~[meɪt] *made~mate* for Experiment 3

In the case of [məʊd]~[məʊt] *mode~moat,* it was found that participants correctly predicted a voiced response in 96% of cases for the voiced option [məʊd] *mode*, and 88% of cases for the voiceless option [məʊt] *moat* (Figure 40, above). Finally, for [meɪd]~[meɪt] *made~mate* (Figure 41, above), participants correctly matched 88% of responses for voiced [d] word-final stop, and 89% of responses for voiceless [t] word-final stop.

A *Fisher's Exact Test* was conducted for each of these three individual vowel nuclei. The results prove consistent (Table 17, below) and demonstrate the same outcome as was previously recorded in Table 16 (above). These results support the causal behaviour of the participants, and establish that participants are being guided to respond on the basis of the original [VOICE] feature present on the recording. The results of the *Chi-Squared* test also remained $P < .0001$. The observed data therefore fits the expected data extremely well.

Table 17: Fisher's Exact Test Results (Final Consonant and Response) for the three word pairs for Experiment 3

Fisher's Exact Test	Probability	Significance	Alternative hypothesis
Left	1.0000		Prob(Final Consonant=[t]) is greater for Response=[d] than [t]
Right	<.0001	*	Prob(Final Consonant=[t]) is greater for Response=[t] than [d]
2-Tail	<.0001	*	Prob(Final Consonant=[t]) is different across Response

4.2.3.3 Results according to sex

Experiment 3 consisted of fifteen females and six males. Given this imbalance, this next section of the analysis focuses on a cross-section of five females and five males. These ten participants were randomly selected using the random number generator in *Microsoft Excel*. The purpose of this analysis is to ensure that no bias was caused by the sex of the participants,

and that the overall trend of results remains the same. This analysis is also important as it allows any sociolinguistic differences related to sex to be identified and accounted for.

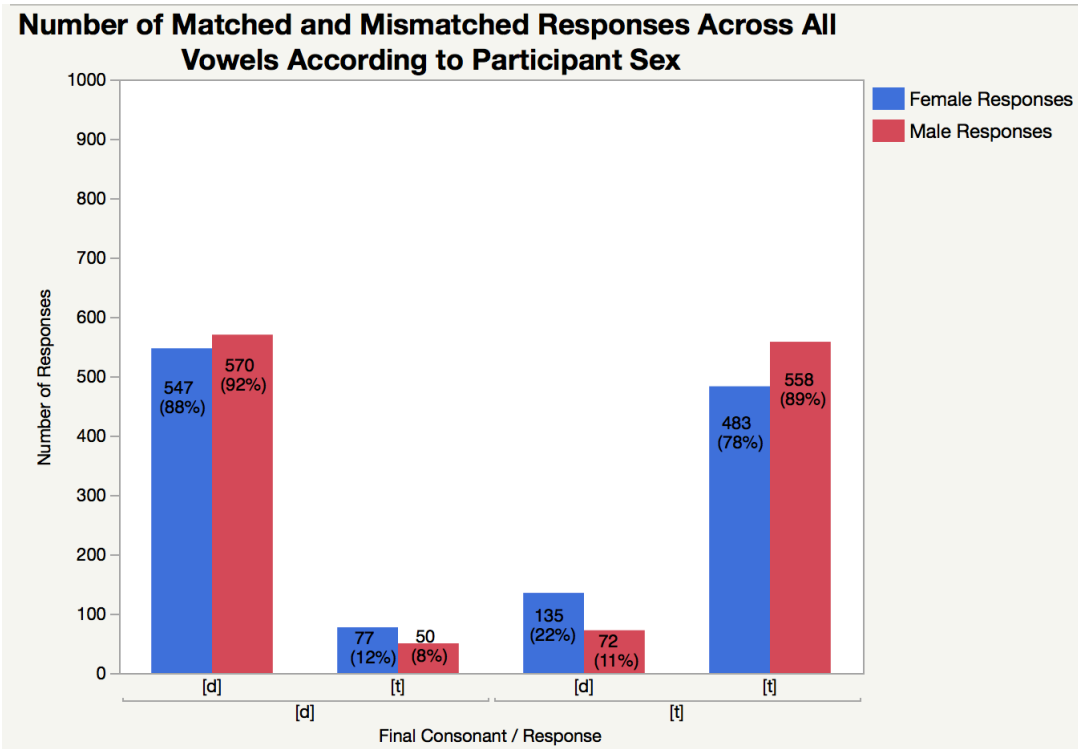


Figure 42: The number of correctly matched and incorrectly mismatched responses across all word-tokens according to sex for Experiment 3

Looking first of all at the overall results (Figure 42, above), males identified an original [d] correctly in 92% of cases and an original [t] in 89% of cases. Conversely, females correctly identified an original [d] in 88% of cases and an original [t] in 78% of cases. Therefore, males appeared to be more accurate overall than females by a small percentage.

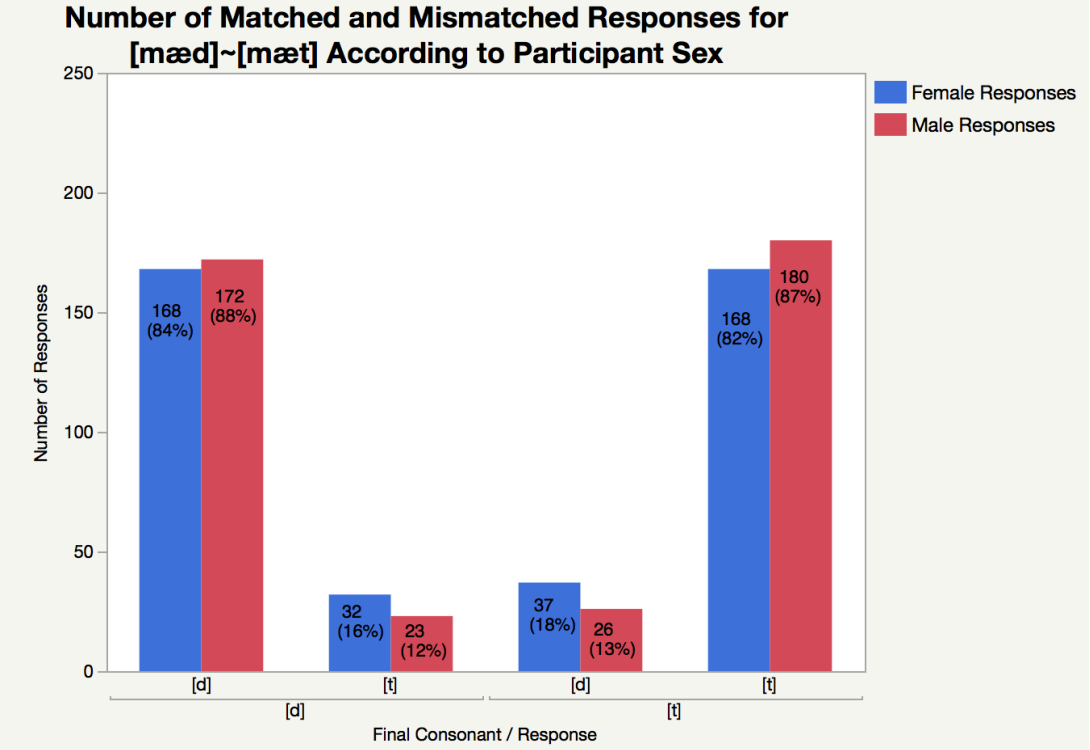


Figure 43: The number of correctly matched and incorrectly mismatched responses for [mæd]~[mæt] *mad~mat* according to sex for Experiment 3

Turning now to consider the word-token [mæd] *mad*, female participants matched the correct response in 84% of cases, compared with 88% for the male participants. Similarly, female participants selected the correct response in 82% of cases for [mæt] *mat*, whereas male participants were accurate in 87% of cases. Therefore, in the case of the vowel nucleus [æ], males once again appeared to be marginally more accurate in predicting voiced and voiceless word-final stops (Figure 43, above).

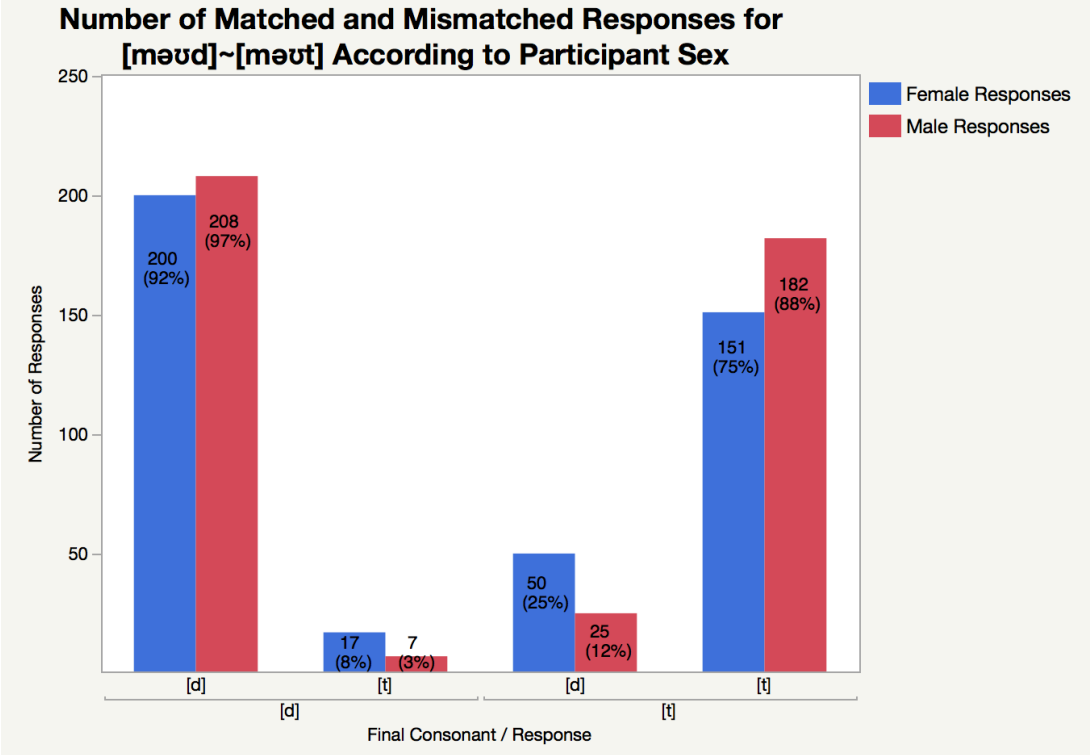


Figure 44: The number of correctly matched and incorrectly mismatched responses for [məʊd]~[məʊt] *mode~moat* according to sex for Experiment 3

In the case of [məʊd]~[məʊt] *mode~moat* (Figure 44, above) female participants correctly matched a voiced response in 92% of cases and voiceless response in 75% of cases. This can be compared with male participants, who correctly matched a voiced response in 97% of cases, and a voiceless response in 88% of cases. Again, males were consistently more accurate for word-tokens containing a [əʊ] vowel nucleus.

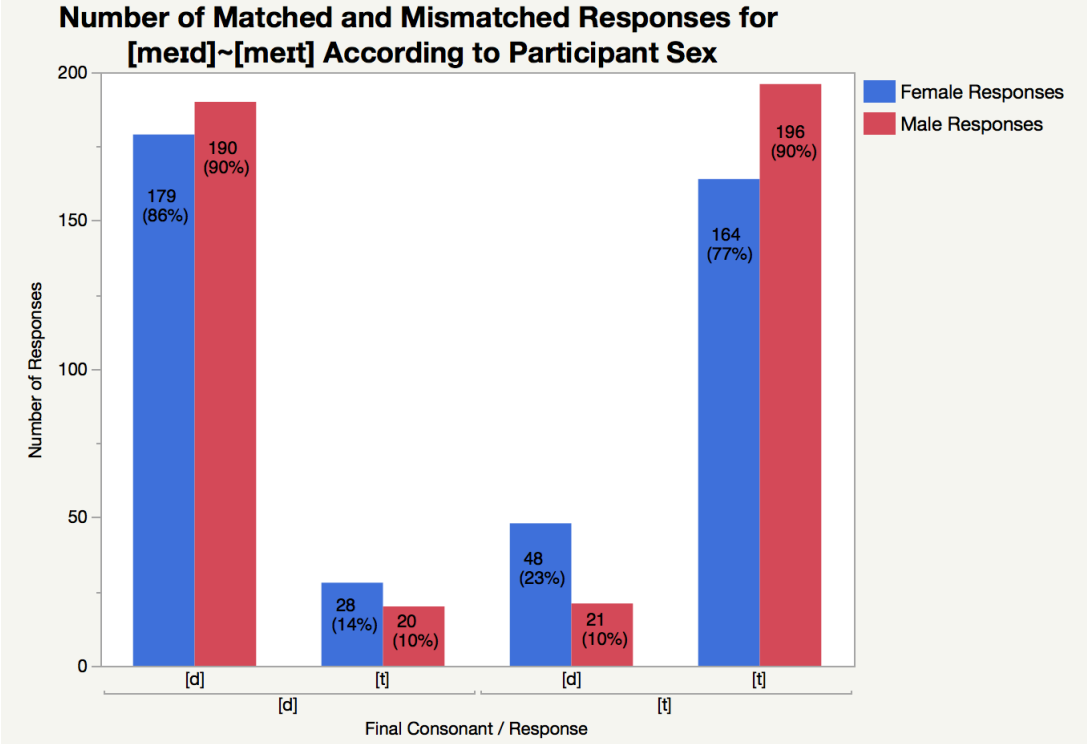


Figure 45: The number of correctly matched and incorrectly mismatched responses for [meɪd]~[meɪt] *made~mate* according to sex for Experiment 3

Finally, for [meɪd]~[meɪt] *made~mate* (Figure 45, above), male participants were correct in 90% of cases for both the word-final voiced and voiceless options. Conversely, female participants correctly matched a voiced response in 86% of cases and a voiceless response in 77% of cases. Therefore, males were once again slightly more accurate in identifying word-final voicing for word-tokens containing the vowel nucleus [eɪ].

The individual results from the *Fisher's Exact Test* for both males and females across the four criteria above remained identical to one another, as illustrated by Table 18 (below). The results of the *Chi-Squared* test also remained $P < .0001$.

Table 18: Fisher's Exact Test Results (Final Consonant and Response) for male and female responses for Experiment 3

Fisher's Exact Test	Probability	Significance	Alternative hypothesis
Left	1.0000		Prob(Final Consonant=[t]) is greater for Response=[d] than [t]
Right	<.0001	*	Prob(Final Consonant=[t]) is greater for Response=[t] than [d]
2-Tail	<.0001	*	Prob(Final Consonant=[t]) is different across Response

Summary of results according to sex: though males appear to be more consistently accurate in matching the correct response with the original word-final [VOICE] feature on the recording, there is no difference in the overall trend of results between male and female participants. As demonstrated by the *Fisher's Exact Tests***,** a consistently high degree of accuracy is apparent for both groups. These results therefore suggest that sex does not have an overriding on the perception of the word-final [VOICE] feature.

4.2.3.4 The relationship between reaction time and vowel length

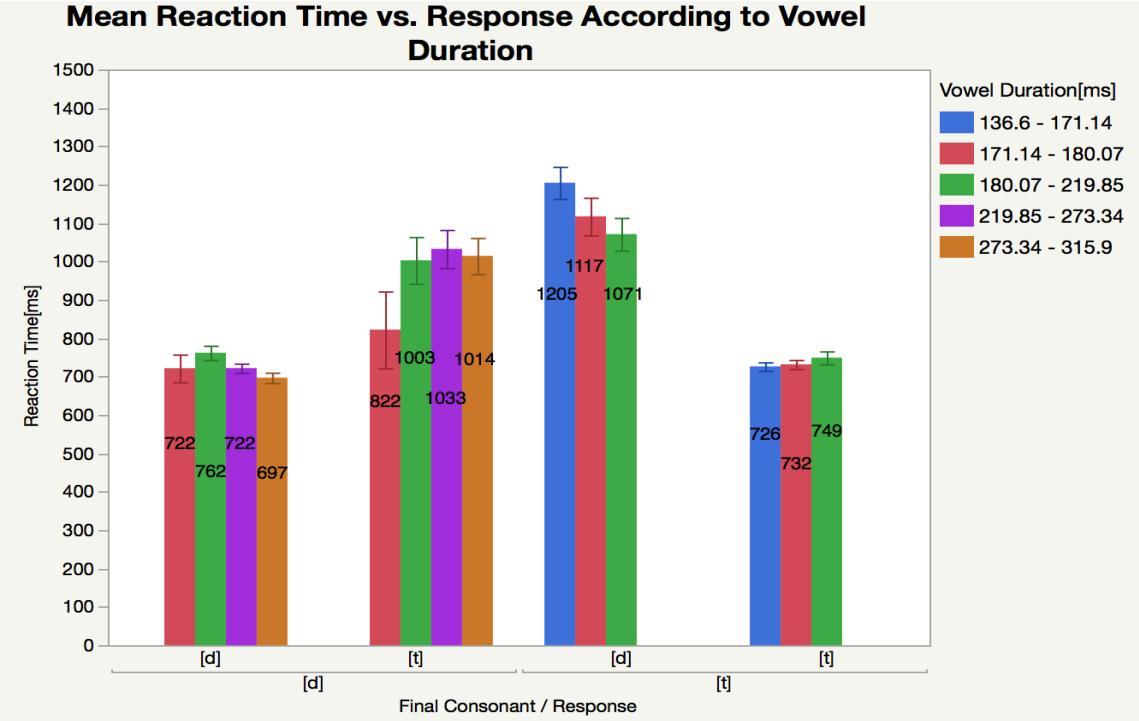


Figure 46: Reaction time distribution according to vowel length for Experiment 3

Finally, let us consider the relationship between the mean reaction times and the varying vowel durations recorded. Unlike Experiments 2 and 4, Experiment 3 is not a gated study. Instead, the vowel durations presented on the stimuli are simply those which were elicited in the recordings. Therefore, it proves interesting to consider the effect that these differing vowel durations may have had on the reaction times taken for the participants to respond to the stimuli. Figure 46 (above) illustrates the relationship between the mean reaction times and the vowel durations that the stimuli contained. Interestingly, reaction times are longer in the instances where participants have selected the *incorrect* or *mismatched* response, i.e. the word-final consonant that was not originally present on the recording. There are several observations to make here. The first is that the reaction times seem, in some cases, excessively long. Vowel duration may be a factor in this. In the case of the 1205ms reaction time, participants have

selected the incorrect response. The vowels contained within this group are among the shortest and would therefore not typically cue a [d] response, and this may have confused participants and caused a longer processing time. Ultimately, a [d] response was selected and therefore the vowel duration was not enough to cue the correct response. However, participants were perhaps aware of the shorter vowel that would normally cue a [t] response, and therefore experienced some confusion leading to a longer reaction time. The same trend is mirrored in the incorrect [t] responses, with longer reaction times perhaps signifying a confusion in reconciling the vowel duration with the correct word-final response. Conversely, in instances where the correct response has been chosen, reaction times are comparably quicker.

4.2.4 Discussion

Whether dividing the data according to the individual vowel nuclei, or looking more generally across all vowels, there is a high degree of accuracy across all contexts demonstrating the participants' ability to correctly identify the original word-final consonant.

Most notably, the results from Experiment 3 suggest that, in English, listeners are perceptually sensitive to subtle changes in vowel duration in CVC words (Klatt, 1976; Denes, 1955, among others). Recall that all of the acoustic information in the final consonant was removed from the stimuli. Despite this, across all contexts, the correctly matched final consonant response was selected in the majority of cases. Male speakers were slightly more accurate than female speakers, possibly due to the stimuli incorporating a male voice, though more research would be required to support or refute this claim. Reaction times were faster in cases where the correct word-final consonant had been selected.

The behaviour of participants overall did not vary depending on which of the three vowels they were presented with. This would suggest that this perception of lengthening occurs, and

is effective, across all vocalic contexts tested. As such, participants demonstrate the underlying representation of vowel length as long or short. Though it is possible that some phonetic information is made available in the word-initial consonant (Coleman 2003) and in the latter parts of the vowel that may allude to the voicing characteristic of the word-final consonant, it seems unlikely that this is overriding the influence of the vowel duration. Nevertheless, Experiment 4 will further control for this variable by employing synthesised speech as opposed to recorded speech.

Overall, the results from Experiment 3 support the predictions posited for this study. The most crucial aspect of these findings is to note that the differences in vowel duration between the word-tokens are small, and these results once again demonstrate a strong perceptual salience of word-final voicing based on preceding vowel duration. It would appear from the data that English listeners are able to use the information provided by vowel duration to accurately categorise a word-final consonant in a CV:C/CVC structure as voiced or voiceless. As such, these results suggest that the duration of the vowel is able to guide the perception of a listener with regard to the nature of the word-final [VOICE] feature.

4.2.5 Conclusion

The results of Experiment 3 provide strong evidence that, when presented only with a CV:/CV fragment of a CVC minimal pair in English, listeners remain able to make accurate judgements regarding the nature of the word-final [VOICE] feature. These results demonstrate once again that listeners are sensitive to fine-grained durational differences in vowels and this can cue the perception of word-final voicing.

CHAPTER FIVE

The effect of vowel duration on the perception of English word-final voicing in non-native speakers: identification tasks for L1 English and L1 German speakers

5.1 Introduction to Experiment 4

The stimuli utilised in Experiments 2 and 3 were devised using recorded speech. As previously discussed, the use of recorded speech is problematic as a speaker cannot fully control their speech production; particularly their speech rate, intonation, pitch, and volume. It is also difficult to completely unify the formation of stimuli in this way, as recorded speech is so variable. This may result in phonetic information containing unintended perceptual cues to remain. To overcome these issues, Experiment 4 comprised of a final forced choice identification task incorporating stimuli that were created using synthesised, computational speech. As with Experiment 2, the stimuli ended in a word-final ambiguous consonant, which participants were required to categorise as a voiced or voiceless sound.

In addition, Experiment 4 incorporated the role of native and non-native speech. Recall that one of the primary research questions that this thesis seeks to address is centred around the role of vowel duration as a cue for voicing in non-native English speakers. Exp1b addressed this area in terms of speech production. Experiment 4 therefore used the variables investigated in Exp1b as the basis for considering speech perception. As with Exp1b, three groups of participants were included; native English speakers, along with two groups of native German speakers who were second language learners of English. As with Exp1b, one group of German participants consisted of speakers who had lived in the UK and experienced an increased

exposure to spoken English. The second group of German participants were based in Frankfurt and therefore had experienced limited exposure to spoken English.

Finally, Experiment 4 addressed the role of the lexicon in speech perception by incorporating both real word and nonword pairs. The findings from Exp1b suggest that there is a greater degree of vowel lengthening and shortening in relation to voicing when speakers are producing words as opposed to nonwords. Experiment 4 therefore sought to decipher whether this is also the case in speech perception, and aimed to provide a more nuanced understanding of the relationship between vowel duration, word-final voicing, and the lexicon in native and non-native English.

5.2 Research questions

Experiment 4 aims to determine how manipulating vowel duration affects participants' perception of word-final voicing, and how this perception differs between native and non-native speakers of English. Additionally, Experiment 4 is interested in whether there is a difference in perception according to whether the stimuli consist of words or nonwords. The lexical status of the stimuli was as follows:

I. Stops
   a. [bæd]~[bæt] *bad~bat* (word V / word VL)
   b. [dæd]~*[dæt] *dad~*dat* (word V/ nonword VL)
   c. *[ʤæd]~*[ʤæt] **jad~*jat* (nonword V/ nonword VL)
   d. *[væd]~[væt] **vad~vat* (nonword V/ word VL)

II. Affricates
   a. [bæʤ]~[bæʧ] *badge~batch* (word V/word VL)
   b. *[ræʤ]~*[ræʧ] **radge~*ratch* (nonword V/nonword VL)

As with the previous experiments, it was expected that stimuli containing comparatively longer vowel durations would cue listeners to perceive a voiced word-final consonant, and stimuli containing shorter vowel durations would cue listeners to perceive a voiceless word-final consonant. Based on the results from Exp1b, native English participant's perception was expected to demonstrate greater sensitivity to vowel duration than non-native English participants. Additionally, non-native English participants who had experienced more exposure to spoken English were expected to show greater sensitivity than those who had experienced limited exposure. Finally, if a lexical effect was present, it was expected that this would cause a shift in perception in the direction of a word as opposed to a nonword, and that word pairs would show a greater degree of lengthening than nonword pairs.

5.3 Methodology

The stimuli: the stimuli required for this identification task were generated using the speech-synthesis program *IPOX* (Dirksen and Coleman, 1994). When devising the stimuli, it was essential to ensure that the only variation between the stimuli pairs was the duration of the vowel. All other aspects of the CVC stimuli needed to be held constant, and therefore synthesising these words computationally allowed for the creation of these controlled stimuli. *IPOX* uses an inbuilt *grammar of English* which is coded according to the different phoneme combinations of English phonology. To generate a word, *IPOX* uses *ASCII* phonemes which can be synthesised into a sequence. The word and nonword tokens found in Table 19 (below) were created in this way. The word-tokens are identical to those employed in Exp1b, incorporating both word-final stops and affricates. Following this, the synthesised word-token can be edited using the *IPOX* command codes. Here, the command code *O AH* was used to remove the aspiration from the final consonant and *O AF* was used to omit the stop burst.

Figures 47-48 (below) demonstrate the result of this omission. As a result of this process, all that remained present was the CV portion. These CV files were then saved.

Table 19: The list of word-tokens used for Experiment 4

English minimal pairs /d/~/t/	**Pair**	**IPA**	
bad~bat	word /d/~word /t/	/bæd/	/bæt/
dad~*dat	word /d/~nonword /t/	/dæd/	/dæt/
*jad~*jat	nonword /d/~nonword /t/	/ʤæd/	/ʤæt/
*vad~vat	nonword /d/~word /t/	/væd/	/væt/
English minimal pairs /ʤ/- /ʧ/	**Pair**	**IPA**	
badge~batch	word /ʤ/~word /ʧ/	/bæʤ/	/bæʧ/
*radge~*ratch	nonword /ʤ/~nonword /ʧ/	/ræʤ/	/ræʧ/

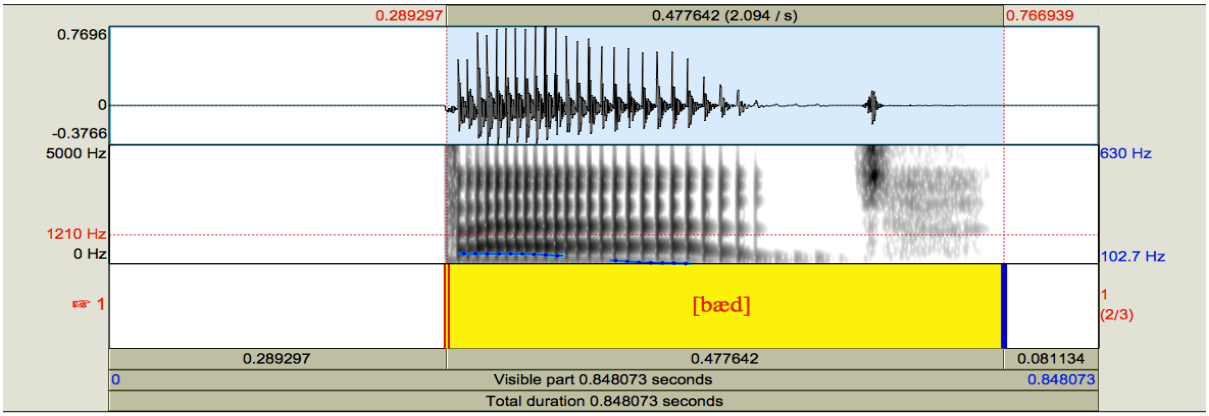


Figure 47: A full synthetic stimulus, here [bæd] *bad*, for Experiment 4

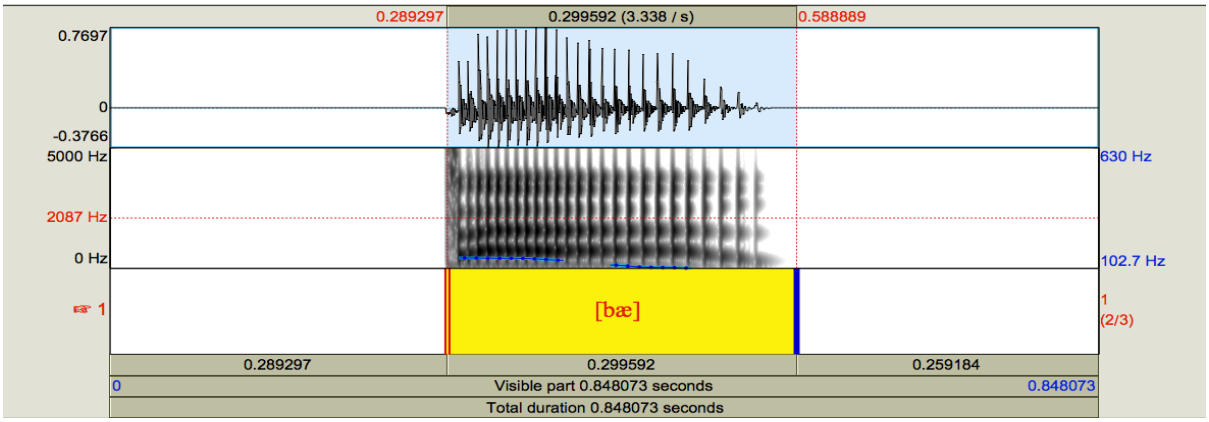


Figure 48: The CV portion of a stimulus, here [bæd] *bad*, (Figure 47), for Experiment 4

Following this, *Praat* was used create gated vowel durations. To begin, the steady state period of the vowel for both the voiced and voiceless token of each word pair was selected. Using the voiced token as a base, glottal pulses were then systematically removed from this steady state period, beginning in the centre of the vowel and then working outwards, moving from left to right of the centre. This process continued until the vowel became as close as possible in duration to the voiceless token of the same word pair. As with the previous experiments, all cuts were made at the zero-crossing boundary to prevent audible clicks. This process is illustrated in Figures 49-50 (below). The individual vowel lengths across the gates for the stimuli pairs are detailed in Tables 20 and 21 (below).

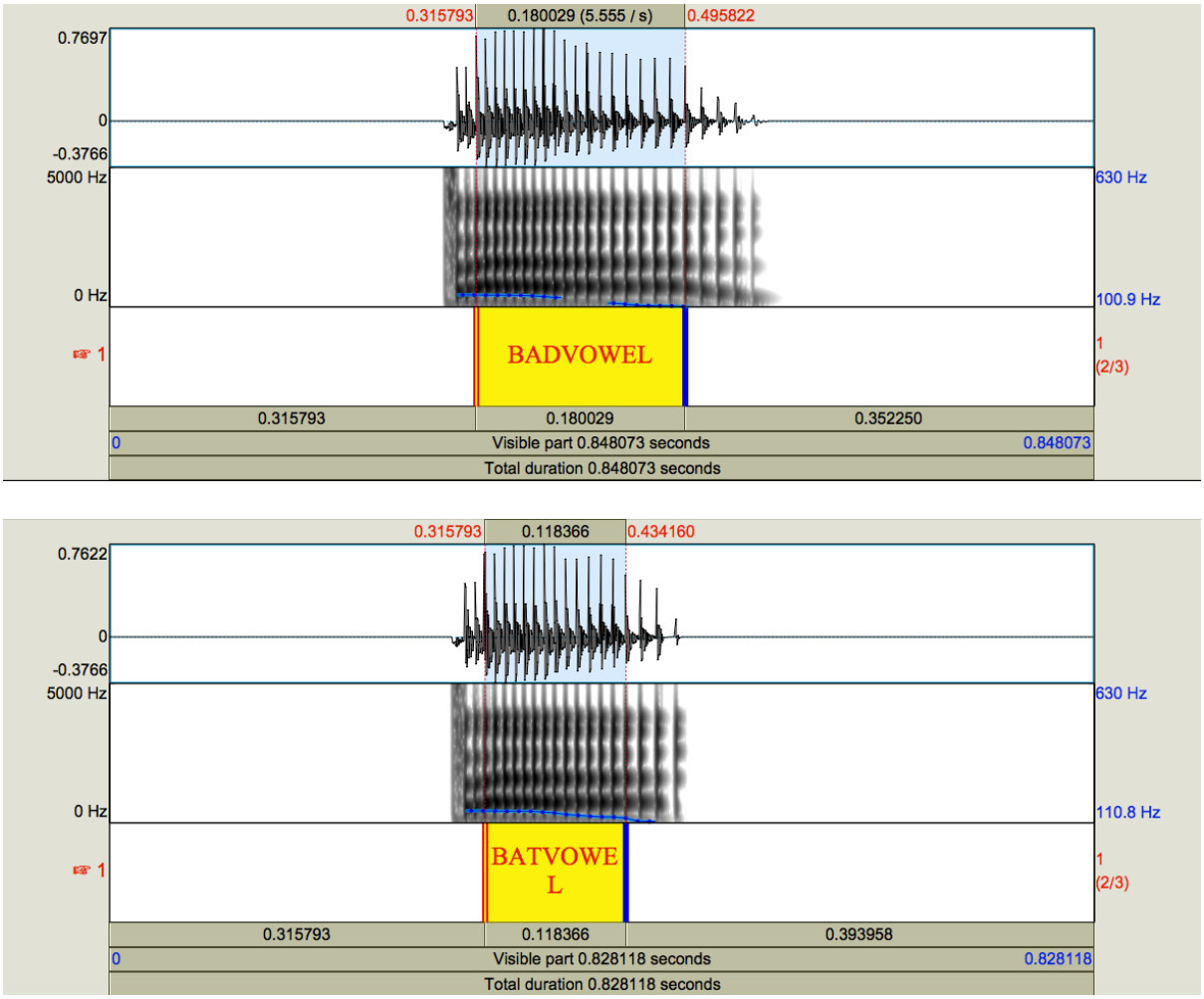


Figure 49: The use of *Praat* to measure the steady state period of the vowel for the word pairs, here [bæd] *bad* and [bæt] *bat*, for Experiment 4

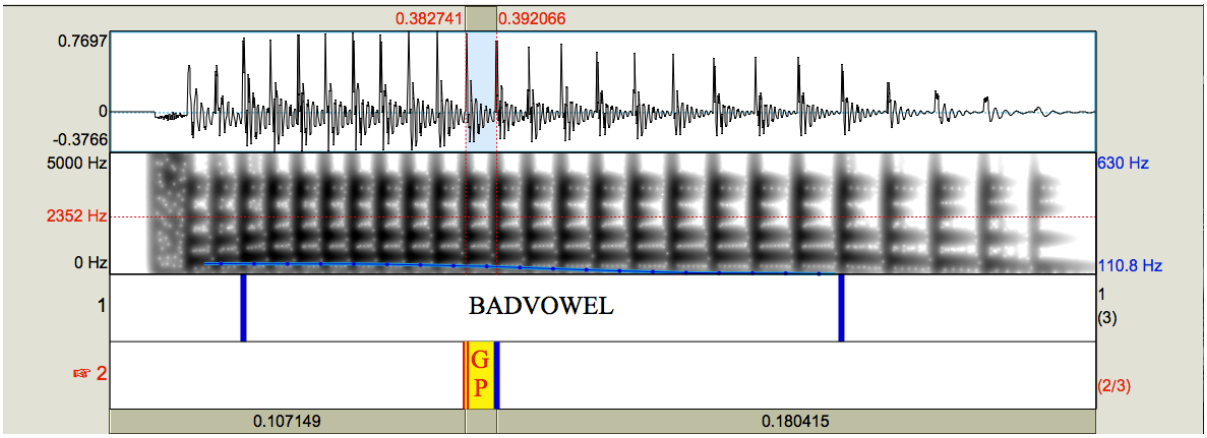


Figure 50: The use of *Praat* to remove glottal pulses systematically from the centre of the voiced file, here [bæd] *bad*, for Experiment 4

Table 20: The individual vowel gate durations for the word-final stop stimuli for Experiment 4[19]

Gate	[bæd]~[bæt] *bad~bat*	[dæd]~*[dæt] *dad~*dat*	*[ʤæd]~*[ʤæt] **jad~*jat*	*[væd]~[væt] **vad~vat*
1	180ms	180ms	182ms	180ms
2	171ms	171ms	173ms	171ms
3	161ms	161ms	163ms	161ms
4	151ms	151ms	153ms	151ms
5	142ms	142ms	144ms	142ms
6	134ms	134ms	136ms	134ms
7	123ms	123ms	125ms	123ms
8	112ms	112ms	113ms	112ms

Table 21: The individual vowel gate durations for the word-final affricate stimuli for Experiment 4[20]

Gate	[bædʒ]~[bætʃ] *badge~batch*	*[rædʒ]~*[rætʃ] **radge~*ratch*
1	194ms	191ms
2	185ms	182ms
3	175ms	172ms
4	165ms	163ms
5	155ms	153ms
6	146ms	144ms
7	137ms	135ms
8	126ms	124ms
9	115ms	113ms

[19] The original *bat* [bæt] file was 118ms, the original **dat* *[dæt] file was 118ms, the original **jat* *[ʤæt] file was 120ms and the original *vat* [væt] file was 118ms

[20] The original *batch* [bætʃ] file was 118ms and the original **ratch* *[rætʃ] file was 118ms

As with Experiment 2, an ambiguous word-final consonant was then created and added to the vowel gates form the individual CV:?/CV? fragments. In order to generate this aspect of the stimuli, the original voiced and voiceless word-final consonants from the synthesised words were selected in *Praat,* inclusive of the closure duration and release burst. Here, boundaries were once again placed where there was a clear shift in the auditory and visual energy of the spectrogram. The vowel transitions into the consonant were retained. As discussed previously, this thesis acknowledges that spectral cues for voicing may remain in the latter part of the vowel. However, presenting the full vowel durations to participants was considered more crucial for the nature of this research that attempting to omit all other potential cues. Cutting glottal pulses from the centre of the steady state period, rather than from the end of the vowel, also ensured that a more natural vocalic shape was retained overall. These word-final consonants were then extracted and saved as two individual files. These files were then converted into a mono file in the same manner used in Experiment 2. In each case, the two stereo channels were overlaid and merged to form one ambiguous consonant with a combined closure duration and release burst. Here, the nature of voicing in the word-final consonant was neutralised. These sound-files were then pasted onto each of the vowel gates. This process is illustrated in Figures 51-53 (below).

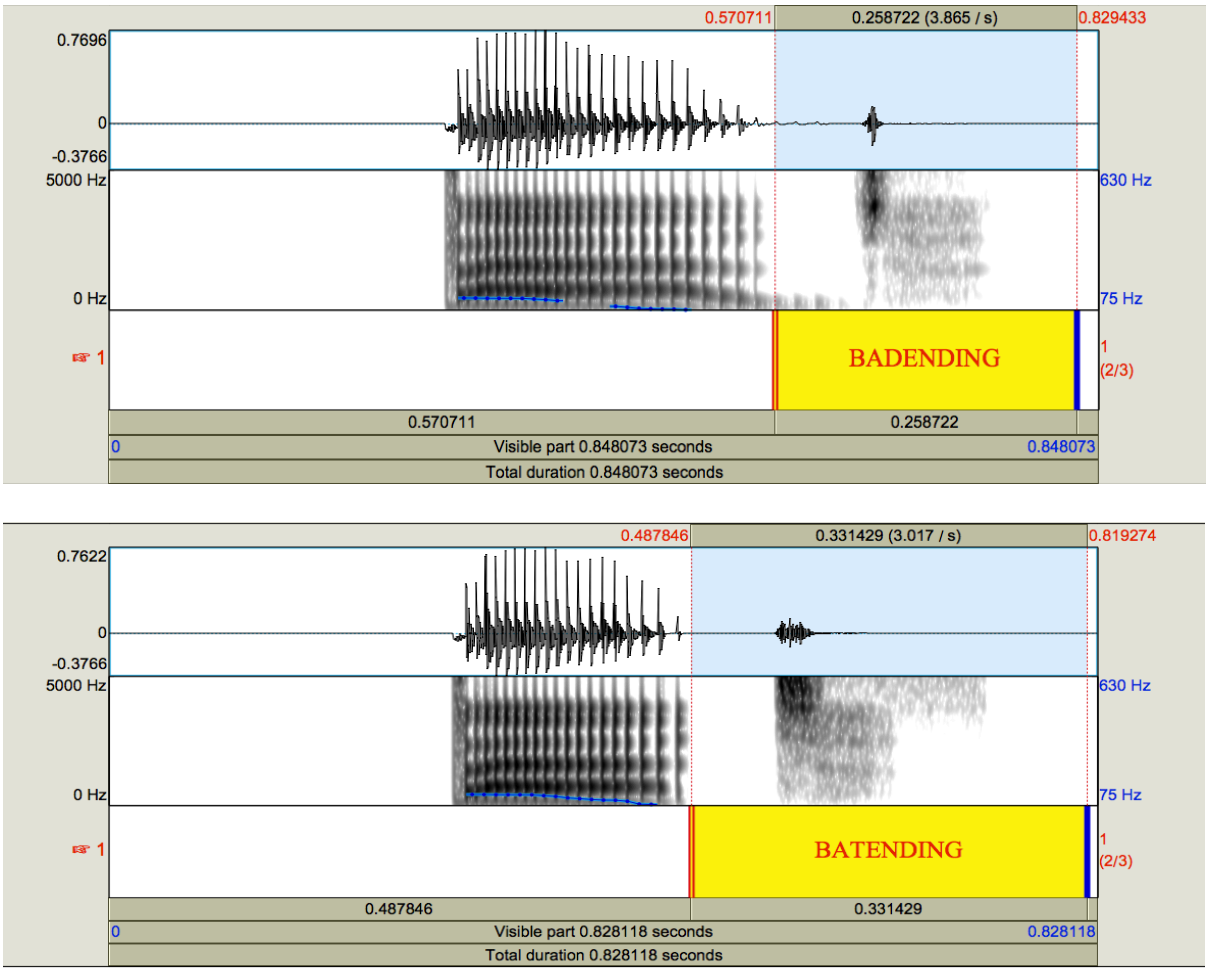


Figure 51: The use of *Praat* to segment the word-final consonant from the word pairs, here [bæd] *bad* and [bæt] *bat*, for Experiment 4

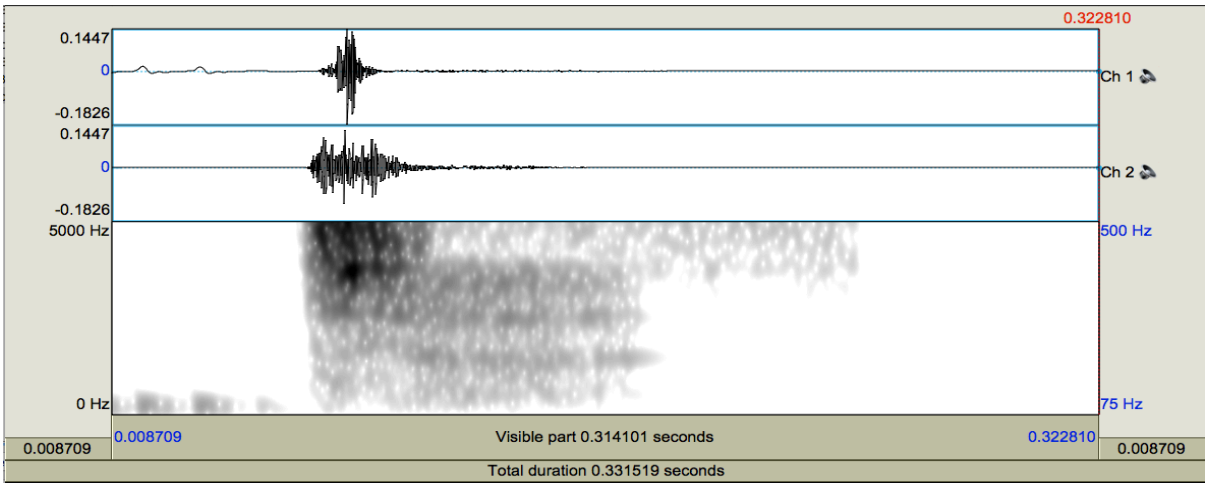


Figure 52: The use of *Praat* to convert the word-final [d] (channel 1) and [t] (channel 2) into two stereo files for Experiment 4

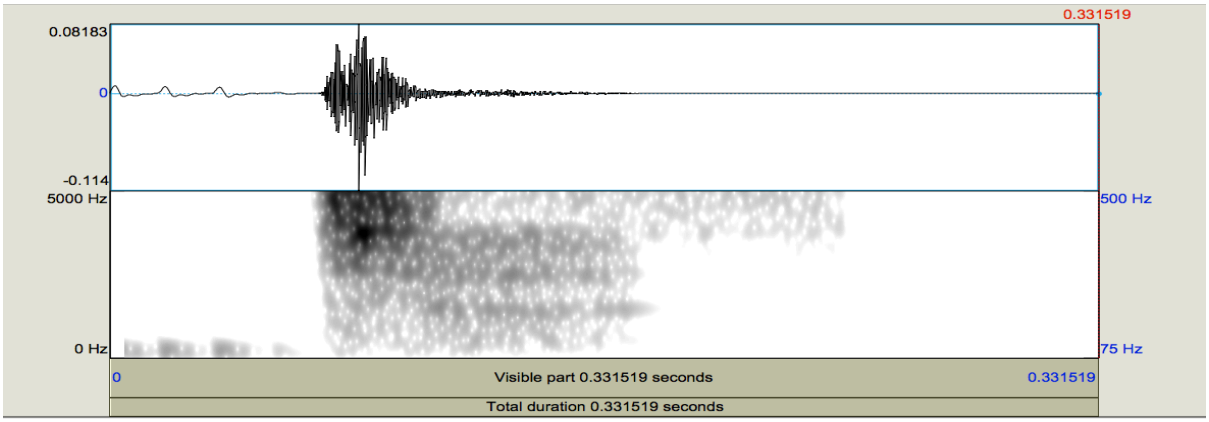


Figure 53: The use of *Praat* to create a single mono file by overlaying channel 1 and channel 2 (Figure 52) for Experiment 4

Finally, a *Praat* script was used to level the pitch across all stimuli to ensure that they contained no unintended rising or falling pitch.

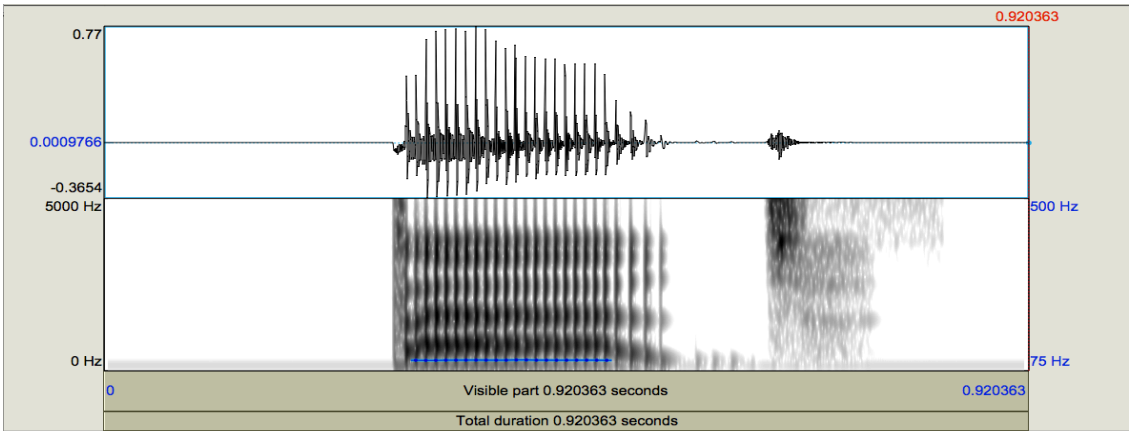


Figure 54: The use of *Praat* to level the pitch across the stimuli, here [bæX], for Experiment 4

The experimental audio file: as with Experiments 2 and 3, *SPLICE* was utilised to generate a combined audio file that allowed the stimuli to be presented to participants at a fixed speed with consistent ordering. In total, one hundred and fifty stimuli were presented to the participants. Here, the stimuli were presented in two blocks. The first block presented the word-final affricates; two word pairs containing nine gates, randomised three times. The second block contained ninety-six stimuli; four word pairs containing eight gates, randomised three times.

The beginning of the experiment was signalled by a visual countdown from *5* to *1*. Participants then heard a *BLEEP*, followed by 500ms of silence before the first auditory stimulus was presented. The visual presentation of the two sound choices appeared simultaneously with the auditory stimuli. Participants then had a 2500ms response window, followed by a 2500ms *PAUSE.* Another *BLEEP* then signalled the beginning of the next trial. All responses were recorded automatically.

Participants: a total of thirty-seven participants took part in this identification task. These speakers were divided into two language groups; L1 English speakers and L1 German speakers (L2 English). Eleven native speakers of Southern British English were recruited from the University of Oxford. Six participants were female and five participants were male. Participants were between the ages of eighteen and thirty. They had no known hearing or language disorders, and all had normal-to-corrected vision. One participant was left-handed. Participants were compensated £5 for their time.

Twenty-six native speakers of German were recruited. Participants were between the ages of eighteen and thirty. They were organised into two groups in accordance with their exposure to English. These criteria were identical to those used in Exp1b. Specifically, for German participants based in Germany, it was crucial that they had never lived in an English-speaking country, that neither of their parents or caregivers were native English speakers, and that they were not studying an English related subject for their degree. However, despite these controls, all participants needed to identify as having good English proficiency as they were informed that the study would be conducted entirely in spoken and written English. Participants who were native speakers of German living in the UK also had to meet a set of criteria. Participants within this group were required to have been resident in the UK, or another English-speaking

country, for a minimum of three years. It was also necessary that they identified as being fluent in spoken English. Seven native speakers of German participated at the University of Oxford in the UK. These participants consisted of five females and two males. None of the participants were left-handed. Nineteen native speakers of German participated at the Goethe University in Germany. These participants consisted of fourteen females and five males. Three participants were left-handed. Participants had no known hearing or language disorders, and all had normal-to-corrected vision. Participants were reimbursed £5/ €5 for their time.

Procedure: participants were recruited by email from both the University of Oxford, UK and Goethe University, Germany. The experiments took place in the Language and Brain Laboratory in Oxford, and the Phonetics Department in Germany. Participants were invited to attend a thirty-minute session of their choice. As with the previous experiments in this thesis, participants were first provided with an information sheet and their written consent was obtained.[21]

The experimental set up was identical to Experiment 2. As previously stated, Experiment 4 consisted of a forced choice identification task. Participants were requested to sit in front of a 17" CRT monitor, attached to which was a pair of headphones and a handheld response box with two buttons. Participants who were left-handed were required to hold the button box upside-down so that their dominant hand performed the same action as the right-handed participants. It was explained to the participants that they would hear a series of English words binaurally and at a comfortable volume; some of these words would be real words and some would be nonsense words. A choice of two sounds would appear on the screen, and participants were requested to listen carefully to the stimuli and decide which sound they perceived hearing

[21] A copy of the information sheet and consent form for Experiment 4 can be found in Appendix D

word-finally. They were instructed that they would have around two and a half seconds to make their decision. They could indicate their choice by pressing the button which corresponded to the side of the screen on which their chosen sound appeared. Prior to the experiment beginning, participants responded to a training block of eight trials. Following the training block, participants were given the opportunity to ask questions for further clarification of the experimental procedure before the experiment began.

As with the previous identification tasks, all participants heard the stimuli in the same order. The recording consisted of two blocks of stimuli, affricates then stops, separated by a short break of eight seconds. The full recording lasted for approximately nine and a half minutes. The first of the two blocks required participants to decide whether they had heard the affricate [ʤ] or [ʧ] word-finally. The second block required a decision to be made between the stops [d] or [t] word-finally. The ordering in which the two sounds appeared on the screen alternated between each session to allow for any variation in dominant sight, but remained consistent within a session. First-responses were recorded and could not be changed, and therefore participants were instructed to go with their first instinct and not to overthink or try to change their decision once a selection had been made.

Following the completion of the experiment, participants were given the opportunity to ask questions which, as with Experiments 2 and 3, could be more specific about the nature of the study.

5.4 Analysis

As with Experiments 2 and 3, the analysis of this experiment was conducted using *JMP (SAS)*. This analysis was concerned with determining the percentage of word-final voiced and voiceless responses recorded at each of the vowel gates, and whether there was a distinct

difference in responses according to the language group being tested. Experiment 4 also considered whether there was a lexical effect present in the data, i.e. whether participants were shifting their responses in the direction of a word versus a nonword regardless of the vowel duration, and whether perception was clearer for words than for nonwords.

In total, five thousand, five hundred and fifty responses were recorded from the thirty-seven participants. As with Experiments 2 and 3, the first stage of the analysis was to clean the results using the same criteria as detailed in Experiment 2.

On the basis of this, two participants and one trial were excluded for failing to demonstrate 80% in overall accuracy. Here, 80% was considered to be the appropriate level of accuracy for both participants and trials as Experiment 4 incorporated both non-native speakers and synthesised stimuli which proved more challenging to perceive. As such, a total of 12% of responses were excluded. This was in line with the percentage of excluded responses for Experiments 2 and 3.

5.4.1 Initial observations

The results across the following sections of analysis have one thing in common; though the correlations are overall as expected, these correlations are demonstrated to a much lower extent than anticipated. It is most likely that these results are due to issues with the synthetic nature of the stimuli, and these limitations will be discussed in more detail at the end of the analysis (cf. Section 5.5 Limitations of the study).

5.4.2 Acoustic input: the effect of vowel duration

Experiment 4 is concerned with determining the extent to which systematically varying vowel duration within stimuli can guide participants' perception towards the presence or

absence of the word-final [VOICE] feature. This first section of the analysis therefore considers this in relation to both word-final stops and affricates. The role of the lexicon and the influence of the word-initial consonants will be considered subsequently.

5.4.2.1 Stops

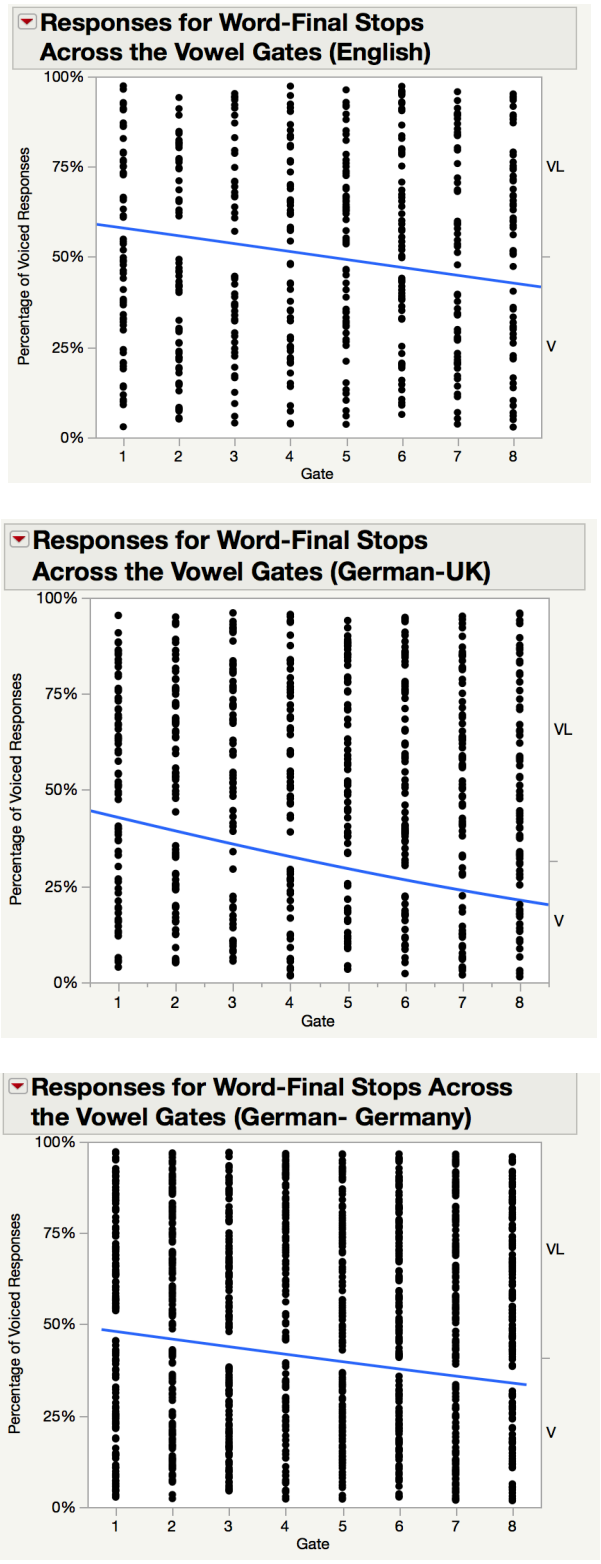


Figure 55: Voiced and voiceless responses for word-final stops across the eight vowel gates according to the three language groups for Experiment 4

Table 22: The percentage of voiced and voiceless responses across the eight vowel gates preceding word-final stops according to the three language groups for Experiment 4

English			German- UK			German-Germany		
Gate	**% V[22]**	**% VL**	**Gate**	**% V**	**% VL**	**Gate**	**% V**	**% VL**
1	57	43	1	43	57	1	48	52
2	62	38	2	40	60	2	45	55
3	52	48	3	33	67	3	43	57
4	49	51	4	36	64	4	41	59
5	46	54	5	29	71	5	43	57
6	39	61	6	22	78	6	38	62
7	55	45	7	25	75	7	38	62
8	42	58	8	23	77	8	31	69

For all three groups, there are consistently more voiced [d] responses perceived at the longest gate (gate 1) than at the shortest gate (gate 8). This is as expected and is in line with the predictions posited for this study. However, the gradient of these results is much shallower than anticipated.

Looking first at the results for the group of the native English participants, there is a statistically significant difference (P = .0185) across the gates. However, there is only a 15% difference in the number of perceived voiced responses between gates 8 and 1. As such, we would have expected this difference to be more exaggerated. The significance of the difference between the gates is also much lower than that of the two groups of German participants. The reason for this could be due to a difficulty with perceiving the synthesised stimuli; the speakers' confusion in hearing this unfamiliar format of speech may be more pronounced due to it being in their native language and thereby not presenting the usual perceptual cues in the normal way.

Considering now the two groups of German participants. The results for both German (UK) (P < .0001) and German (Germany) (P = .0004) groups show a statistically significant difference across the gates, with a high degree of statistical significance between the responses for the German (UK) participants in particular. However, the participants from the German

[22] Here, *V* refers to *voiced* and *VL* refers to *voiceless*

(UK) group demonstrate just a 20% difference in the number of voiced results perceived between gates 8 and 1, and the German (Germany) group just 17%. Both groups of German participants also perceive less than 50% voiced responses at the longest vowel gate; 43% and 48% respectively. This is in line with the prediction that German participants will perceive more voiceless responses overall, as guided by their L1 phonology. As expected, German participants based in the UK have more differentiated end points than those based in Germany. In this way, German listeners who have received more exposure to spoken English show an increased sensitivity to vowel duration as a perceptual cue for word-final voicing.

These results are illustrated in Figure 55 (above) and the numerical values can be found in Table 22 (above).

5.4.2.2 Affricates

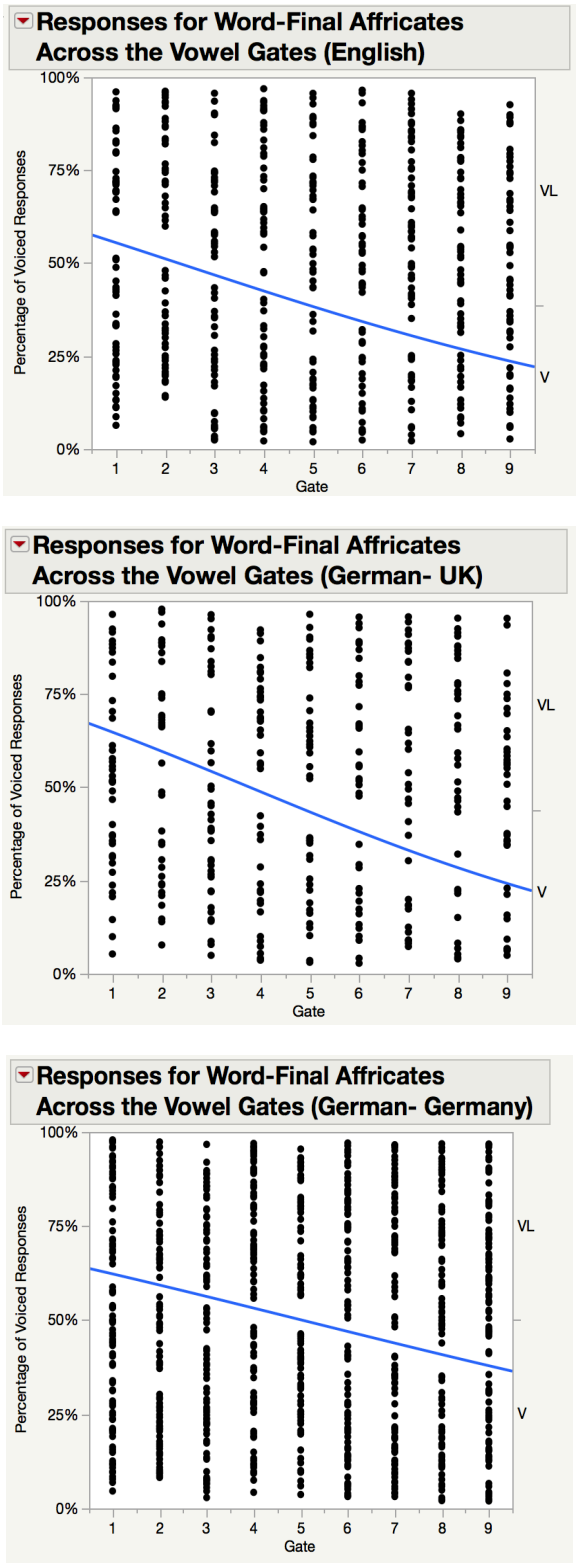


Figure 56: Voiced and voiceless responses for word-final affricates across the nine vowel gates according to the three language groups for Experiment 4

Table 23: The percentage of voiced and voiceless responses across the nine vowel gates preceding word-final affricates according to the three language groups for Experiment 4

English			German- UK			German-Germany		
Gate	**% V**	**% VL**	**Gate**	**% V**	**% VL**	**Gate**	**% V**	**% VL**
1	53	47	1	70	30	1	59	41
2	48	52	2	48	52	2	61	39
3	49	51	3	64	36	3	62	38
4	47	53	4	44	56	4	45	55
5	43	57	5	45	55	5	54	46
6	33	67	6	40	60	6	45	55
7	24	76	7	29	71	7	49	51
8	25	75	8	27	73	8	41	59
9	26	74	9	27	73	9	35	65

Let us first consider the results for the group of English speakers. The spread of results across the gates show a high statistically significant difference ($P < .0001$), illustrating that participants perceived significantly more voiced [ʤ] responses at gate 1 and significantly more voiceless [ʧ] responses at gate 9. There is a difference of 27% between the number of voiced responses perceived at the longest and shortest gates. However, the clarity of this trend is not as expected. Participants perceived voiced responses in 53% of cases at gate 1, and as such perceived 47% of voiceless responses at this gate. Again, it was expected that this trend would be much more exaggerated.

Interestingly, the results from the group of German (UK) participants follow a trend more akin to what would have been expected for the native English participants. Once again, there is a highly significant statistical difference ($P < .0001$) across the gates, however in this instance there is a 43% difference between the number of voiced responses perceived at gates 1 and 9; 16% higher than the native English group. As such, participants perceived a voiced response in 70% of cases at gate 1, and in 27% of cases at gate 9. Here, vowel duration appears to be guiding the participants' perception of voicing, and as these speakers of German have had an increased exposure to spoken English they look to be demonstrating an increased sensitivity to this L2 perceptual cue.

Once again, there is a highly significant statistical difference across the gates for the German (Germany) group ($P < .0001$). However, there is just a 24% difference between gate 1 and gate 9, resulting in this being the least differentiated group.

These results are illustrated in Figure 56 (above) and the numerical values across the gates can be found in Table 23 (above).

5.4.3 The effect of the lexicon

Having now evaluated the overall results in accordance with whether the stimuli ended in an ambiguous word-final stop or affricate, we now turn to the impact that lexical status and bias may have had on the data. Recall that the stimuli were grouped into the following pairs:

III. Stops
   a. [bæd]~[bæt] *bad~bat* (word V / word VL)
   b. [dæd]~*[dæt] *dad~*dat* (word V/ nonword VL)
   c. *[ʤæd]~*[ʤæt] **jad~*jat* (nonword V/ nonword VL)
   d. *[væd]~[væt] **vad~vat* (nonword V/ word VL)

IV. Affricates
   a. [bæʤ]~[bæʧ] *badge~batch* (word V/word VL)
   b. *[ræʤ]~*[ræʧ] **radge~*ratch* (nonword V/nonword VL)

The following section of the analysis will look to consider these individual pairs in order to determine whether lexical information has interacted with the acoustic information provided by the vowel duration.

5.4.3.1 Stops

[bæd]~[bæt] *bad~bat*

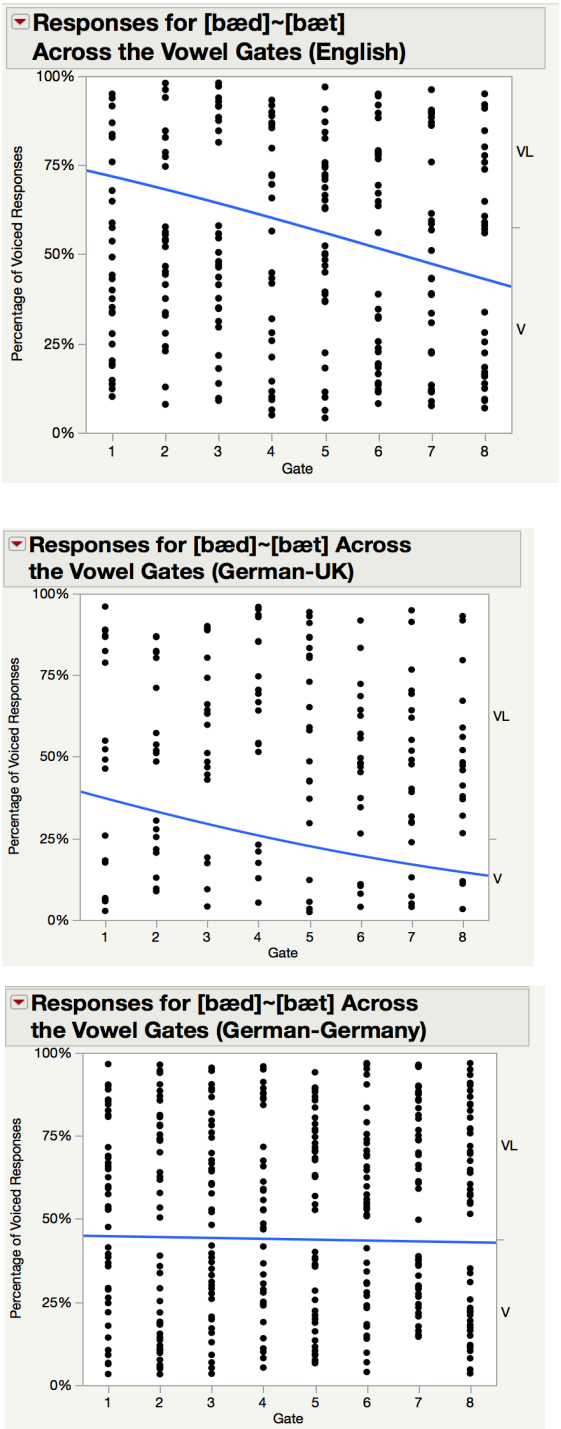


Figure 57: Voiced and voiceless responses across the eight vowel gates for [bæd]~[bæt] *bad~bat* for Experiment 4

[dæd]~*[dæt] *dad~*dat*

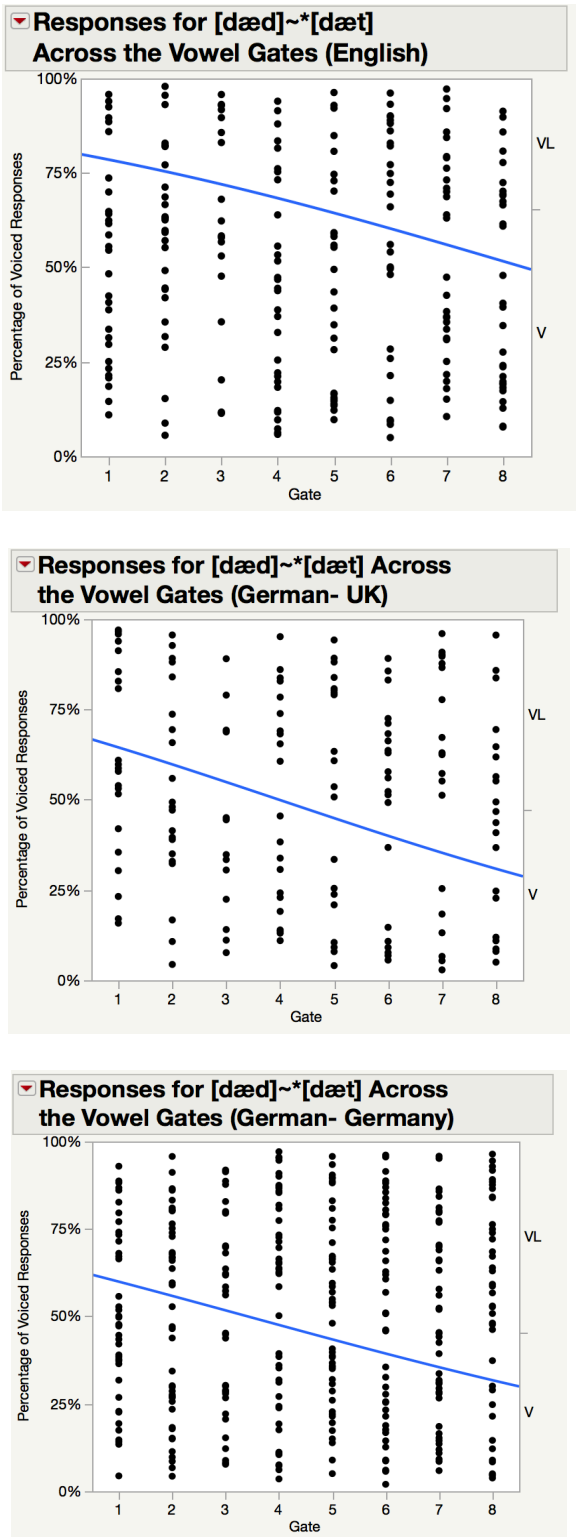


Figure 58: Voiced and voiceless results across the eight vowel gates for [dæd]~*[dæt] *dad~*dat* for Experiment 4

***[ʤæd]~*[ʤæt] *_jad_~*_jat_**

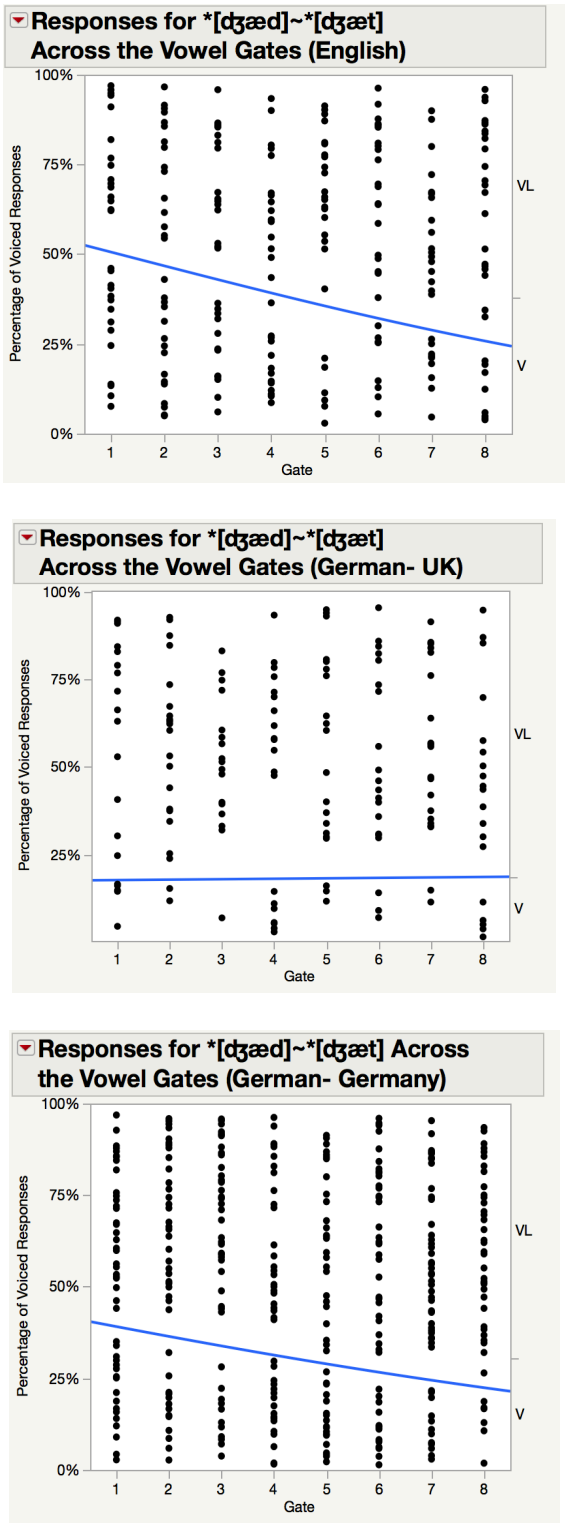
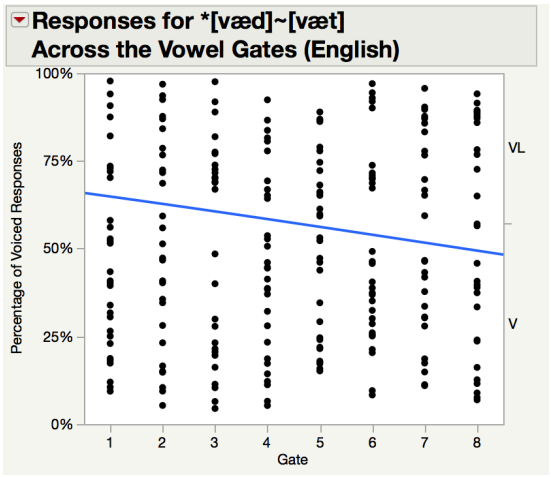
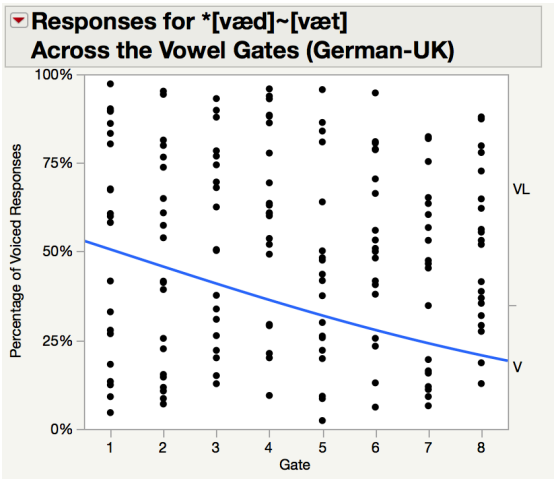


Figure 59: Voiced and voiceless results across the eight vowel gates for *[ʤæd]~*[ʤæt] **jad*~**jat* for Experiment 4

***[væd]~[væt] *_vad~vat_**





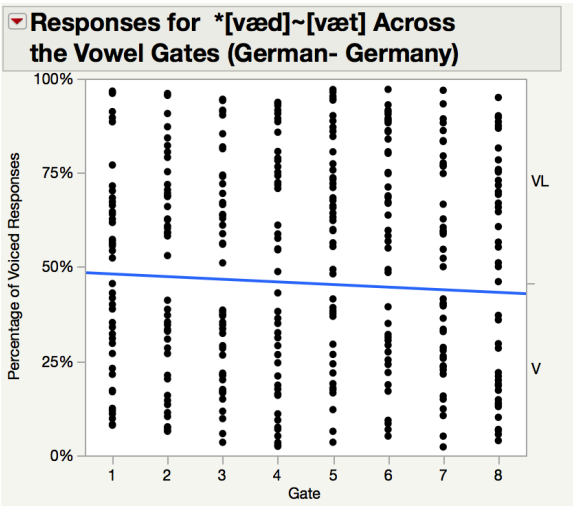


Figure 60: Voiced and voiceless results across the eight vowel gates for *[væd]~[væt] **vad~vat* for Experiment 4

Table 24: The percentage of voiced and voiceless responses across the eight vowel gates for [bæd]~[bæt] *bad~bat* according to the three language groups for Experiment 4

English			German- UK			German-Germany		
Gate	**%V**	**%VL**	**Gate**	**%V**	**%VL**	**Gate**	**%V**	**%VL**
1	76	24	1	38	62	1	39	61
2	69	31	2	38	62	2	50	50
3	67	33	3	21	79	3	47	53
4	52	48	4	29	71	4	40	60
5	52	48	5	19	81	5	48	52
6	52	48	6	20	80	6	43	57
7	53	47	7	19	81	7	41	59
8	44	56	8	15	85	8	43	57

Table 25: The percentage of voiced and voiceless responses across the eight vowel gates for [dæd]~*[dæt] *dad~*dat* according to the three language groups for Experiment 4

English			German- UK			German-Germany		
Gate	**%V**	**%VL**	**Gate**	**%V**	**%VL**	**Gate**	**%V**	**%VL**
1	81	19	1	62	38	1	63	37
2	77	23	2	62	38	2	55	45
3	58	42	3	64	36	3	50	50
4	73	27	4	52	48	4	40	60
5	72	28	5	38	62	5	47	53
6	50	50	6	33	67	6	39	61
7	53	47	7	33	67	7	43	57
8	56	44	8	38	62	8	27	73

Table 26: The percentage of voiced and voiceless responses across the eight vowel gates for *[ʤæd]~*[ʤæt] **jad~*jat* according to the three language groups for Experiment 4

English			German- UK			German-Germany		
Gate	**%V**	**%VL**	**Gate**	**%V**	**%VL**	**Gate**	**%V**	**%VL**
1	47	53	1	30	70	1	42	58
2	50	50	2	10	90	2	29	71
3	46	54	3	5	95	3	25	75
4	45	55	4	33	67	4	37	63
5	25	75	5	15	85	5	39	61
6	30	70	6	14	86	6	34	66
7	33	67	7	10	90	7	24	76
8	26	74	8	30	70	8	12	88

Table 27: The percentage of voiced and voiceless responses across the eight vowel gates for *[væd]~[væt] **vad~vat* according to the three language groups for Experiment 4

English			German- UK			German-Germany		
Gate	**%V**	**%VL**	**Gate**	**%V**	**%VL**	**Gate**	**%V**	**%VL**
1	70	30	1	43	57	1	48	51
2	58	42	2	52	48	2	47	53
3	50	50	3	48	52	3	52	48
4	68	32	4	24	76	4	48	52
5	52	48	5	43	57	5	37	63
6	62	38	6	20	80	6	40	60
7	52	48	7	38	62	7	50	50
8	45	55	8	10	90	8	45	55

The individual trends described in this section are depicted in Figures 57-60 (above). The detailed percentages of the responses for each of the languages and word-initial consonant groups can be found in Tables 24-27 (above).

Let us first consider the results for the word/word pair [bæd]~[bæt] *bad~bat*. For the native English speakers, the results across the vowel gates are statistically significant (P = .0029) and there is a 32% difference in the number of voiced [d] responses at the longest and shortest vowel gates. These results can be compared with both groups of German participants. German participants based in the UK (P = .0267) demonstrated a 23% difference in the number of voiced responses perceived at the longest and shortest vowel gates. Conversely, German participants based in Germany (P = .8139) produced just a 4% difference in the opposite direction to that which was expected, with no statistically significant difference between responses. Overall, both groups of German participants perceived predominantly voiceless responses across the gates, regardless of the phonetic information being provided by the vowel duration. These results support predictions as no lexical bias was expected to have occurred due to both targets being words as opposed to nonwords. As expected, native English participants demonstrate the greatest difference between the vowel gates. Also, though German participants from both groups perceived predominantly voiceless responses, speakers of

German based in the UK once again show more differentiated results than German speakers based in Germany.

In the case of the word/nonword pair [dæd]~*[dæt] *dad*~**dat*. It was expected that more voiced responses would be perceived overall, as [dæd] *dad* is a real word ending in a voiced stop, as opposed to the nonword *[dæt] **dat* which ends in a voiceless stop. This certainly seems to be the case for native English speakers, who even at the shortest vowel gate still perceive 56% of voiced responses overall. Note that there is a 25% difference between gates 1 and 8 (P = .0036). The results for the German speakers do not appear to be as affected by the lexical status of the stimuli. Here, there is an overall decline in the number of voiced responses recorded between the longest and shortest vowel gates. German participants based in the UK demonstrate a 24% difference in the number of voiced responses across the longest and shortest vowel gates (P = .0040), whilst German participants based in Germany exhibit a 36% difference (P = .0004). Both groups of German speakers therefore show significantly differentiated end points, with German participants based in Germany showing the most significant difference across the three groups. Here, the percentage of voiced responses at the longest gate was the highest of all four word pairs for both groups of German speakers. This may be driven by the lexical representation of the word [dæd] *dad*, and the absence of this lexical representation for the nonword *[dæt] **dat.*

As with the word/word pair, for the nonword/nonword pair *[ʤæd]~*[ʤæt] **jad*~**jat* it was expected that participants would be primarily guided by the acoustic information provided to them by the vowel duration, due to the absence of lexical bias. The results for the native English participants were statistically significant (P = .0085), but the trend was not as consistent as that which was demonstrated for the word/word pair. There is overall difference of 21% in the number of voiced responses perceived between the longest and shortest gates.

The end-point for voiced responses is also under 50% which is not as expected. In the case of the German participants, both groups perceived more voiceless responses overall. Once again, this is most likely due to them being guided by their L1 phonology. However, interestingly, native English participants also perceived a majority of voiceless responses, though to a lesser degree than the German participants. The results for the German (UK) participants proved statistically insignificant (P = .9230) with no obvious trend across the vowel gates, and no difference in the percentage of voiced responses recorded between the longest and shortest gates. Conversely, the results for the German (Germany) participants did prove statistically significant (P = .0183) with a 30% difference in the number of voiced responses perceived across gates 1 and 8. As with [dæd]~*[dæt] *dad*~**dat*, this is not as expected as we anticipated that German participants who have received more exposure to spoken English would demonstrate more differentiation than German participants who have received less exposure.

Finally, we turn to the analysis for the nonword/word pair *[væd]~[væt] **vad*~*vat*. Here, we expected the opposite trend to the word/nonword pair. More specifically, it was expected that more voiceless responses would be recorded overall. This is because the word [væt] *vat* is expected to cue more voiceless responses than the nonword *[væd] **vad*. However, this does not appear to have been the case for the native English speakers. The results across the vowel gates for the English participants were not statistically significant (P = .1111) and though the end-points exhibit a difference of 25% in the perception of voiced responses, there is no clear coherence across the interim gates. German participants who were based in the UK did demonstrate a significant difference across the vowel gates (P = .0075), with a 33% difference in the number of voiced responses perceived between the longest and shortest vowel gates. These participants perceived an overall higher number of voiceless responses, with the only exception being at gate 2. German participants based in the Germany demonstrated results with

no significant difference across the vowel gates (P = .5300) and just a 3% difference in the number of voiced responses perceived across gates 1 and 8. Overall, more voiceless responses were again perceived by this group, with the only exception being at gate 3. Once again, it appears that the German speakers are being guided by their L1 phonology, with German speakers based in the UK demonstrating more differentiated end-points than German speakers based in Germany.

5.4.3.2 Affricates

[bædʒ]~[bætʃ] *badge~batch*

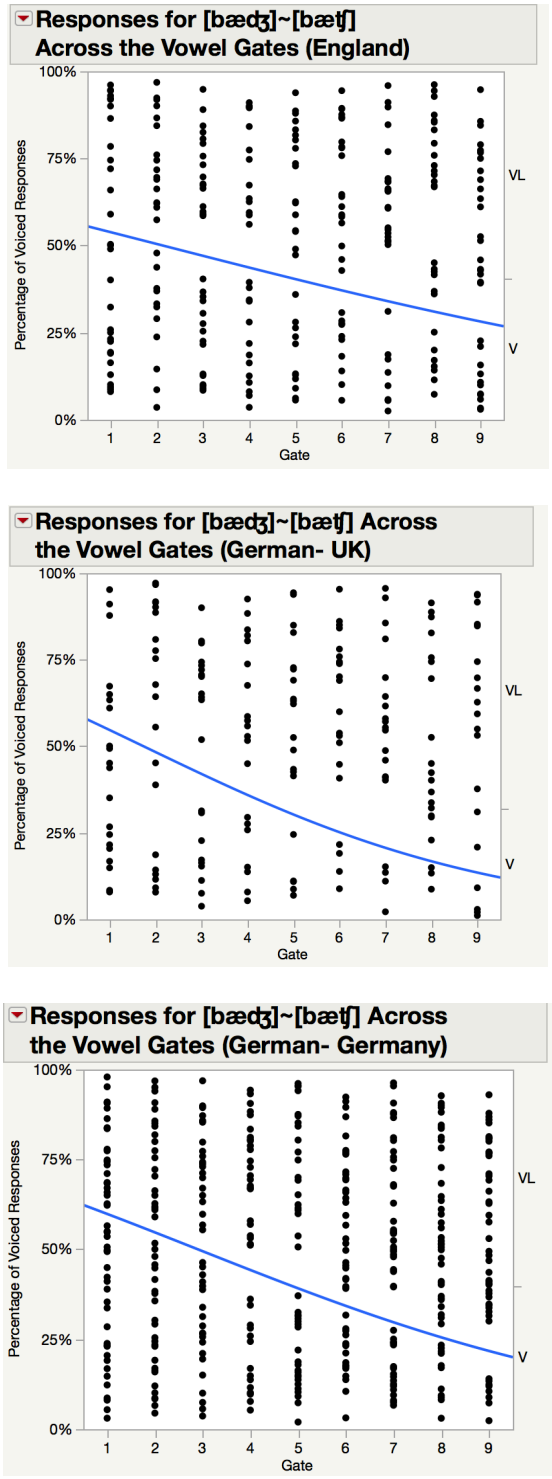


Figure 61: Voiced and voiceless results across the nine vowel gates for [bædʒ]~[bætʃ] *badge~batch* for Experiment 4

[rædʒ]~*[rætʃ] *radge~ratch*

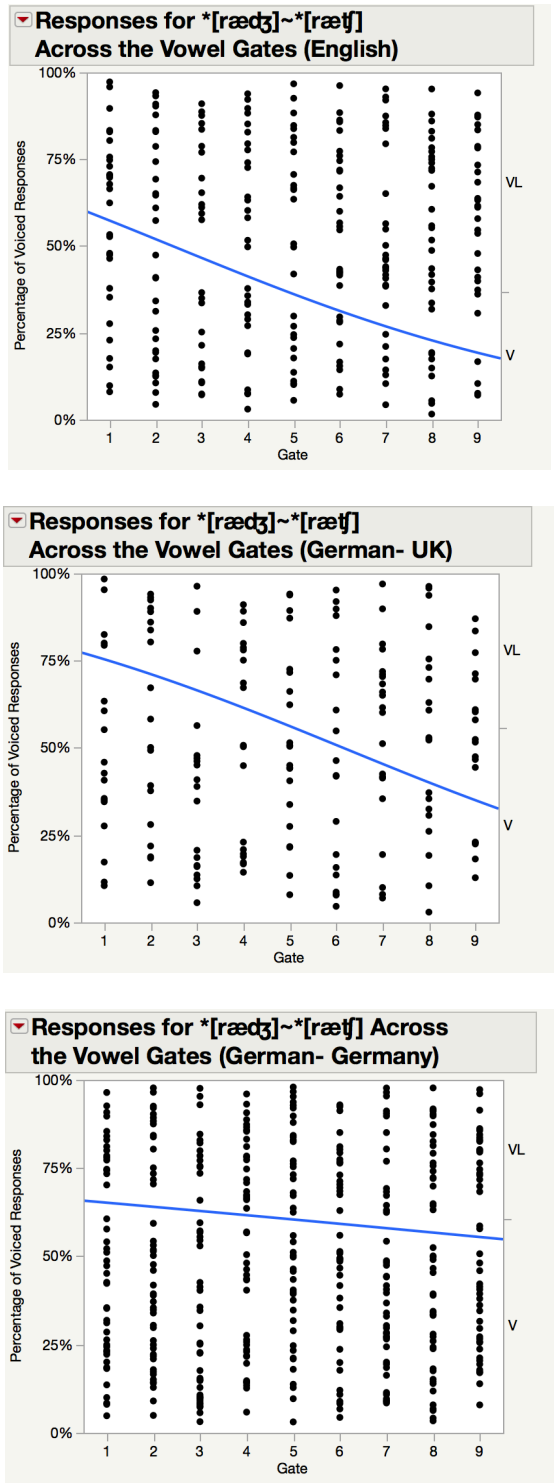


Figure 62: Voiced and voiceless results across the nine vowel gates for *[rædʒ]~*[rætʃ] **radge~*ratch* for Experiment 4

Table 28: The percentage of voiced and voiceless responses across the nine vowel gates for [bædʒ]~[bætʃ] *badge~batch* according to the three language groups for Experiment 4

English			German- UK			German-Germany		
Gate	**% V**	**% VL**	**Gate**	**% V**	**% VL**	**Gate**	**% V**	**% VL**
1	58	42	1	65	35	1	55	45
2	43	57	2	40	60	2	54	46
3	48	52	3	43	57	3	53	47
4	47	53	4	33	67	4	37	63
5	43	57	5	29	71	5	54	46
6	37	63	6	24	76	6	31	69
7	27	73	7	20	80	7	34	66
8	23	77	8	14	86	8	22	78
9	37	63	9	20	80	9	18	82

Table 29: The percentage of voiced and voiceless responses across the nine vowel gates for *[rædʒ]~*[rætʃ] **radge~*ratch* according to the three language groups for Experiment 4

English			German- UK			German-Germany		
Gate	**% V**	**% VL**	**Gate**	**% V**	**% VL**	**Gate**	**% V**	**% VL**
1	48	52	1	75	25	1	63	37
2	53	47	2	55	45	2	67	33
3	50	50	3	86	14	3	71	29
4	47	53	4	55	45	4	52	48
5	43	57	5	62	38	5	53	47
6	29	71	6	57	43	6	60	40
7	22	78	7	38	62	7	65	35
8	27	73	8	40	60	8	60	40
9	14	86	9	33	67	9	52	48

The individual trends described in this section are depicted in Figures 61-62 (above), and the detailed percentages of the responses for each of the language and word-initial consonant groups can be found in Tables 28-29 (above).

Let us first consider the results from the group of native English participants. There is a statistically significant difference for both the word/word pair [bædʒ]~[bætʃ] *badge~batch* (P = .0046) and the nonword/nonword pair *[rædʒ]~*[rætʃ] **radge~*ratch* (P < .0001). This difference is significantly high in the case of the nonword/nonword pair. In the case of word/word pair, participants selected a voiced [dʒ] response in 58% of cases at the longest gate (gate 1), 10% more than they selected in the case of the nonword/nonword pair. However, there

is only a 21% difference between the number of voiced responses selected at the longest and shortest gates for the word/word pair. With the exceptions of gates 2 and 9, there is a steady decline in the number of voiced responses as the vowel gates get shorter; therefore, these results overall support the predictions outlined for Experiment 4.

Conversely, for nonword/nonword pair there is a 34% difference in the number of voiced responses perceived at the longest and shortest vowel gates. However, the results for the first four gates linger around the 50% mark, which is again a much lower maximum percentage than was expected. Nevertheless, from gate 4 onward, with the exception of gate 8, there is a steady decrease in the number of voiced responses perceived as the gates get shorter in duration.

Turning now to consider the two groups of German results, in the case of the word/word pair both groups demonstrated highly significant statistical differences across the gates ($P < .0001$). However, in the case of the nonword/nonword pair, only the German (UK) results produced a significantly different result across the gates ($P = .0003$), whereas the German (Germany) results did not ($P = .1783$). In the case of both the word/word and nonword/nonword pairs, the group of German (UK) participants demonstrated results more in line with those which were expected for the native English participants. For the word/word pair, the German (UK) participants demonstrated a 45% difference between the shortest and longest vowel gates, and a 42% difference for the nonword/nonword pair. Conversely, the German (Germany) speakers exhibited a 37% difference in the case of the word/word pair, and just an 11% difference in the case of the nonword/nonword pair. Interestingly, the German (UK) participants produced results which matched the predictions for the group of native English speakers more closely than the native English speakers did, demonstrating the largest range of voiced responses between the longest and shortest gates overall.

It does not appear that the lexical status has affected the results for native English or German (UK) participants. However, for German participants based in Germany, responses for the word/word pair demonstrated a steady decline in the number of voiced responses perceived at each gate as the vowel durations got shorter, with the exception of an increase at gate 5. Conversely, for the nonword/nonword pair, these participants demonstrated no coherent trend and produced results which showed no statistical significance. It could therefore be posited that this may have been caused by the absence of an L2 lexical representation for the nonwords, and as such L2 perceptual cues from the vowel duration could not be utilised.

5.4.4 Results according to sex

As with Experiment 3, an uneven number of males and females across took part in Experiment 4. This section of the analysis will therefore once again consider a cross-section of five males and five females randomly chosen from each language group to determine whether sex causes an overt effect on the behaviour and perception of the participants. As with Experiment 3, the participants were selected using a random number generator in *Microsoft Excel*. For the group of German participants recorded in the UK, only two male speakers took part. Therefore, though no statistical significance can be drawn from this small number of participants, these results will be incorporated to give a fuller overview.

This analysis according to sex will be on a smaller scale than the full analysis, and will focus only on the overall results for stimuli ending in word-final affricates and stops. The individual word pairs will not be included in this analysis, as the general trend for male and female responses can be gauged from the overall results for the each of the two word-final consonant categories.

5.4.4.1 Stops

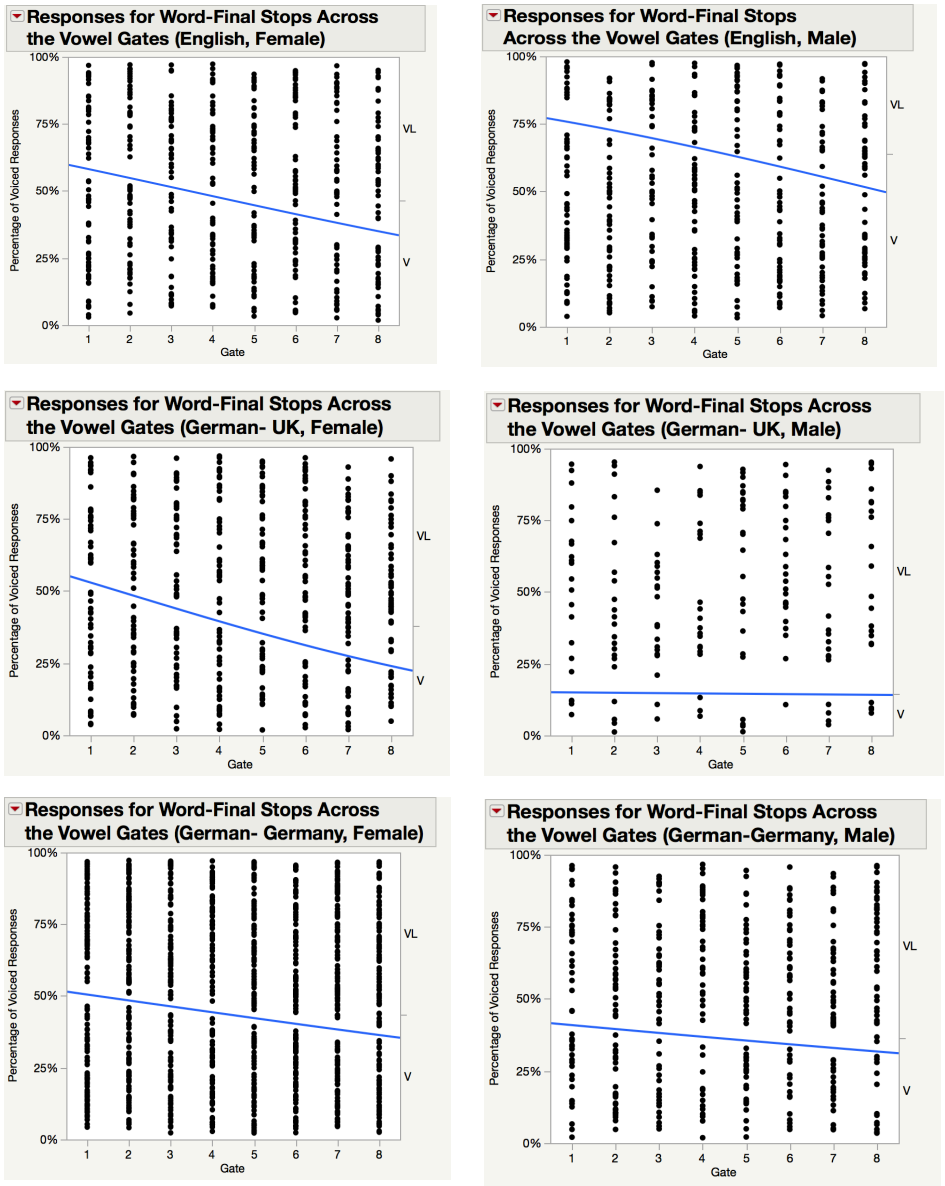


Figure 63: Voiced and voiceless responses across the eight vowel gates for word-final stops according to participant sex for Experiment 4

Table 30: The percentage of voiced and voiceless responses across the eight vowel gates for word-final stops according to the three language groups and participant sex for Experiment 4

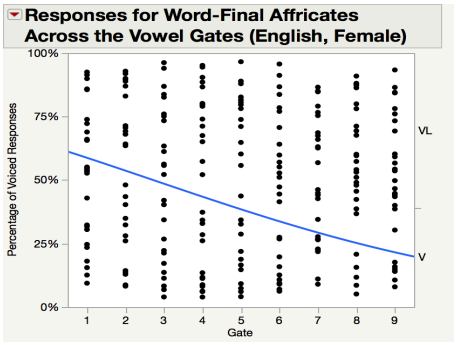
English						German- UK						German-Germany					
	Male			Female			Male			Female			Male			Female	
Gate	%V	%VL	Gate	%V	% VL	Gate	%V	%VL	Gate	%V	%VL	Gate	%V	%VL	Gate	%V	%VL
1	76	24	1	61	39	1	17	83	1	53	47	1	42	58	1	50	50
2	72	28	2	56	44	2	17	83	2	50	50	2	41	59	2	47	53
3	67	33	3	46	54	3	10	90	3	41	59	3	37	63	3	46	54
4	69	31	4	52	48	4	17	83	4	42	58	4	29	71	4	47	53
5	59	41	5	43	57	5	17	83	5	34	66	5	43	57	5	43	57
6	61	39	6	38	62	6	4	96	6	29	71	6	29	71	6	43	57
7	64	36	7	34	66	7	17	83	7	28	72	7	42	58	7	38	62
8	46	54	8	41	59	8	19	81	8	25	75	8	27	73	8	34	66

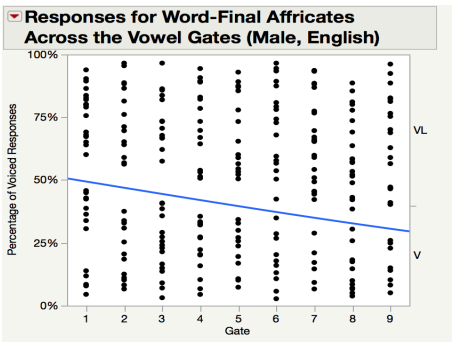
Each of the trends discussed here are illustrated in Figure 63 (above) and Table 30 (above) which contains the percentage of voiced and voiceless responses at each of the nine vowel gates for the three language groups according to sex.

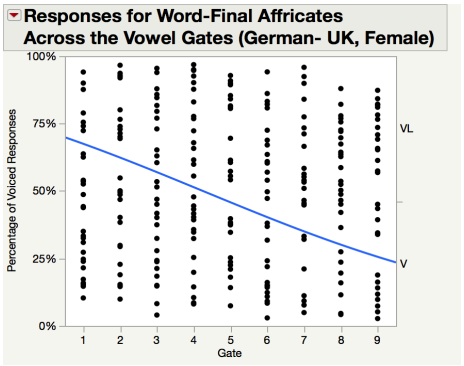
First considering the native English responses, we see that female participants ($P = .0005$) show a consistent decline in the number of voiced responses perceived as the vowel gates get shorter, with exceptions at gates 4 and 8. Overall, the trend in female responses was as expected, and they demonstrated a 20% difference between the end-point stimuli. Conversely, the male participants ($P = .0005$) perceived a majority of voiced responses throughout, with the exception of gate 8. However, we do see a 30% difference between the end-point stimuli; a higher differentiation than for the female participants, though this difference exhibits the same significance.

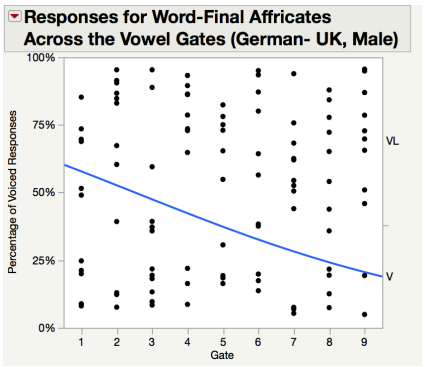
Neither male German participants based in the UK ($P = .9170$) nor Germany ($P = .1875$) demonstrate statistically significant differences across the vowel gates. Instead, both groups of male participants indicate perceiving a majority of voiceless responses across the gates. This is particularly true for the German (UK) participants. Conversely, the perception of the German female participants is more differentiated. In the case of female German participants based in the UK ($P < .0001$), a majority of voiceless responses are still indicated with the exception of gate 1. However, here there is a decrease of 28% in the number of voiced responses perceived at the longest and shortest gates, which is the highest percentage difference demonstrated across all German groups. Female German participants based in Germany ($P = .0022$) also show a significant difference across the vowel gates, albeit less significant than for the female German (UK) speakers. A majority of voiceless responses are perceived.

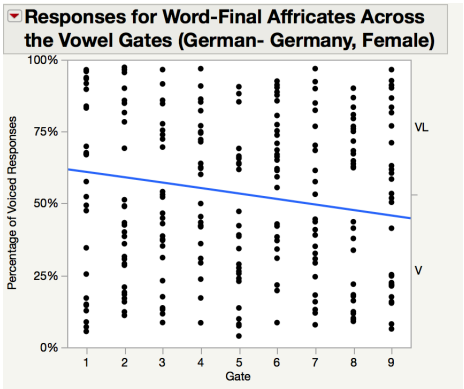
5.4.4.2 Affricates











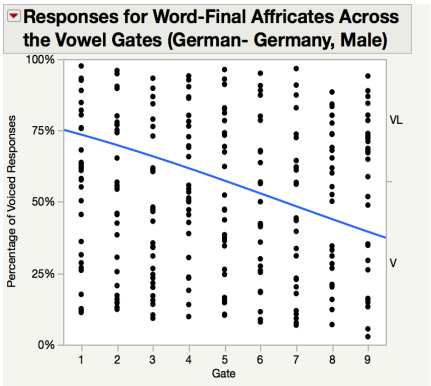


Figure 64: Voiced and voiceless results across the nine vowel gates for word-final affricates according to participant sex for Experiment 4

Table 31: The percentage of voiced and voiceless responses across the nine vowel gates for word-final affricates according to the three language groups and participant sex for Experiment 4

English						German- UK						German-Germany					
	Male			Female			Male			Female			Male			Female	
Gate	%V	%VL	Gate	%V	% VL	Gate	%V	%VL	Gate	%V	%VL	Gate	%V	%VL	Gate	%V	%VL
1	43	57	1	60	40	1	64	36	1	72	23	1	70	30	1	52	48
2	44	56	2	48	52	2	33	67	2	54	46	2	66	34	2	69	31
3	55	45	3	44	56	3	75	25	3	60	40	3	76	24	3	64	36
4	46	54	4	48	52	4	27	73	4	50	50	4	57	43	4	44	56
5	41	59	5	46	54	5	42	58	5	47	53	5	60	40	5	63	37
6	34	66	6	36	64	6	25	75	6	47	53	6	58	42	6	32	68
7	24	76	7	30	70	7	25	75	7	31	69	7	50	50	7	63	37
8	37	63	8	19	81	8	33	67	8	24	76	8	41	59	8	50	50
9	32	68	9	22	78	9	18	82	9	30	70	9	37	63	9	41	59

Each of these trends discussed here are illustrated in Figure 64 (above) and Table 31 (above) which once again contains the percentage of voiced and voiceless responses at each of the nine vowel gates for the three language groups according to sex.

Let us first consider the native English responses. The female responses demonstrate end-points that are more in support of predictions than the male responses, with a higher degree of differentiation in the percentage of voiced responses recorded at the longest gate (gate 1) and a higher percentage of voiceless responses recorded at the shortest gate (gate 9). The female participants thereby produced a significant difference across the gates ($P = .0001$). Conversely, male responses ($P = .0446$) indicate less significance, with only an 11% difference between the end-point stimuli, compared with a 38% differentiation for female participants.

For German speakers based in the UK, both male ($P = .0102$) and female ($P < .0001$) responses show a significant difference in the percentage of voiced responses across the gates, 46% and 42% respectively. This significance is particularly high for female speakers. Therefore, the prediction that German speakers would consistently perceive more voiceless responses does not appear to have been the case for male nor female participants in this group.

In the case of German participants based in Germany, male responses ($P = .0002$) demonstrate a more differentiated perceptual curve with a 33% difference between the end-point stimuli. The female responses were not statistically significant ($P = .1146$) with an 11% difference between the end-point stimuli. For German speakers based in Germany, the expected predictions were also unsupported in that neither male nor female participants perceived a majority of voiceless responses overall.

5.4.4.3 Summary of results according to sex

Based on this subsection of results, it appears that male and female participants have behaved differently. For native English participants, the female responses more consistently support the predictions outlined for Experiment 4 than the male responses for word-final affricates, though males demonstrated more differentiated end-points for word-final stops. For the German participants based in the UK, with regard to stimuli ending in word-final stops, male participants fell more in line with predictions by perceiving a majority set of voiceless responses. For female participants, a majority of voiceless responses were also recorded, though they exhibited a more differentiated perceptual curve across the vowel gates. Both male and female responses demonstrated significant differences across the vowel gates for affricates. Conversely, for German participants based in Germany, both males and females perceived a majority of voiceless responses across the vowel gates for stimuli ending in word-final stops. Female participants demonstrated a significant difference across the vowel gates, whilst male participants did not. In the case of affricates, female participants did not exhibit a significant difference across the vowel gates, whereas males did. Neither group perceived a majority of voiceless responses.

In conclusion, controlling for the sex of participants is an important consideration to make in planning an experiment. Though the above differences in results are not necessarily a direct result of a participant sex alone, this could be a consideration for future studies. Ensuring a similar number of male and female participants creates more robust support for any overarching linguistic patterns found in a more generalised analysis, such that these findings cannot be explained by sociolinguistic factors such as sex. As such, Experiment 5 contained the same number of male and female participants.

5.5 Limitations of the study

There are two primary limitations acting on the results of Experiment 4; the methodology used to generate the stimuli, and the total number of participants who took part.

Firstly, it seems that there are issues with using computationally-generated speech to form the basis of the stimuli. This appears to be the case given that the data trends are generally in support of predictions, but the extent to which these trends are demonstrated are consistently less than expected across the results. Experiment 4 utilised this method of stimuli formation to address the limitations of using recorded speech in Experiments 2 and 3. As stated previously, the use of recorded speech may cause issues with the reduced degree to which the variation between stimuli can be controlled. Specifically, the volume, pitch, and intonation cannot easily be made uniform across the stimuli using recorded speech due to its natural variation. As such, it was unclear as to whether there could be phonetic information encoded within these aspects which may provide cues for the presence or absence of the word-final [VOICE] feature, for instance in the word-initial consonant (Coleman, 2003). Comparatively, the stimuli in Experiment 4 were all completely uniform, with the only controlled differences occurring in the duration of the vowel. However, having now analysed these results, it appears that there are also significant issues with using synthetically produced speech in perception experiments. The stimuli in Experiment 4 appear to be limited in the extent to which participants could perceive differences in the word-final ambiguous consonants. This may well be because the stimuli lack a certain *human* quality. As previously discussed, all participants underwent a training block which intended to familiarise them with the sound of synthetic speech, however this short practice session may not have been sufficient. Therefore, although there are issues in controlling the parameters of recorded speech, this method may be preferable in perception experiments such as this. When conducting future research, it is important to consider the

extent to which human participants can perceive differences in synthesised speech, and the effect that this has on the overall accuracy of responses. Importantly, the data from native and non-native participants has been directly compared under the same experimental conditions. Therefore, though there may have been issues with the perception of the stimuli, this issue has been offset by the same degree between the language groups. Directly comparing native and non-native data in this way has enabled more robust conclusions to be drawn.

Secondly, there were issues with the recruitment of participants for Experiment 4. The fieldwork conducted in Germany yielded nineteen participants, and obtaining native German speakers in the UK proved difficult, with just nine participants taking part. Similarly, only eleven native English speakers participated. A higher and more even number of participants would be required to make more widespread inferences on the basis of these results. The results of Experiment 4 should be therefore be considered as preliminary findings for a future, larger-scale, study.

5.6 Discussion

The trends noted in the analysis will now be further discussed. Let us begin by considering the results of Experiment 4 on the basis of vowel duration. Following this, the role of the lexicon will be examined.

The durational differences between the shortest and longest gates for the voiced~voiceless pairs are indicated in Table 32 (below).

Table 32: The durational difference between the shortest and longest vowel gate for each stimuli pair for Experiment 4

Word-token	IPA	Durational Difference (ms)
badge~batch	[bædʒ]~[bætʃ[	79
*radge~*ratch	*[rædʒ]~*[rætʃ]	78
bad~bat	[bæd]~[bæt]	68
dad~*dat	[dæd]~*[dæt]	68
*jad~*jat	*[dʒæd]~*[dʒæt]	69
*vad~vat	*[væd]~[væt]	68

Recall that the stimuli in Experiment 4 consisted of the same tokens that were employed in Exp1b. In this production study, native English speakers consistently produced the greatest variation in vowel duration; producing longer vowels prior to voiced word-final consonants and shorter vowel durations prior to voiceless word-final consonants. It was therefore expected that the group with the greatest sensitivity to vowel duration as a cue for word-final voicing would be the native English participants. However, this was not consistently found to be the case. Instead, for stimuli ending in stops, English participants demonstrated the least difference in perception. In light of the results from Experiment 2 and 3, which suggest that English participants are astute at using vowel duration as an indication of word-final voicing, it seems likely that these findings are due to confusion caused by the synthesised nature of the stimuli. However, in the case of word-final affricates, English participants do demonstrate a high degree of differentiation across the gates. It may therefore be inferred that the nature of voicing for word-final affricates is more easily determined than that of word-final stops. A larger scale study would be necessary to support this notion. It may also be inferred that the confusion caused by the sound of the synthesised stimuli had a greater effect on the English data than the German data. This is likely due to German speakers being more disconnected from their non-native language, and therefore more readily able to focus on the experimental stimuli.

Regarding the German participants, it was expected that there would be a positive correlation between the amount of spoken English that the participants had been exposed to, and the extent to which these participants would be guided by the vowel duration. This prediction was based on the findings of Exp1b. German participants based in the UK demonstrated the highest degree of differentiation across the vowel gates in the case of both word-final stops and affricates. Of the three language groups, German participants based in Germany showed the least percentage differentiation across the vowel gates for stimuli ending in affricates, and in the case of word-final stops, exhibited a smaller perceptual curve than German participants based in the UK. These results support the prediction that the extent to which non-native participants are sensitive to vowel duration as a cue for voicing is likely to increase with the amount of exposure to spoken English they have received.

Participants from both German groups indicate perceiving a majority of voiceless responses throughout the vowel gates for word-final stops. However, interestingly, their results are more differentiated for word-final affricates. As such, German participants appear to be being guided by the L2 perceptual cues more in the case of word-final affricates, compared with word-final stops. Recall that, though German phonology only permits voiceless word-final stops in the surface form, it does contain both underlyingly voiced and voiceless stops. Conversely, German does not contain voiced affricates underlyingly, unless borrowed into the language. Therefore, it could be that in the case of word-final stops, German participants perceive less differentiation between the longer and shorter gates as they are subconsciously familiar with the presence of underlyingly voiced stops. They are therefore being primarily guided by their L1 phonology to produce an overall majority of voiceless responses; as they would if the stimuli were in German. Conversely, they would not be familiar with underlyingly voiced affricates. Therefore, they may demonstrate more differentiation between the gates for the

word-final affricate stimuli due to being influenced to a greater extent by the L2 acoustic information being provided by the vowel duration.

We will now turn to consider the results of the individual word and nonword pairs. Let us first interpret the results from stimuli ending in word-final stops. In the case of the real word pair [bæd]~[bæt] *bad~bat*, the results were as expected. Lexical bias was not relevant in the case of this pair, as both tokens were real words. The native English participants produced the greatest differentiation across the vowel gates, and as expected there was no obvious shift in the direction of either [bæd] *bad* or [bæt] *bat*. Conversely, German participants perceived a majority of voiceless responses overall at each vowel gate. This demonstrates that German speakers are primarily guided by their L1 phonology, leading them to perceive more voiceless responses overall. For German participants based in the UK, a significant difference across the vowel gates was obtained. However, for the German based participants, no significant difference was found. These results therefore once again support the prediction in that a higher degree of exposure to spoken English appears to lead to an increase in L2 sensitivity to vowel duration as a cue for word-final voicing.

In the case of [dæd]~*[dæt] *dad~*dat*, all three groups demonstrated significant differences in the perception of word-final voicing across the vowel gates. It was expected that there would be a higher number of voiced responses overall as participants were expected to be biased by the lexical status of [dæd] *dad*, as opposed to *[dæt] **dat*. This was the case for native English participants, who produced a majority of voiced responses across vowel gates. German participants showed no evidence of lexical bias, and demonstrated a decline in the number of voiced responses as the vowel gates got shorter. These results indicate that native English participants are influenced by the lexicon, and provide evidence that this influence may

override the acoustic information from the vowel duration in determining word-final voicing. However, German participants do not demonstrate this tendency.

Turning now to the nonword/nonword pair, *[dʒæd]~*[dʒæt] **jad~*jat*, it seems that an absence of any lexical representation has caused confusion across the language groups. It was expected that this stimuli pair would produce results similar to the real word pair, with the vowel duration causing a primary influence on the results in the absence of any lexical bias. However, all three language groups perceived a majority of voiceless responses across all vowel gates. English participants showed the greatest amount of differentiation according to the p-values, followed by the German participants based in Germany. German participants based in the UK did not exhibit a significant difference. The reason for these patterns are likely to be due to the absence of a real word causing overall confusion and the need for extra processing, and as such these results suggest that both acoustic information and lexical representation are important in establishing word-final voicing in an experimental setting. As previously discussed in Experiment 2, this notion may once again support research conducted by Rubin et al. (1976) who found that a phoneme is detected more quickly when it is part of a word than if it appears as part of a nonword. Here, it is possible that the participants are not allowing the creation of a nonword to influence their categorisation of the word-final phoneme.

The final stop-pair, *[væd]~[væt] **vad~vat*, was expected to cue more voiceless responses overall due to a lexical bias in the direction of the word [væt] *vat* as opposed to *[væd] *vad*. This was not found to be the case for the English participants, who perceived a majority of voiced responses across the vowel gates, with the only exception found at the shortest gate. As in the previous cases, both groups of German participants selected a majority of voiceless responses across the gates, though this is more likely to be guided by their L1 German phonology more-so than an interaction with their L2 English lexicon. For the group of native

English participants, it could be that word frequency is at play.[23] Consider that [væt] *vat* is a lower frequency word in English. Therefore, participants may have been guided more-so by the acoustic information provided by the vowel duration than the influence of the lexicon. This was not the case for the higher frequency word [dæd] *dad* which did produce results in the expected direction. Relative frequency counts for these two word-tokens were five hundred and sixty-seven for [dæd] *dad*, and one hundred and ten for [væt].

Let us now turn to the results for stimuli ending in affricates. In the case of native English participants, there was more differentiation between the end points for */rædʒ/~*/rætʃ/ **radge*~**ratch* than for [bædʒ]~[bætʃ] *badge~batch*. However, more voiceless responses are perceived overall for both stimuli pairs. For both groups of German participants, we see more differentiation in the case of the real word pair as opposed to the nonword pair. For the real word pair, German participants based in the UK demonstrated a more significant difference in the perception of voicing across the vowel gates than English participants. German participants based in Germany also demonstrated a higher differentiation at the end-points of the data than the native English participants. It once again appears that an increased exposure to English leads to more native-like perception, as German participants based in the UK demonstrated a more significant difference in their results compared with German participants based in Germany. Regarding the nonword/nonword pair, of the two groups of German participants, those based in the UK once again demonstrated the greatest difference between the end-points of the data, with German-based participants demonstrating no significant difference. However, neither group have perceived a majority of voiceless responses in this case, with a majority of voiced responses being recorded for the Germany-based participants group throughout the gates, and up to gate 7 for the UK-based participants. This does not support the previously-

[23]The CELEX database was used to ascertain frequency counts

attested to notion that native speakers of German are more likely to perceive overall more voiceless responses due to their L1 phonology. However, these trends could once again be caused by the confusion of having no lexical representation present at all; as was inferred to be the case for *[ʤæd]~*[ʤæt] **jad*~**jat*. This is particularly true for German participants based in the UK, who in the absence of lexical bias, appear to have been increasingly guided by the acoustic information provided by the vowel duration.

To summarise, the extent to which the lexicon may interact with vowel duration in the perception of word-final voicing appears to be dependent on several factors. For English participants, with regard to word-final stops, word frequency looks to be an important factor. The word/nonword stimuli containing the higher frequency real word ([dæd] *dad*) demonstrates more of a lexical bias effect than the matched-pair containing the lower frequency real word ([væt] *vat*). Absence of a clear lexical entry also appears to have caused confusion in the case of *[ʤæd]~*[ʤæt] **jad*~**jat*. For the two groups of German participants, in the absence of any clear lexical bias, participants are primarily guided by their L1 phonology. This proved to be the case for word-final stops but not for word-final affricates. As previously discussed, this is likely to be due to the rules governing their underlying phonology, and the existence of underlyingly voiced stops but not affricates. In the case of affricates, German participants demonstrate more differentiated end-points which appear to have been driven by the L2 acoustic cue provided by the vowel duration. However, in the case of the nonword/nonword affricate pair, German participants based in Germany showed no coherence across their responses, and this is likely due to an absence of lexical representation combined, with less exposure to spoken English. Here, the differences between English and German listeners support Broersma (2005:3900) who concluded that the native production of non-native contrasts in unfamiliar positions *'may hardly ever be attained'*. In summary, though a

subtle lexical effect is found in the data based upon lexical status, these results do not indicate that the lexicon and lexical bias are able to consistently override the effect of vowel duration in all cases.

5.7 Conclusion

The results from Experiment 4 suggest that vowel duration has systematically influenced the participants' perception of word-final voicing to varying extents. Specifically, the extent to which this has been the case appears dependent on both the language group to which the participant belongs, and the category of the word-final consonant being investigated. However, the most crucial finding here is that across a vocalic difference of 68ms-79ms, participants demonstrate evidence that they can distinguish between the end-points of the stimuli as either voiced or voiceless, word-finally.

Overall, native English participants show a greater sensitivity to vowel duration as cue for word-final voicing than non-native English participants. However, there are some exceptions to this trend. For example, when looking at the overall results for stops we see that the German participants based in the UK demonstrated the most differentiation between the end-points of the stimuli, and English participants demonstrated the least. Here, the native English participants seem to have had issues with the perception of the synthesised stimuli, resulting in the non-native speakers outperforming them. These results are reminiscent of the Broersma (2005) study in which the native English participants were misled by uninformative vowel durations, whilst the Dutch participants performed in an English native-like manner.

Generally, the UK-based German participants demonstrated more sensitivity to vowel duration as a cue for word-final voicing than the Germany-based participants. Therefore, sensitivity to vowel duration as a cue for word-final voicing does appear to increase for L2

English speakers in direct correlation with how much exposure to the L2 language they have received. However, both groups of German speakers perceived more voiceless responses overall in the case of word-final stops than they did for affricates. Here speakers appear to be being guided by their L1 phonology, as underlyingly voiced stops exist in German phonology and underlyingly voiced affricates do not.

Overall, for native English participants, a lexical effect was found in the stimuli ending in word-final stops such that words of a higher frequency caused more of a shift in the direction of a real word than words of a lower frequency. Specifically, the stimuli pair [dæd]~*[dæt] *dad~*dat* caused more of a categorical shift towards the word [dæd] *dad* than the stimuli pair [væd]~*[væt] *vad~*vat* did towards [væt] *vat*. An effect of lexical status was also noted between the word/word pair [bæd]~*[bæt] *bad~*bat* and the nonword/nonword pair *[ʤæd]~*[ʤæt] **jad~*jat*. In both of these cases, it was expected that no lexical effect would occur due to the absence of lexical bias in either direction. As such, it was expected that there would be a gradual shift in the perception of voiced to voiceless responses around the centre of the vowel gates. This was the case for the word/word pair, but less so nonword/nonword pair in which mostly voiceless responses were perceived. This may be due to the absence of any lexical representation in the nonword/nonword pair causing processing difficulties when interacting with the acoustic cue provided by the vowel duration.

A clear lexical effect was not made apparent for either group of German participants in the case of word-final stops. In the case of word-final affricates, a weak lexical effect may be inferred as German participants based in Germany showed no significance in their results for the nonword/nonword pair. Again, this could be due to an absence of lexical representation. Conversely, the lack of prevalence of voiced affricates in underlying German phonology appears to have lead German participants based in Germany to be primarily guided by L2 cues

in the word/word pair. This is most likely due to them having less exposure to spoken English, compared with German participants based in the UK who have perceived a majority of voiceless responses in the case of the word/word affricate pair. This is likely to be due to UK-based participants' L1 phonology overriding L2 cues due to having had more exposure to spoken English.

Overall, Experiment 4 supports the notion that vowel duration acts as a primary cue for word-final voicing in English. Additionally, it provides a more nuanced understanding of how this cue may interact with L2 phonology, and the interaction that this process has with the lexicon.

5.8 Overall conclusions for chapters four and five

Overall, this series of forced choice identification tasks have produced some significant and important insights into the relationship between vowel duration, voicing, and the lexicon in both native and non-native English speech perception.

Using a variety of experimental methodologies to create stimuli, the three identification tasks have exhibited that vowel duration has had a consistent and systematic influence on the response of participants. This is in support of literature which characterises vowel duration as a primary perceptual cue for word-final voicing in English (Klatt, 1976; Denes, 1955, among others). The results also infer that the extent to which the lexicon can be demonstrated to override this acoustic cue is limited, and as such, challenge the results of iconic papers such as Ganong (1980). Exploring these relationships and the extent to which vowel duration can be demonstrated to cue word-final voicing in L2 English has also provided a more nuanced understanding of non-native speech perception. The findings of Experiment 4, as with Exp1b, suggest that there is an increase in participants' sensitivity to vowel duration as a perceptual cue for word-final voicing in speakers who have had an increased exposure to spoken English.

Additionally, the varying methodologies used across these three experiments aim to have provided useful insights for future studies to take into consideration; specifically, the benefits and limitations of using recorded and synthetic speech as the basis for stimuli in perception experiments. Based on the results it seems that recorded speech is preferable, as despite its variability, using synthesised speech as part of a perception experiment looks to cause additional processing difficulties for participants. Future studies may consider overcoming this issue by trialling different methods of computationally generating stimuli.

As previously discussed, this thesis does not deny that the nature of speech perception is complex, and that there are many possible cues aside from vowel duration that may influence

phoneme categorisation. However, despite these nuances, the results from this series of identification tasks do support the notion that vowel duration acts as both a primary and necessary cue for word-final voicing distinctions in English.

CHAPTER SIX

An exploration into the relationship between vowel duration and the lexicon in English word-final voicing: A lexical decision task for native speakers

6.1 Introduction to Experiment 5

Experiments 2 and 4 have previously investigated the interaction between acoustic information and the influence of the lexicon in perceiving the nature of English word-final voicing. The results of these studies suggest that vowel duration has a primary perceptual effect on its characterisation, and lexical bias has been considered as having a secondary effect, dependent on factors such as word frequency. Lexical representation has thus far proven to work alongside the acoustic information provided by the vowel duration, leading to more exaggerated effects of vowel duration in relation to word-final voicing when lexical representation is clear. Experiment 5 aimed to readdress this interaction with a specific focus on determining whether vowel duration can interact with the lexicon in establishing semantic relationships between auditory primes and visual targets.

Experiment 5 consisted of a lexical decision task with semantically related fragment priming. A lexical decision task involves priming participants with semantically or phonologically related auditory stimuli, before visually presenting them with a target string of letters forming either a real or nonsense word. Participants must subsequently identify this presentation as a word or nonword. The reaction time taken for a participant to make this decision is measured. Regarding semantic priming, previous studies have found that participants' reaction times are faster and more accurate in classifying a real word when the

target is preceded by a semantically related priming word, relative to a prime that is semantically unrelated (Meyer and Schvaneveldt, 1971; Neely, 1976).

In developing the methodology for Experiment 5, two key pieces of research were consulted. Despite these papers not being focused on the same field of research as Experiment 5, they provided invaluable guidance regarding the design and execution of the lexical decision task reported here. As such, this section gives a brief overview of these papers.

Roberts et al. (2013) aimed to decipher more information about how the specification of phonological information is stored in the mental lexicon. More specifically, is phonological information stored with a full specification or an underspecified representation? The paper focused on a possible asymmetry in the perception of word-medial mispronunciations, for example *[tɛmə] **temo*r, as opposed to [tɛnə] *tenor*. Two experiments were conducted; however only the first experiment was relevant in informing Experiment 5. It consisted of a semantically primed lexical decision task constructed using a Latin square design, such that no stimulus was repeated. The auditory primes were comprised of disyllabic words; correct pronunciations and mispronunciations. Participants consisted of thirty-three native speakers of British English. The visual targets were presented immediately following the offset of the auditory prime, after which participants were required to make their decision on whether the visual target was a real word in English, or not. Experiment 5 was specifically influenced by the incorporation of the Latin square method, the concept of using semantic priming, the ordering of the presentation of stimuli, and the nature of participants used in Roberts et al. (2013). As such, this piece of literature provided an invaluable resource in developing the methodology of the current study.

The second paper which informed Experiment 5 was written by Friedrich et al. (2008) and investigated neurophysiological evidence for underspecified lexical representations. Two

experiments were conducted. The results of the first experiment were in support of neurophysiological evidence that phonological information is stored in the form of underspecified representations in the mental lexicon. The second experiment subsequently consisted of a word fragment priming study. Native German participants were presented with the auditory onset of a word, followed immediately by the visual presentation of a real word or nonword. The real word visual targets were presented as fragments;

> 'In combination with the initial syllable fragment extracted from the spoken version of that word (identity condition); in combination with the syllable fragment extracted from the corresponding pseudoword (variation condition); once in combination with an unrelated word fragment and once in combination with an unrelated pseudoword fragment (control condition).
>
> (Freidich et al., 2008:1551)

This research, along with the findings from Experiment 3, informed the concept of using only the word-initial consonants and following vowel fragments of full words for Experiment 5. By presenting the participants with this fragment in both monosyllabic and disyllabic words, the activation of semantic relationships was systematically reliant on the interaction between the vowel duration and word-final voicing.

The literature discussed in this section informed the concept that auditory fragment priming using varying vowel durations may cue the relevant word-final voicing feature needed to complete word or nonword primes in English. Incorporating semantically related or unrelated visual targets was designed to allow for more nuanced inferences to be made regarding the interaction between phonetic, phonological, and lexical information in speech perception.

6.2 Research questions

Experiment 5 endeavoured to determine whether preceding vowel duration can prime the word-final voicing characteristic of an auditory prime, leading to the faster recognition of a semantically related visual target.

It was expected that participants would show significantly faster reaction times in identifying a real word target for the semantically related experimental primes than for the semantically unrelated control primes when there was a complementary relationship between vowel duration and word-final voicing. Conversely, when there was a mismatch between vowel duration and word-final voicing, the experimental primes were not expected to be significantly faster than the control primes.

6.3 Methodology

The recording: a thirty-six-year-old male speaker with a Southern British English dialect was recorded in a sound-proof booth in the Language and Brain Laboratory at the University of Oxford. A male speaker was chosen for the same reasons as detailed in Experiment 1. This recording was made using a Rode NT-USB microphone. *Audacity* was used to capture the recordings, using a mono 44.1khz sampling rate. One hundred and sixty words and nonwords were recorded in total, comprising of eighty pairs of corresponding voiced and voiceless word-final endings. For example, [ʧaɪːd] *chide* and *[ʧaɪt] **chite.*

Words were both monosyllabic and disyllabic; seventy-three pairs consisting of monosyllabic words, and seven pairs consisting of disyllabic words. As with the previous experiments in this thesis, the speaker was instructed to read out the list of words at a fixed speed and to control their pitch and intonation throughout the recording as much as possible. The series of words were recorded twice, and the second recording was used to form the stimuli

as the speaker had acclimatised to the format of the recording and their speech was less variable in this instance.

A note on the recording: this thesis has previously noted the limitations of using recorded speech as the basis for experimental stimuli. These limitations include the fact that experimenters have far less control of the variation in pitch, volume, and intonation of the resulting stimuli. Additionally, it is impossible to elicit completely naturalised speech under experimental conditions. Of course, the same word can never be produced more than once in an identical fashion, and this means that variation between one potential stimulus base and another is unavoidable.

However, the issues faced in Experiment 4 regarding the clarity of the synthesised stimuli were judged to be of greater detriment to the results overall than the limitations of using recorded speech. Though the overall results of Experiment 4 supported the proposed predictions, the extent to which this was the case was believed to be limited by the extra processing time required for participants to reconcile the sound of synthesised speech and match it with a human equivalent. Therefore, for Experiment 5 recorded speech was considered to be the best method of forming the stimuli. This was especially the case given the bimodal nature of Experiment 5, and the additional processing required of the participant to not only be primed by a word-initial consonant and vowel fragment, but also to match this with a semantically related lexical target.

The stimuli design: participants were presented with the word-initial consonants and vowel fragment of a series of word pairs[24]. These fragment pairs formed the auditory primes. No

[24] The full list of stimuli can be found in Appendix F

word-final ambiguous consonant was incorporated into these primes. This formation of the stimuli was informed by the results from the previous three identification tasks. Specifically Experiment 3, which yielded the most accurate responses from participants. Experiment 3 contained stimuli consisting of CV:/CV fragments and did not incorporate an ambiguous word-final consonant. The primes employed in Experiment 5 contained an equal number of instances in which there was a complementary relationship between the vowel duration and word-final voicing necessary to cue a real word in English, or a mismatch between the vowel duration and word-final voicing. In the latter case, on the basis that vowel duration is a primary cue for word-final voicing in English, it was expected that the activation of a real word in English would be inhibited.

For example, participants were presented with [ʧaɪː] designed to cue *chide* when presented with the visual target *scold* or *[ʧaɪ] designed to cue **chite*, thereby inhibiting the activation of a real word due to the shorter vowel duration. As these primes were auditory, no orthographic condition was presented. However, participants were presented with an orthographic visual target which consisted of either a nonword or a real word which was either semantically related or unrelated to the auditory prime. An equal number of semantically related and unrelated real word targets were presented, and these two conditions were matched with the same number of nonword targets. Having been presented with an auditory prime, participants were tasked with identifying the target as a real or nonsense word in English. Further details regarding the selection of these auditory primes and visual targets are discussed below.

The primes and targets: the auditory stimuli consisted of pairs of experimental primes and control primes which related to the same visual target. Here, the experimental primes were

those which were semantically related to the target, whereas the control primes were semantically unrelated to the target. Each pair contained a complementary and mismatched instance of the relationship between vowel duration and voicing.

The auditory primes were therefore differentiated on the basis of *lexicality* and *relatedness.* Here, lexicality refers to whether the stimuli consisted of a fragment expected to activate a word or nonword in English. For example, [ʧaɪː] *chide* or *[ʧaɪ] **chite*. Recall that the real word primes were comprised of a vowel length which was complementary to the word-final voicing feature required to activate a word in English. Conversely, the nonword primes contained a mismatch between the vowel length and word-final voicing feature required to activate a word. Relatedness refers to whether the fragment was designed to cue a prime which was semantically related or unrelated to the target. For example, the experimental primes [ʧaɪː]/*[ʧaɪ] were presented in relation to the target *scold* which is semantically related to *chide.* The control primes [ʤæː]/*[ʤæ] were also presented in relation to the target *scold* which do not have a semantically related connotation.

Recall that participants would be asked to identify the visual target as a real word or nonword. As such, it was predicted that the identification of the target word would be fastest when there was a complementary relationship between vowel duration and voicing and a semantic relationship between the prime and target.

To investigate a range of vocalic contexts, it was important that an even spread of vowels from across the vowel space were incorporated. Therefore, the primes consisted of eighty stimuli, each containing twenty of the following vowels; /æ/, /eɪ/, /əʊ/, and /aɪ/. In addition to incorporating a range of vowels, the endings of the word and nonword primes were designed to cue a range of forty-six stops, twenty-eight fricatives, and six affricates. This enabled the

results of this study to be representative of a range of phonetic contexts and ensured that the findings of Experiment 5 would not be constrained by a set of stimuli that were too narrow.

The control primes were both phonologically and semantically unrelated to the experimental primes. They contained both a different vowel nucleus and were designed to cue a different word-final minimal pair to that of the corresponding experimental prime.

The CELEX database was once again used to ascertain the word frequencies of the desired prime, along with noting other lexical items which may be elicited by the same fragment. For example, the prime [ʧaɪː] may activate [ʧaɪːd] *chide*, but it may also activate [ʧaɪːv] *chive* or [ʧaɪːm] *chime*. Here, it was important to establish that there was a strong semantic relationship between the desired experimental prime and the visual target, and no such semantic relationship possible for the control prime.

In order to achieve this, an online relatedness survey was conducted which employed a Likert scale. Each of the auditory primes and visual targets were presented in a list format with a corresponding seven-step scale. For example, *chide* and *scold*. On the seven-step scale, one represented pairs which were judged to be totally unrelated, and seven represented words which were judged to be synonymous with one another. Twenty-four participants took place in this survey; fifteen females and nine males. Twenty-one participants were between the ages of eighteen to thirty, one was between the ages of thirty-one to fifty, and one participant was aged over fifty-one. Twenty-two participants were native English speakers, and two were second-language English speakers.

Following the relatedness survey, each word pair was assigned a score out of one hundred and sixty-eight; twenty-four participants rating each word pair out of seven. Then, the ten top-scoring voiced and voiceless word-final pairs were chosen to form the twenty experimental

prime-target pairs. The remaining sixty prime-target pairs were used for the control and nonword primes.

The following design was subsequently employed in Experiment 5 (Table 33, below).

Table 33: An example of the stimuli design used for Experiment 5

Experimental prime	**Experimental prime**	**Control prime**	**Control prime**
V	**VL**	**V**	**VL**
CHIDE	*CHITE	JAZZ	*JASS
[ʧaɪ:]	*[ʧaɪ]	[ʤæ:]	*[ʤæ]
SCOLD	SCOLD	SCOLD	SCOLD
Nonword target	**Nonword target**	**Nonword target**	**Nonword target**
ENGAGE	*ENGAICH	STOVE	*STOAF
[ɪngeɪ:]	*[ɪngeɪ]	[stəʊ:]	*[stəʊ]
*TROYP	*TROYP	*TROYP	*TROYP

This design involved devising twenty blocks of eight stimuli as detailed in Table 33 (above); four relating to real word targets and four relating to nonword targets. To reiterate, the targets were visually presented to the participants in full orthographic form and the primes were audibly presented.

Due to the design of a lexical decision task, the nonword targets were only necessary for balancing out the number of word and nonword targets that participants would be required to identify as real or nonsense words in English. Therefore, following the completion of Experiment 5, the results from these nonword targets were discarded as they did not contribute to the analysis of this study.

The stimuli: The one hundred and sixty stimuli, eighty word and nonword target pairs, were edited using *Praat*. The CV:/CV fragment for each of the auditory primes were selected. A boundary was placed at the end of the vowel, where both a visual and auditory change in energy

signalled the end of the final glottal pulse. As with the previous experiments, these fragments were then extracted and saved as individual files. As such, both the closure duration and release burst of the word-final consonants were omitted. The full vowel was retained despite the possibility of it containing perceptual cues for voicing. Again, it was deemed important for the nature of this research that the naturally occurring, unedited vowel duration be presented in full to participants. This process is illustrated in Figures 65-66 (below).

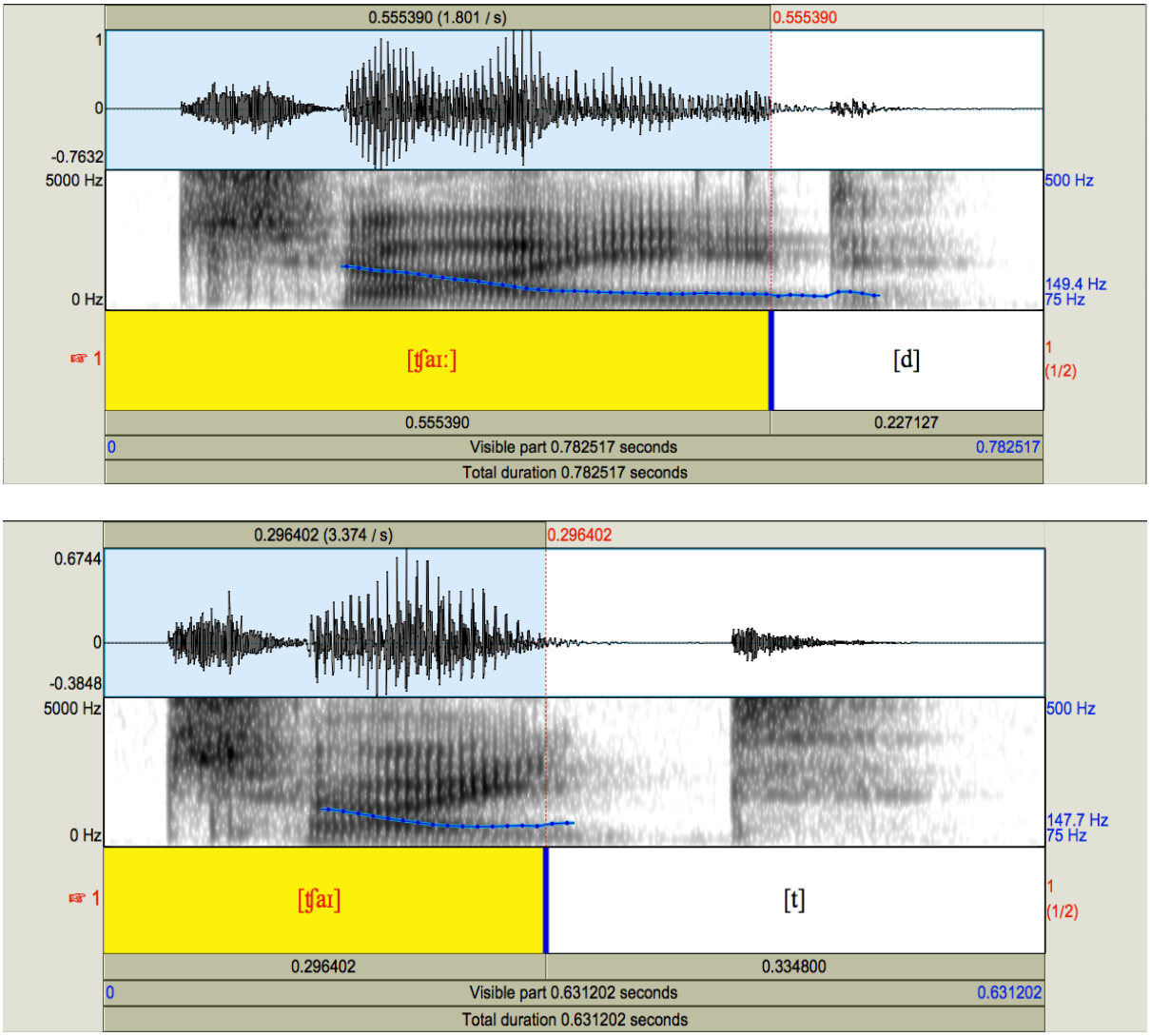


Figure 65: The use of *Praat* to segment the CV:/CV portion of each stimulus, here [ʧaɪː] *chide* and *[ʧaɪ] **chite*, for Experiment 5

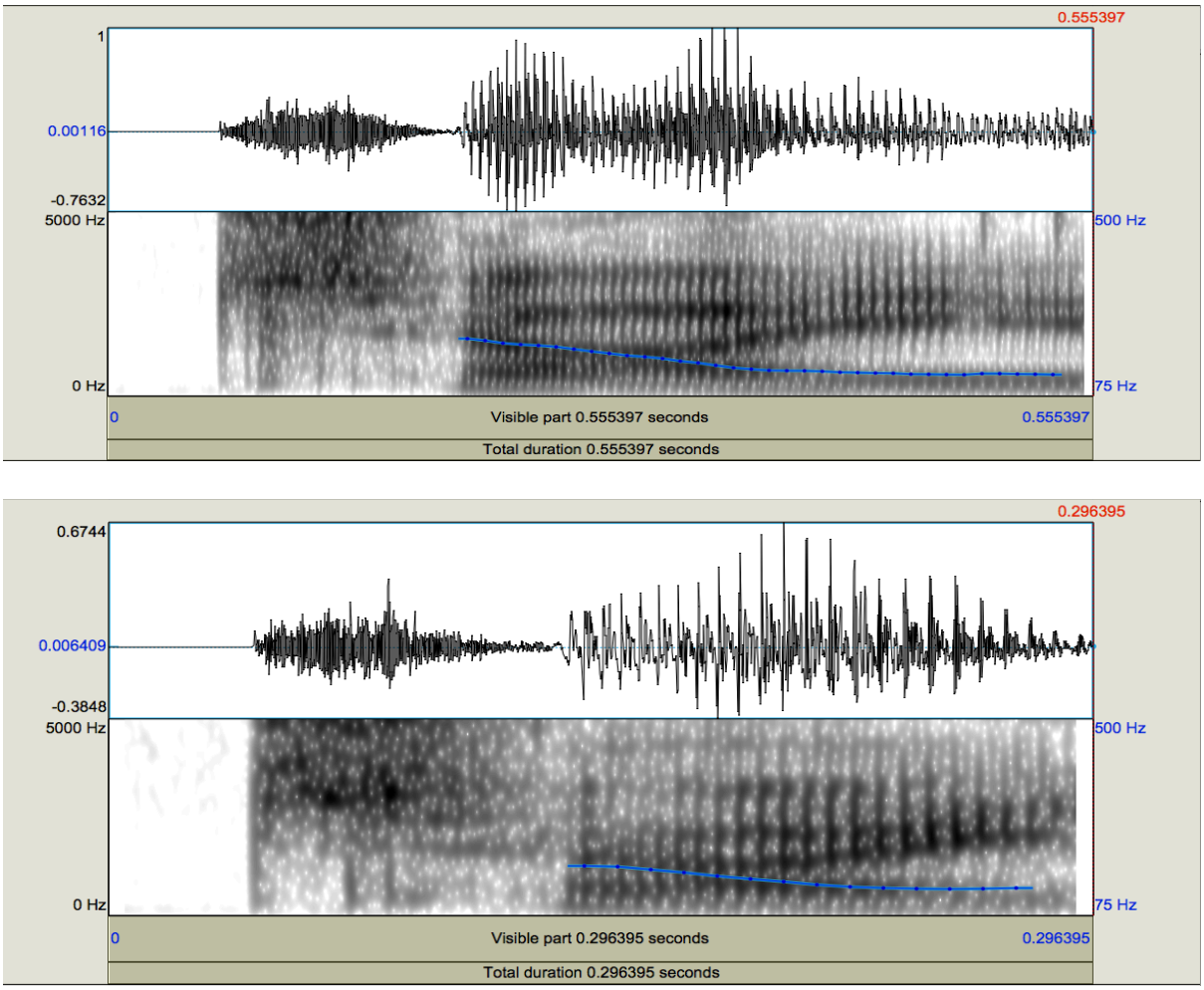


Figure 66: The extracted CV:/CV portion of each stimulus, here [ʧaɪ:] and [ʧaɪ] (Figure 65), for Experiment 5

The experimental audio file: *SPLICE* was utilised to form four separate *.wav* files containing an equal number of word and nonword targets organised into the following groups; WORD/RELATED, WORD/UNRELATED, NONWORD/RELATED, NONWORD/UNRELATED. An example of the correspondence of these groups to the experimental and control primes is demonstrated as follows, here all participants were presented with the visual target *scold:*

I. WORD/RELATED: [ʧaɪː] expected to activate *chide* (experimental)

II. WORD/UNRELATED: [ʤæː] expected to activate *jazz* (control)

III. NONWORD/RELATED: [ʧaɪ] designed to activate **chite* (experimental)

IV. NONWORD/UNRELATED: [ʤæ] designed to activate **jass* (control)

Each participant was assigned to one of the four groups, and as such only heard one of the four primes. To begin the experiment, participants were aurally presented with a series of five *BLEEP*s whilst simultaneously seeing a countdown from *5* to *1* on the computer screen in front of them. Participants then heard another *BLEEP* followed by a 500m *PAUSE*. The first audio prime was then presented, and immediately following the offset of the auditory prime participants saw the visual target on screen for 300ms. They were then given a 2500ms response window before the next trial began, as indicated by a *BLEEP*.

Participants: forty-eight participants took part in Experiment 5. All participants were native speakers of Southern British English. Twenty-four participants were female, and twenty-four participants were male. All participants were between the ages of eighteen and thirty. They had no known hearing or language disorders, and all had normal-to-corrected vision. One participant was left-handed.

As with previous experiments, participants were recruited from the University of Oxford. For Experiment 5 this recruitment was conducted in two ways, the first of which was by email and the second was through a research-pool of applications from the Experimental Psychology Department at the University of Oxford. Participants who signed up through the Psychology

Research Pool were granted 2 Credits towards their course, and all participants were reimbursed £5 for their time.

Procedure: participants were divided into the four groups; WORD/RELATED, WORD/UNRELATED, NONWORD/RELATED, NONWORD/UNRELATED.

Recall that participants in the WORD/RELATED group were primed with stimuli comprising of a complementary relationship between vowel duration and word-final voicing. They also saw semantically related visual targets. For example, in Table 33 (above) we see that *scold* was primed by [ʧaɪː], with a complementary longer vowel duration expected to cue a voiced [d], resulting in the activation of the word [ʧaɪːd] *chide.* For the same target, *scold*, the NONWORD/RELATED group of participants were primed with stimuli comprising of a mismatch between vowel duration and word-final voicing, [ʧaɪ] **chite*. For the control primes, the WORD/UNRELATED and NONWORD/UNRELATED groups of participants were presented with the same real word target, *scold*, however the auditory primes in both cases were semantically unrelated to this target. For instance, in Table 33 (above) we see that the initial CVː fragment of the real word [ʤæː] *jazz* is presented, along with the mismatched CV fragment of the nonword [ʤæ] **jass.*

As with the previous experiments, participants were invited to attend a session at the Language and Brain Laboratory, and each session lasted for around fifteen minutes. Again, up to eight participants could attend a session at any one time. Written consent from all participants was obtained prior to the start of each experiment[25].

Experiment 5 took place in a quiet room. Participants were invited to sit in front of a 17" CRT monitor, attached to which was a pair of headphones and a button box with two buttons

[25] A copy of the information sheet and consent form for Experiment 5 can be found in Appendix E

labelled *yes* and *no*. Partitions were placed such that the participants could not see the responses of others around them. They were instructed to put on their headphones and hold the button box such that each of their thumbs were assigned to one of the two buttons. The researcher explained that that they would hear a series of audio sequences. These sequences would be made up of fragments of English real or made-up words, and they would be presented to the participants binaurally through the headphones at a fixed volume and speed. Regardless of this audio input, the participants were instructed to focus on the computer screen in front of them. It was explained that on the computer screen, a series of words and nonwords would be presented visually. Participants were asked to decide whether they considered the word on the screen to be a real word in English, or a made-up word. It was explained that the difference between real and nonwords would be obvious, for example they may see the word *scold* or **troyp*. No abbreviated words were included in the study. Participants registered their decision by using the button box connected to the computer. If, for example the word *scold* appeared on the screen, participants were expected to select the *yes* button. Conversely, if the nonword **troyp* was presented, participants were expected to select the *no* button. Left-handed participants held the box the opposite way around so that their dominant hand was still performing the same selection as the right-handed participants.

Prior to the experiment beginning, participants responded to a training block of eight trials. This gave participants the opportunity to familiarise themselves with the format of the study and check that the volume of the auditory stimuli were at a comfortable level. Following the training block, participants were also given the opportunity to ask questions. As was the case in prior experiments, all permissible questions up to this point were regarding the format of the study, and could not relate to the aims and objectives of what the study hoped to find.

Each experimental block was three minutes in length. There was no repetition of stimuli. Each block contained forty trials, such that each participant saw twenty real word and twenty nonword targets pseudorandomised across the conditions. Related real word primes occurred in 12.5% of trials.

As with previous studies in this thesis, once the experiment was completed participants were given the opportunity ask any final, more specific, questions about the nature of the study before their session came to an end.

6.4 Analysis

As with the previous studies in this thesis, *JMP (SAS)* was used to conduct the analysis of the results for Experiment 5. This analysis was concerned with determining whether there was a statistically significant difference in the average reaction times taken to identify a real word between the experimental primes and the corresponding control primes. Therefore, whether there is was an effect on participant responses based on lexicality and relatedness. Results were therefore coded according to whether they were a WORD or NONWORD, RELATED or UNRELATED.

In total, one thousand, nine hundred, and twenty responses were recorded from across the forty-eight participants. As with the previous experiments, these results were cleaned according to the criteria outlined for Experiment 2.

Due to the nature of this experiment, no participants or trials were excluded on the basis of accuracy. Specifically, because participants were provided with fragments, they were required to not only complete a prime based on its vowel duration, but also match this with a semantically related target. Due to the complexity of this process, excluding participants who were not at least 80% percent accurate, as with the identification tasks, did not prove viable.

The Latin square method was used for the analysis of Experiment 5. This design is useful in that it allows indirect comparisons to be made between two groups of competing experimental primes by incorporating the role of two corresponding groups of control primes. In this way, one can determine whether there is an interaction between lexicality and relatedness in terms of the average reaction times taken by participants to recognise a visual target as a real word in English. This method also overcomes a potential issue with Experiment 5 relating to the length of the stimuli. Specifically, primes which contain a long vowel are longer by nature than those which contain a short vowel, such that any effect present in the results could arguably be attributed to the prime length, independent of the vowel length. Comparing an experimental stimulus with a corresponding control stimulus, as opposed to directly comparing the effect of vowel duration in two paired experimental stimuli, acts as a control condition for this and mitigates the effect of prime length independent of vowel length.

Four groups of results were incorporated into the analysis, each comprising of the reaction times for one of the four sets of primes. The mean reaction times from participants for the WORD and NONWORD conditions were then compared to determine whether there was an interaction between lexicality and relatedness. The structure of this comparison is illustrated in Figure 67 (below).

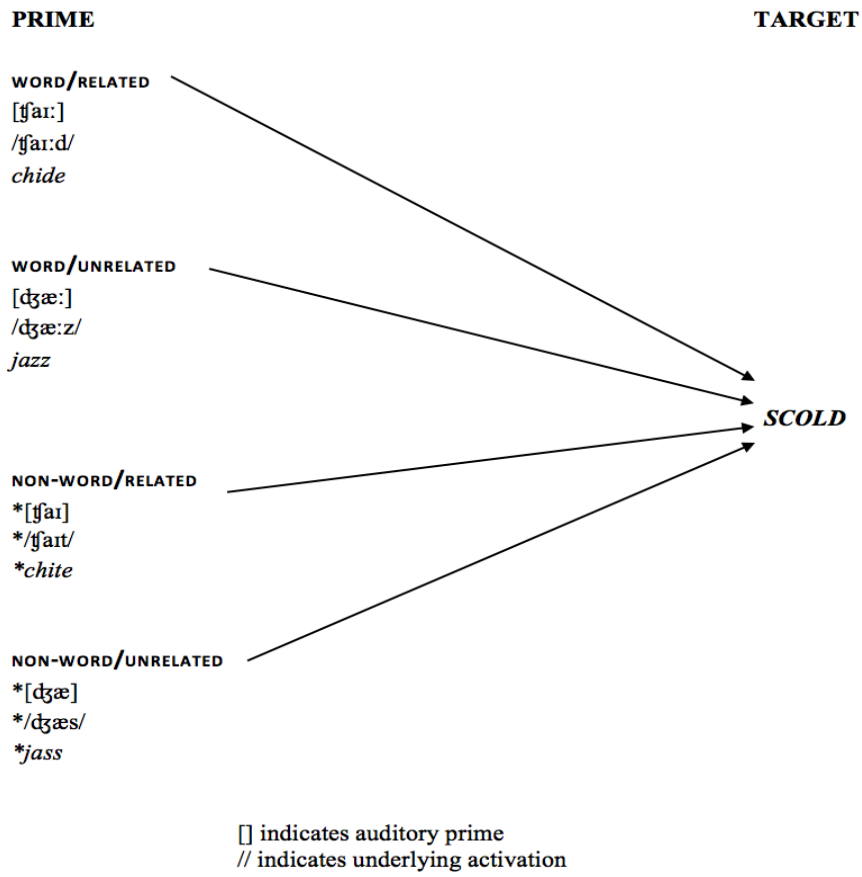


Figure 67: A demonstration of the relationship between the primes and targets for Experiment 5

Recall that the forty primes which related to nonword targets were excluded from the analysis. Recall also that for the WORD condition, it was expected that the RELATED experimental prime would demonstrate a significantly faster reaction time than the UNRELATED control prime. Conversely, for the NONWORD prime condition, it was expected that there would be no significant difference between the reaction times for the RELATED experimental prime and the UNRELATED control prime.

This prediction draws upon the notion that, for the WORD condition, there is a complementary relationship between the vowel duration and the word-final underlying [VOICE] feature required to activate a real word in English. Specifically, based on the example in Figure 67 (above), the underlying voiced stop /d/ was expected to be cue by the longer vowel duration

in the experimental prime, thereby activating the real word [ʧaɪːd] *chide*. Equally, when presented with the control prime, the longer vowel duration was expected to cue the underlying voiced fricative /z/, activating the word real [ʤæːz] *jazz*. As the target word *scold* was semantically related to the word [ʧaɪːd] *chide*, the reaction time in the case of the WORD/RELATED experimental prime was expected to be faster than the WORD/UNRELATED control prime, in which there was no semantic relationship with the target.

Conversely, it was expected that both the experimental and control primes for the NONWORD condition would inhibit the cue necessary for the correct word-final [VOICE] feature to be assigned. This was due to the mismatch between vowel duration and voicing, thereby not activating a real word in English. In Figure 67 (above), the real words [ʧaɪːd] *chide* and [ʤæːz] *jazz* were therefore expected to be inhibited. Instead, the shorter vowel duration was expected to cue the voiceless word-final stop /t/ and fricative /s/, triggering the nonwords *[ʧaɪːt] **chite* and *[ʤæːs] **jass*. Therefore, neither the experimental nor the control prime would motivate a semantic relationship with the target *scold*. As such, there was not expected to be a significant difference between the reaction times taken to recognise the visual target as a word in the NONWORD/RELATED and NONWORD/UNRELATED groups.

A mixed model analysis was conducted in *JMP*. The fixed effects analysed were *relatedness* and *lexicality*; i.e. whether the prime was related or unrelated to the target, and whether the prime consisted of a real word or nonword. The report incorporated both *visual target* and *participant* as random effects. Figure 68 (below) illustrates the reaction time data obtained from Experiment 5.

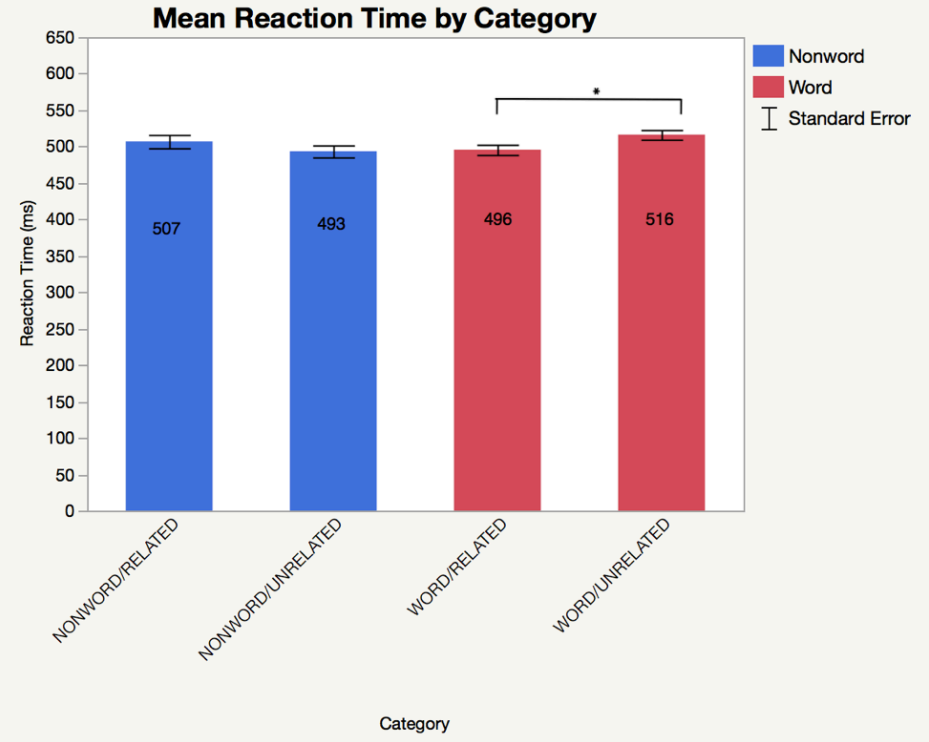


Figure 68: Mean reaction times according to lexicality for Experiment 5

Figure 68 (above) demonstrates that the predictions outlined above were indeed supported. There was an interaction of lexicality and relatedness (F = .0220). This interaction was driven by a facilitatory effect of priming in the WORD category regarding RELATED words versus UNRELATED words (with mean reaction times of 496 ms vs. 516ms, respectively). In the NONWORD category, UNRELATED words were recognised faster than their RELATED counterparts (with mean reaction times of 493 ms vs. 507 ms, respectively). Critically, when conducting a more targeted contrast between the corresponding experimental and control primes, a difference was found in the behaviour of participants between the reaction times originating from primes with a complementary relationship between vowel duration and voicing, and those without.

Looking firstly at those with a complementary relationship, the average reaction times for the WORD/RELATED experimental primes and WORD/UNRELATED control primes were compared and a significant difference was found (P = .045), with the WORD/RELATED primes

demonstrating a faster reaction time than the WORD/UNRELATED primes. As expected, in the case of primes comprising of a conflicting relationship between vowel duration and voicing, no significant difference in reaction times was found between the NONWORD/RELATED experimental primes and NONWORD/UNRELATED control primes ($P = .2145$).

The crucial finding of Experiment 5 is therefore that there was an effect on participant reaction times which was dependent on the relationship between vowel duration and word-final voicing. This is all the more impressive as the stimuli consisted purely of the word-initial consonant and vowel fragments. Here, native English speakers have again demonstrated their ability to determine word-final voicing based on the acoustic information provided to them by preceding vowel duration. Where this relationship between vowel duration and voicing was complementary, a real word was activated. Participants were then able to successfully ascertain a semantic relationship with the target word. This recognition of the target as a real word was significantly quicker when there was a semantic relationship present than when there was not. Equally, when a mismatch occurred between the vowel duration and the word-final [VOICE] feature necessary to create a real word, participants either did not activate a lexical entry, or this activation was slower due to the extra processing time taken to reconcile the mismatch. As such, the time taken to recognise the target as a real word took longer, and this resulted in no significant difference between the primes which were semantically related to the target and the primes which were not semantically related to the target.

Although the primary measure of significance is the reaction time data, the frequency and positioning of errors across the data also offers some insight into the findings of Experiment 5. Here, an *error* refers to any response in which a participant has incorrectly selected a real word target as a nonword, or conversely, a nonword target as a real word. This section of the analysis will therefore centre on determining the location of these errors across the data, focusing on

the frequency of errors within the four stimuli blocks, and determining how a comparison of these four sets of results can provide further insight into the findings of Experiment 5.

In total, twenty-nine errors were recorded across the data. The highest number of errors were found in the control NONWORD/UNRELATED group (41%). This was followed by the corresponding experimental block WORD/UNRELATED (31%). Conversely, both the experimental and control groups for the RELATED primes, including the test items, contained the same frequency of errors (14%). Figure 69 (below) illustrates this spread of results.

These findings will be further discussed in the subsequent section.

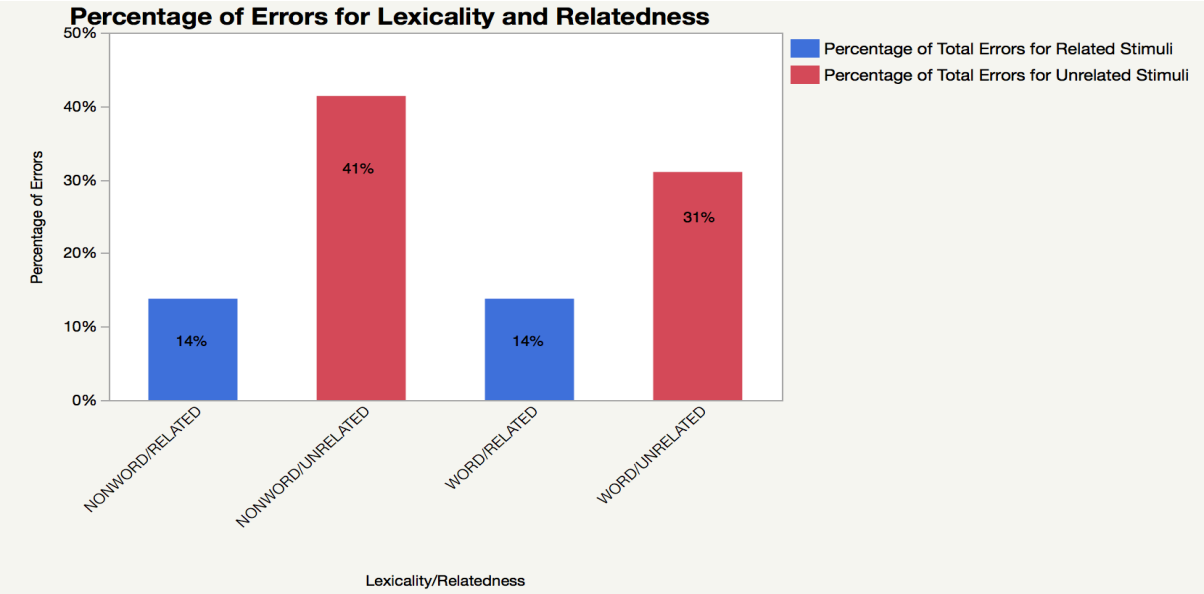


Figure 69: Number of errors found within the four experimental groups for Experiment 5

6.5 Discussion

Experiment 5 hypothesised that reaction times would be significantly faster for the experimental primes than for the corresponding control primes when there was a complementary relationship between vowel duration and word-final voicing. But, when there was a mismatch between vowel duration and word-final voicing, the experimental primes were not expected to be significantly faster than the control primes.

The findings support these hypotheses. The results showed that there was a significant difference in the mean reaction time taken to identify the target as a real word between the WORD/RELATED experimental primes and the WORD/UNRELATED control primes. However, no significant difference in the mean reaction time was found between the NONWORD/RELATED experimental primes and the NONWORD/UNRELATED control primes. This illustrates that when there is a complementary relationship between the vowel duration and the word-final voicing necessary to activate a real word in English, reaction times are faster when there is a semantic relationship established between the prime and target. This supports the aforementioned research conducted by Meyer and Schvaneveldt (1971) and Neely (1976) which found that real word targets are recognised faster when they have a semantic relationship with the prime. However, in the case of a mismatch between the vowel duration and word-final voicing necessary to activate a real word in English, there was no effect of a semantic relationship on either the RELATED or UNRELATED categories.

These results therefore support the notion that there is indeed an interaction between vowel duration, word-final voicing, and the lexicon. For native speakers of English, these findings suggest that the lexicon contains a longer vowel with the ability to activate an underlyingly voiced word-final consonant. Similarly, it contains a shorter vowel with the ability to activate an underlyingly voiceless word-final consonant. When presented with a prime which has a complementary relationship between the vowel duration and the voicing characteristic required to cue a real word in English, this activation occurs. However, when there is a mismatch between vowel duration and word-final voicing, the activation of a real word in English is inhibited. This once again illustrates that native speakers of English are sensitive to the cue that vowel duration provides in relation to the nature of word-final voicing. The findings complement Whalen (1991:355) who, having conducted a lexical decision task focused on

subcategorical mismatches, found that *'phonetic analysis resolves the conflict within the stimulus before searching the lexicon'*. This conclusion was made based upon reaction time latencies where a mismatch occurred. Whalen's (1991) study employed monosyllabic fricatives differentiated on the basis of place of articulation, lexicality, location (initial fricative or final fricative), and changeability (the change in lexical status when the fricative identity is altered). Examples of stimuli included, [mɛs]~[mɛʃ] *mess~mesh*, [gɪʃ]~[gɪs] **gish~*giss*, [səʊk]~[ʃəʊk] *soak*~*shoak, and [sɪlk]~[ʃɪlk]**shilk~silk*. Each test item contained a mismatch between the fricative identity, so [mɛs] *mess* would appear in one case, and then [mɛʃ] *mesh* in another case. A lexical effect was demonstrated by the greater error rate for nonwords compared with words. However, nonwords which were stable (i.e. did not change lexical status) showed lexical decisions that were similar in speed to the real words. As such, only changeable nonwords were slower. The results from Experiment 5 support Whalen (1991) in that a phonetic mismatch must be resolved before lexical access can occur, as demonstrated by the insignificant difference shown between the two NONWORD groups. These findings indicate that where there is a mismatch between the vowel duration and the voicing required to activate a real word in English, the real word either remains inactivated, or is slower to be activated due to the time taken for the participant to resolve the mismatch. For example, from *[ʧaɪːt] **chite* to [ʧaɪːd] *chide*. The analysis of errors supports the idea that the lexicon is subsequently searched following this phonetic resolution, as the highest number of errors occurred within the two UNRELATED groups. Therefore, a possible reason for these reaction time latencies, in addition to resolving a phonetic mismatch, could be because participants were confused by the absence of a semantic relationship between the auditory prime and visual target in the UNRELATED groups, leading them to select an incorrect response. However, this insinuates that a real word was eventually activated as participants were able to more accurately establish a

relationship between the prime and target where there was a semantic relationship present, even in the NONWORD cases. Of course, the spread of errors across the data is small, and as such vowel duration appears to be the primary cue influencing the reaction times from participants.

Due to the nature of stimuli formation, it must once again be acknowledged that it is not impossible that cues for word-final voicing were retained in the word-initial consonant (Coleman, 2003) or in spectral cues from the vowel, however, the potential presence of these cues does not negate the clear effect of vowel duration as determined by Experiment 5. Instead, these additional cues may be attributed to the complexity of speech perception, and the idea that there is no single perceptual cue which works in isolation when determining the nature of word-final voicing in English.

6.6 Conclusion

Experiment 5 has further demonstrated that there is an interaction between acoustic and lexical information. Participants were able to conceptualise the nature of word-final voicing on the basis of vowel duration, and this resulted in either the activation or the inhibition of a real-word in English. Additionally, participants were able to identify the presence or absence of a semantic relationship between the prime and the target with which they were being presented. Evidence from the analysis of errors suggests that the lexicon influenced perception in that fewer errors were made when there was a semantic relationship between the prime and target, regardless of whether the prime was designed to cue a real word or nonword. As previously discussed, this suggests that a lexical item is eventually activated even in cases where there is a mismatch between vowel duration and voicing. An additional strength of Experiment 5 is therefore the indirect nature of a lexical activation by partially matching prime. This demonstrates that phonetic information must be resolved prior to a lexical item being activated,

but equally lexical information demonstrates its importance in providing perceptual guidance for word-final voicing in English.

As no significant difference in mean reaction times were found between the control and experimental NONWORD groups, the potential, yet slower, activation of a real word in the lexicon was not able to override the effect of the mismatch between vowel duration and word-final voicing. These findings therefore complement the results from Experiments 2, 3, and 4 in supporting the notion that vowel duration is a primary and salient perceptual cue for word-final voicing in English, and lexical bias has a secondary effect.

CHAPTER SEVEN

Final discussion and conclusions

7.1 The relationship between vowel duration and word-final voicing

The [VOICE] feature is contrastive in English in all positions of a word. Therefore, this thesis was initially concerned with determining whether native speakers of English anticipate the nature of the [VOICE] feature for a word-final obstruent, and consequently lengthen the preceding vowel in production. Exp1a was a pilot study measuring native English speakers' production of vowel duration prior to the word-final stops [d] and [t]. The findings from Exp1a found that speakers consistently lengthened their vowel duration prior to the voiced word-final consonant, relative to the voiceless word-final consonant, with an average of 81ms difference. Furthermore, these results are consistent across underlying vowel contrasts, for example, monomoraic [æ] and bimoriac [eɪ] or [əʊ]. Relative lengthening occurs across all contexts. The results from Exp1a support the notion that the voicing characteristic of a word-final obstruent consistently correlates with preceding vowel duration, and this complements findings from previous research (House, 1961; Klatt, 1976; de Jong, 1991). Recall that several theories have been proposed as to why this tendency may occur in speech production. Raphael (1975) hypothesised that an increased duration of muscular activity is required to produce a voiced consonant, and this increase in energy leads to the production of a longer vowel. Similarly, Klatt (1976) suggests that speakers tend to open their glottis earlier prior to a voiceless consonant, preventing any excess voicing to be generated during articulation. This results in a shorter vowel.

Despite the myriad of phonetic reasons as to why this tendency may occur, from a phonological perspective the most notable finding from Exp1a is the evidence that it provides

for the prominence of the word-final [VOICE] feature, and the salience of underlying vowel duration. In support of the FUL model of speech perception, these findings suggest that speakers have a mental, abstract conception of the word-final [VOICE] feature. In accordance with the phonological rules of English, and the underlying representation of vowel duration, speakers are able to produce this feature word-finally as either voiced or voiceless. As such, the results from Exp1a consistently support the hypothesis that speakers are able to plan their utterance such that they can anticipate the nature of the word-final [VOICE] feature, and as such, vary the production of their preceding vowel duration accordingly.

Turning now to the extent to which native English listeners interpret vowel length as a perceptual cue for the word-final [VOICE] feature. Experiments 2-5 demonstrate that listeners have an ability to use vowel duration as a primary perceptual cue for word-final voicing across a variety of ambiguous contexts. Vowel length is not represented in terms of features but as the number of moras, and this distinction is underlyingly binary; either *short* or *long*. The question we investigated here is whether listeners interpret preceding vowel duration as either underlyingly *short* or *long* and use this phonetic information to anticipate the nature of the word-final [VOICE] feature. The findings of this thesis suggest that listeners are able to consistently make this distinction, to anticipate the nature of voicing word-finally.

In particular, Experiment 3 provides evidence that listeners are able to accurately predict word-final voicing, even when they are only provided with the CV:/CV fragment of CVC minimal pairs. Throughout this forced choice identification task, participants were consistently able to determine whether an original recording had ended with a word-final [d] or [t] and were accurate in 91% and 88% of cases, respectively. The stimuli used in Experiment 3 were identical to those used in Exp1a.

Of course, this thesis does not intend to claim that vowel duration is the only cue used by listeners to characterise word-final voicing in English. Multiple factors are likely to influence the perception of the word-final [VOICE] feature. This notion has been reflected throughout the literature. For example, as previously referenced, Hogan and Rozsypal (1980:1770) list several other acoustic factors, namely *'voice bar duration, silent closure duration, and burst/frication duration'.* Similar research has also focused on VOT (Pisoni and Lazarus, 1974) and transition length (Keating and Blumstein, 1978). Additionally, Coleman (2003) suggests that information regarding word-final voicing may be present in word-initial consonants.

Instead of denying the complex nature of speech perception and attempting to attribute a listener's behaviour to one-single factor, this thesis aims to decipher the extent to which vowel duration can be inferred as a primary perceptual cue for word-final voicing in English. In order to achieve this, the experimental stimuli were controlled to various degrees; utilising both recorded and synthesised speech, and incorporating both ambiguous and omitted word-final sounds. Indeed, the overarching findings from Experiments 2-5 consistently provide evidence that listeners are able to interpret vowel duration as a primary cue in categorising the word-final [VOICE] feature. This once again supports a model of perception in which listeners are able to extract the relevant acoustic information from the vowel duration necessary to guide their perception of the phonological [VOICE] feature word-finally.

However, factors aside from acoustic information also have the capacity to influence phoneme categorisation. Specifically, this thesis addresses the role of the lexicon and its interaction with vowel duration in determining word-final voicing. Eminently, the Ganong Effect (1980) suggests that speech perception occurs in the direction of a real word as opposed to a nonword. Experiments 2 and 4 incorporated stimuli consisting of recorded speech and synthesised speech, respectively. Both experiments contained an equal paring of words and

nonwords. The gated stimuli in Experiment 2 spanned a 65ms difference from the longest gate to the shortest gate, and incorporated the bimoriac vowel [eɪ]. A word-final ambiguous stop was presented at the end of the stimuli and this, along with the word-initial consonants, was identical across the stimuli. Though some weak lexical effects were noted, largely dependent on word frequency, a clear shift in the direction of a word versus a nonword was not evident. As such Experiment 2 concluded that participants were being primarily guided by vowel duration in making their choice between hearing a word-final [d] or [t]; perceiving steadily fewer voiced responses as the vowel gates got shorter. Comparatively, the stimuli in Experiment 4 consisted of gated synthesised speech. It incorporated two types of word-final ambiguous consonant; merges between the stop-pair [d]~[t] and the affricate-pair [ʤ]~[ʧ] such that the nature of voicing could not be identified in isolation. Here, the moraic vowel [æ] was incorporated as the vowel nucleus. For the stop-pair, there was an average durational difference of 69ms between the longest gate and the shortest gate. Similarly, for the affricate-pair there was a 79ms difference in vowel duration between the longest gate and the shortest gate. Despite some confusion originating from the synthesised nature of the stimuli, the findings from Experiment 4 once again support the notion that vowel duration is a primary perceptual cue for word-final voicing, as evidenced by a drop in the percentage of voiced responses perceived at each gate as the gates got progressively shorter. Some lexical effects were noted; for example, a majority of voiced responses were perceived across the vowel gates for the word pair of [dæd]~*[dæt] *dad*~**dat*. This demonstrates a potential lexical bias in the direction of the word [dæd] *dad*, as opposed to the nonword *[dæt] **dat*. However, the effect of lexical bias once again appears dependent on factors such as word frequency. This was evidenced by the finding that the same tendency exhibited by [dæd]~*[dæt] *dad*~**dat* was not apparent in the case of [væt]~*[væd] *vat*~**vad*. According to the CELEX database, the word frequency for [dæd] *dad*

is five hundred and sixty-seven, and just one hundred and ten for [væt]. This would suggest that lower frequency words are not activated as readily as high frequency words, and thereby behave more similarly to nonwords. Overall, the results of Experiment 4 once again support the notion that vowel duration acts as a primary perceptual cue for word-final voicing in native English listeners.

In light of the results from Experiment 2-4, Experiment 5 aimed to further explore the influence of the lexicon using a lexical decision task. It was designed to test whether vowel duration would have the capacity override the influence of lexical relationships between an auditory prime and visual target. A range of monomoraic and bimoriac vowels, [æ], [eɪ], [əʊ] and [aɪ], formed the vowel nuceli. If vowel duration acted as a primary perceptual influence, it was expected that the WORD/RELATED experimental primes would demonstrate a significantly faster reaction time than the WORD/UNRELATED control primes. Conversely, it was expected that there would be no significant difference between the reaction times for the NONWORD/RELATED experimental primes and the NONWORD/UNRELATED control primes. However, if lexical bias was exhibiting an overriding effect, the participants across these two groups would have behaved similarly, and there would be a significant difference found in both instances. Here, vowel duration would not have influenced the inhibition or longer processing times associated with a nonword. The former was found to be the case, and this once again supports the overarching finding that vowel duration is both a primary and necessary cue for word-final voicing in English.

Crucially, Experiments 2-5 demonstrate that native English listeners are perceptible to small changes in vowel duration, and are able to use this phonetic information and their underlying representation of vowel length to guide their phonological representation of the word-final [VOICE] feature. This is all the more impressive given the forced experimental setting in which

the participants were being asked to demonstrate this ability, and the varying stimuli formation methods and methodologies employed. Notably, these findings support research which attributes vowel duration as having the ability to provide both a primary and necessary perceptual cue for word-final voicing in English (Ainsworth, 1972; Denes, 1955; Klatt and Cooper, 1975; Raphael, 1972, among others).

Finally, this thesis aimed to determine the extent to which non-native speakers of English are able to anticipate and process this phonological tendency. Specifically, Experiments 1 and 4 incorporated native German speakers who were L2 English speakers. Recall that the word-final alternation between voiced and voiceless endings is contextually neutralised in German. As such, employing native German speakers in this research has enabled inferences to be made regarding the influence of native phonology on the non-native processing and interpretation of vowel duration as a cue for word-final voicing.

Exp1b directly compared CVC utterances, containing the monophthong [æ], from both native English speakers and native German speakers. The native German speakers were grouped in accordance with their exposure to spoken English, with speakers having been recorded either in the UK or Germany. As such, two sub-groups were formed. German speakers recorded in Germany, who had received limited exposure to spoken English, demonstrated no significant difference in vowel length between words ending in voiced and voiceless obstruents when compared with native English speakers and German speakers based in the UK, who had received an increased exposure spoken English. Interestingly, German speakers who had lived and worked in the UK demonstrated a more native-like production, though the durational differences in vowel length did remain smaller than the English speakers overall. This tendency was also less significant for affricates than for stops, in line with the underlying phonology of German (Table 34, below). The most notable aspect of these results is that they indicate that,

with increased exposure to a second language, speech production becomes more native-like; even regarding fine-grained phonetic differences such as variations in vowel duration. This suggests that speakers have the ability to develop and adapt their phonology to phonetic mapping in accordance with a second language.

Table 34: Mean vowel durational differences for stimuli ending in stops and affricates across the three language groups for Exp1b

Language group	Stops (ms)	Affricates (ms)
UK	130	127
German- Germany	30	15
German- UK	100	59

However, the extent to which this proves true appears to be constrained by native phonology. Here, a lexical effect is present. Recall that although the surface form of German does not permit voiced obstruents in a word-final position, its phonology does contain underlyingly voiced stops. Conversely, voiced affricates do not occur underlyingly in German, unless borrowed into the language. The results of Exp1b reflect the influence of L1 phonology on L2 production in that German speakers do not demonstrate significant differences in vowel duration in the case of nonwords ending in an affricate, regardless of their exposure to spoken English. In the case of words ending in an affricate, German speakers recorded in the UK continue to behave in a more English native-like manner. This suggests that non-native speakers are guided by primarily by L2 perceptual cues where no underlying voicing contrast occurs in their native phonology only when they have a preconceived idea of how a word should sound. This item is subsequently activated in their L2 lexicon, and the translation of this tendency does not appear to stretch to nonwords.

Turning now to speech perception, in addition to native English speakers, Experiment 4 also incorporated German speakers based in the UK, and German speakers based in Germany. The

stimuli were identical to those employed in Exp1b. The findings from non-native English listeners for stimuli ending in word-final stops indicate that German listeners perceived a majority of voiceless responses regardless of the amount of exposure to spoken English they had received. However, for affricates, a gradual decrease was evident in the number of voiced responses recorded as the gates got shorter. Here, the German listeners look to have been guided by their L1 native phonology in cases where they have an underlying voiced category prevalent in their native language, as is the case for stops. This is not the case for affricates and as such, listeners appear to be guided primarily by L2 phonetic cues.

Non-native English listeners do not demonstrate strong evidence to suggest that lexical bias can override the phonetic information being provided by the vowel duration. However, lexical status has influenced the perception of word-final affricates. In the case of the word pair [bædʒ]~[bætʃ] *badge~batch*, German listeners based in the UK perceived a majority of voiceless word-final sounds. Conversely, in the nonword pair *[rædʒ]~*[rætʃ] **radge~*ratch,* the same speakers were primarily guided by the vowel duration, thereby showing sensitivity to this L2 cue. For German speakers based in Germany, a significant difference in perception was evident in the case of the real word affricate-pair, but an insignificant difference was found in the nonword affricate-pair. Here, it appears that for non-native speakers with more exposure to spoken English, L1 phonology governs perception where an L2 lexical entry can be activated. Where a lexical entry cannot be activated, as in the case of a nonword, or where a speaker has received less exposure to spoken English, L2 phonology governs perception. Interestingly, this presents an asymmetry with the production data wherein native German speakers were primarily guided by their L1 phonology in the case nonwords ending with an affricate, regardless of the amount of exposure to spoken English that they had received.

Overall, German participants based in the UK demonstrated perceptual behaviour more similar to that of the native English speakers when compared with those participants based in Germany. These findings complement the production results from Exp1b, and this research suggests that acquiring a more native-like perception of vowel duration and its relationship with voicing in an L2 could be a learned phenomenon. However, the extent to which non-native speakers can interpret differences in vowel duration and thereby characterise the ambiguous word-final [VOICE] feature appears to be dependent on the presence of an existing category in the underlying phonology of their native language. In perception, where underlying native phonology permits a voicing distinction, non-native speakers appear to be primarily constrained by their L1 phonology. Conversely, where there is no such tendency, listeners are more likely to be influenced by English phonology, as driven by the acoustic information provided to them. The notion that underlying phonology influences perception supports the previous findings by researchers such as Jongman (1992) who demonstrated that speakers are guided not only by the surface form of an utterance, but also by their underlying representation of the [VOICE] feature.

7.2 Suggestions for further research

Extending the inclusion of both native and non-native participants to all experiments would provide the opportunity for more direct comparisons to be made between the perception of the word-final [VOICE] feature in L1 and L2 languages. In particular, rerunning Experiment 5 or developing a similar lexical decision task to include non-native English speakers may reveal more information regarding the nature of L2 perception.

Another key area to develop is the formation of stimuli. The three identification tasks reported here aimed to incorporate a range of methodologies, utilising both recorded and

synthesised speech along with presenting either an absent or ambiguous word-final consonant. Developing different methods of synthesising stimuli would be particularly useful to decipher whether a more natural sounding outcome can be achieved. This may have the potential to overcome the perceptual issues that participants experienced with the stimuli in Experiment 4. Similarly, some benefit may be gained from rerunning Experiment 4 and increasing the number of non-native participants. This would provide a greater body of data from which more widespread inferences could be made. In the same manner, recording more speakers in Exp1b would increase the reliability and richness of the data acquired.

Finally, incorporating other languages in which duration is contextually neutralised word-finally would provide further comparisons. For example, investigating Turkish or Russian would expand our understanding of the more nuanced ways in which different languages may behave when interacting with L2 perceptual cues.

7.3 Final remarks

As previously stated, this thesis does not deny the complex nature of speech perception and does not intend to suggest that vowel duration alone is solely the most influential cue for word-final voicing in English. Instead, the findings of this thesis have provided a more comprehensive analysis of the extent to which listeners use vowel duration to characterise word-final voicing across a range of ambiguous contexts. This thesis has explored the representation and processing of vowel length in relation to word-final voicing in both native and non-native English. We have focused on the salience of the underlying representation of vowel length in characterising the [VOICE] feature word-finally. Regardless of whether a vowel nuclei is monomoraic or bimoriac, the results prove consistent that the interaction between vowel duration and voicing plays a key role in English speech production and perception.

The incorporation of both the role of the lexicon, and the influence of non-native speech have provided additional lenses through which we may view this phonological tendency. The overarching findings of this research suggest that vowel duration is consistently used as a primary perceptual cue in determining word-final voicing, even when manipulated such that the influence of the lexicon has the potential to override this phonetic information. The role of the lexicon is not negligible, as lexical effects have been identified throughout the research, particularly in relation to real words versus nonwords. Lexical bias is also present in the data, though this is dependent on both word frequency. Non-native speakers of English appear to demonstrate an increased level of English native-like production and perception when they have had an increased exposure to English, suggesting that this could be a learned phenomenon. However, native phonology does constrain this non-native perception, with the primary evidence for this tendency coming from the different ways in which German speakers perceive word-final stops and affricates.

It is intended that the findings of this thesis have provided a thorough evaluation of the representation and processing of vowel length as a perceptual cue for word-final voicing in English. The findings of this research may therefore be situated within the wider research fields of phonetics, phonology, and psycholinguistics; contributing to these areas by producing significant and exciting insights into the field of speech perception.

REFERENCES

Ainsworth, W. A. 1972. 'Duration as a cue in the recognition of synthetic vowels' in *Journal of the Acoustical society of America*, Vol. 51, pp. 648-651.

Beguš, G. 2017. 'Effects of ejective stops on preceding vowel duration' in *Journal of the Acoustical Society of America*, Vol. 142, pp. 2168-2184.

Belasco, S. 1953. 'The influence of force of articulation of consonants on vowel duration' in *Journal of the Acoustical Society of America*, Vol. 25, pp. 1015–1016.

Best, C. T. & Tyler, M. D. 2006. 'Nonnative and second-language speech perception: Commonalities and complementarities' in Munro, M. J. & Bohn O- S. (eds.) *Second language speech learning: The role of language experience in speech perception and production.* Amsterdam: John Benjamins.

Broadbent, D. E. & Gregory, M. 1964. 'Accuracy of recognition for speech presented to the right and left ears' in *Quarterly Journal of Experimental Psychology*, Vol. 16, pp. 359-360.

Broersma, M. 2005. 'Perception of familiar contrasts in unfamiliar positions' in *Journal of the Acoustical Society of* America, Vol. 117, pp. 3890–3901.

Broersma, M. 2010. ‘Perception of final fricative voicing: Native and nonnative listeners’ use of vowel duration’ in *Journal of the Acoustical Society of America*, Vol. 127, pp. 1636-1644.

Browman C. P. & Goldstein, L. 1986. ‘Towards an articulatory phonology’ in *Phonology Yearbook*, Vol. 3, pp. 219–252.

Bryden, M. P. 1963. ‘Ear preference in auditory perception’ in *Journal of Experimental Psychology*, Vol. 65, pp. 103-105.

Burton, M. W., Baum, S. R. & Blumstein, S. E. 1989. ‘Lexical effects on the phonetic categorization of speech: The role of acoustic structure’ in *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 15, pp. 567-575.

Chaney, R. B. & Webster, J. C. 1965. *Information in certain multidimensional acoustic signals.* Report No. 1339. United States Navy Electronics Laboratory Reports: San Diego, California.

Chang, W. 2000. ‘Geminate vs. Non-Geminate consonants in Italian: Evidence from a Phonetic Analysis’ in *University of Pennsylvania Working Papers in Linguistics*, Vol. 7, pp. 53-63.

Charles-Luce, J. 1997. ‘Cognitive factors involved in preserving a phonemic contrast’ in *Language and Speech*, Vol. 40, pp. 229–248.

Chen, M. 1970. 'Vowel Length Variation as a Function of the Voicing of the Consonant Environment' in *Phonetica*, Vol. 22, pp. 129-159.

Chomsky, N. 1965. *Aspects of the Theory of Syntax.* Cambridge, Massachusetts: MIT Press.

Chomsky, N. & Halle, M. 1968. *'The Sound Pattern of English'*. New York: Harper & Row.

Coleman, J. 2003. 'Discovering the acoustic correlates of phonological contrasts' in *Journal of Phonetics*, Vol. 31, pp. 351-372.

Connine, C. M., Titone, D. & Wang, J. 1993. 'Auditory Word Recognition: Extrinsic and Intrinsic Effects of Word Frequency' in *Journal of Experimental Psychology*, Vol. 19, pp. 81-94.

Crowder, R. G. 1982. 'Decay of auditory memory in vowel discrimination' in *Journal of Experimental Psychology: Learning, Memory and Cognition,* Vol. 8, pp. 153-162.

Crowther, C. S. & Mann, V. 1992. 'Native language factors affecting use of vocalic cues to final consonant voicing in English' in *Journal of the Acoustical Society of America*, Vol. 92, pp. 711-722.

Crowther, C. S. & Mann, V. 1994. 'Use of vocalic cues to consonant voicing and native language background: The influence of experimental design' in *Perception & Psychophysics*, Vol. 55, pp. 513-525.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K. & Hogan, E. M. 2001. 'Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition' in *Language and Cognitive Processes*, Vol. 16, pp. 507–534.

De Jong, K. 1991. 'An Articulatory Study of Consonant-Induced Vowel Duration Changes in English' in *Phonetica*, Vol. 48, pp. 1-17.

Denes, P. 1955. 'Effect of Duration of the Perception of Voicing' in *Journal of the Acoustical society of America*, Vol. 27, pp. 761-764.

Diehl, R. L., Lotto, A. J. & Holt, L. L. 2004. 'Speech perception' in *The Annual Review of Psychology*, Vol. 55, pp. 149-179.

Dirksen, A. & J. S. Coleman. 1994. 'All-Prosodic Synthesis Architecture' in *Proceedings of the Second ESCA/IEEE Workshop on Speech Synthesis*, New Paltz, NY. Sept. 12-15, 1994, pp. 232-235.

Edge, B. A. 1991. 'The production of word-final voiced obstruents in English by L1 speakers of Japanese and Cantonese' in *Studies in Second Language Acquisition*, Vol. 13, pp. 377-393.

Eggermont, J. J. 2015. *Auditory Temporal Processing and its Disorders*. Oxford: OUP.

Eilers, R. E. 1977. 'Context-sensitive perception of naturally produced stop and fricative consonants by infants' in *Journal of the Acoustical Society of America*, Vol. 61, pp. 1321-1336.

Eimas, P.D. 1963. 'The relation between identification and discrimination along speech and non-speech continua' in *Language and Speech*, Vol. 6, pp. 206-217.

Eimas, P. D. 1974. 'Auditory and linguistic processing of cues for place of articulation by infants' in *Perception & Psychophysics*, Vol. 16, pp. 513-521.

Eimas, P. D., Siqueland, E. R., Jusczyk, P. W. & Vigorito, J. 1971. 'Speech perception in infants' in *Science,* Vol. 171, pp. 303-306.

Engstrand O. & Krull D. 1994. 'Durational correlates of quantity in Swedish, Finnish and Estonian: Cross language evidence for a theory of adaptive dispersion' in *Phonetica*, Vol. 51, pp. 80-91.

Esposito A. & Di Benedetto, M. G. 1999. 'Acoustical and perceptual study of gemination in Italian stops' in *Journal of the Acoustical Society of America*, Vol. 106, pp. 2051-2062.

Fant, G. 1960. *Acoustic Theory of Speech Production With Calculations based on X-Ray Studies of Russian Articulations.* The Hague: Mouton.

Fledge, J. E. 1984. ‘The effect of linguistic experience on Arabs' perception of the English /s/ vs. /z/ contrast’ in *Folia Linguistica*, Vol. 18, pp. 117-138.

Fledge, J. E. 1988. ‘The development of skill in producing word-final English stops: Kinematic parameters’ in *Journal of the Acoustical Society of America*, Vol. 84, pp. 1639-1652.

Fledge, J. E. 1999. *‘The relation between L2 production and perception’*, ICPhSS99, pp. 1273-1276. San Francisco.

Fledge, J. E. & Port, R. 1981. ‘Cross-language phonetic inference: Arabic to English’ in *Language and Speech*, Vol. 24, pp. 125-146.

Flege, J.E., Schmidt, A.M. & Wharton, G. 1995. ‘Age of learning affects rate-dependent processing of stops in a second language in *Phonetica*, Vol. 53, pp. 143-161.

Fodor, J. 1983. *The modularity of mind*. MIT Press, Cambridge: MA.

Forster, K. I. 1976. ‘Accessing the mental lexicon’ in Wales, R. J. & Walker, E. (eds.) *New Approaches to language mechanisms*, pp. 257-287. Amsterdam: North-Holland.

Foss, D. J. & Gernsbacher, M. A. 1983. ‘Cracking the dual code: Toward a unitary model of phoneme identification’ in *Journal of Verbal Learning and Verbal Behavior*, Vol. 22, pp. 609-632.

Fowler, C. A. 1981. 'Production and perception of coarticulation among stressed and unstressed vowels' in *Journal of Speech, Language & Hearing Research*, Vol. 46, pp. 127-139.

Fowler, C. A. 1984. 'Segmentation of coarticulated speech in perception' in *Perception & Psychophysics*, Vol. 36, pp. 359-368.

Fowler, C. A. 1994. 'Speech perception: direct realist theory' in Asher, R. E. (eds.) *The Encyclopaedia of Language and Linguistics*, pp. 4199-4203. Oxford: Pergamon.

Fowler, C. A. 1996. 'Listeners do hear sounds, not tongues' in *Journal of the Acoustical Society of America*, Vol. 99, pp. 1730-1741.

Fowler, C. A. 2006. 'Compensation for coarticulation reflects gesture perception, not spectral contrast' in *Perception & Psychophysics*, Vol. 68, pp. 161-177.

Fox, R. A. 1984. 'Effect of Lexical Status on Phonetic Categorization' in *Journal of Experimental Psychology: Human Perception and Performance,* Vol. 10, pp. 526-540.

Frauenfelder, U.H., Segui, J. & Dijkstra, T. 1990. 'Lexical effects in phonemic processing: Facilitatory or inhibitory?' in *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 16, pp. 77–91.

Friedrich, C. K., Lahiri, A. & Eulitz, C. 2008. 'Neurophysiological Evidence for u Underspecified Lexical Representations: Asymmetries With Word Initial Variations' in *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 34, pp. 1545-1559.

Fujisaki, H. & Kawashima, T. 1969. 'On the modes and mechanisms of speech perception' in *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo*, Vol. 28, pp. 67-73.

Fujisaki, H. & Kawashima, T. 1970. 'Some experiments on speech perception and a model for the perceptual mechanism' in *Annual Report of the Engineering Research Institute*, *Faculty of Engineering, University of Tokyo*, Vol. 29, pp. 207-214.

Galantucci, B., Fowler, C. A. & Turvey, M. T. 2006. 'The motor theory of speech perception reviewed' in *Psychonomic Bulletin & Review*, Vol. 13, pp. 361-377.

Ganong, W. F. 1980. 'Phonetic categorization in auditory word perception' in *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 6, pp. 110-125.

Goldinger, S. D. 1997. 'Words and voices: Perception and production in an episodic lexicon' in Johnson, K & Mullennix, J. W. (eds.) *Talker Variability in Speech Processing*, pp. 33-66. San Diego: Academic Press.

Goto, H. 1971. ‘Auditory perception by normal Japanese adults of the sounds "l" and "r."’ in *Neuropsychologia*, *Vol. 9*, pp. 317–323.

Greenlee, M. E. 1978. *‘Learning the phonetic cues to the voiced/voiceless distinction: an exploration of parallel processing in phonological change’*, Ph.D. dissertation, University of California, Berkeley.

Grosjean, F. 1980. ‘Spoken word recognition processes and the gating paradigm’ in *Perception & Psychophysics*, Vol. 28, pp. 267-283.

Halle, M., Stevens, K. N. & Oppenheim, A. V. 1967. ‘On the mechanism of glottal vibration for vowels and consonants’ in *Massachusetts Institute of Technology, Research Laboratory of Electronics, Quarterly Progress Report*, Vol. 85, pp, 267–277.

Ham, W. 2002. *Phonetic and phonological aspects of geminate timing.* New York: Routledge.

Hansen, B. B. 2004. ‘Production of Persian geminate stops: Effects of varying speaking rate’ in Agwuele, A., Warren, W. & Park, S. H. (eds.) *Proceedings of the 2003 Texas Linguistics Society Conference: Coarticulation in Speech Production and Perception*, pp. 86–95. Somerville, MA: Cascadilla Proceedings Project.

Harnad, S. 2003. ‘Categorical Perception’ in *Encyclopedia of Cognitive Science.* Nature Publishing Group/Macmillan.

Haskins Laboratories Quarterly Progress Report. 1956. No. 21, Appendix 5, 1-2.

Herd, W., Jongman, A. & Sereno, J. 2010. 'An acoustic and perceptual analysis of /t/ and /d/ flaps in American English' in *Journal of Phonetics.* Article in Press.

Hogan, J. T. & Rozsypal, A. J. 1980. 'Evaluation of vowel duration as a cue for the voicing distinction in the following word final consonant' in *Journal of the Acoustical Society of America*, Vol. 67, pp. 1764-1771.

House, A. S. 1961. 'On Vowel Duration in English' in *Journal of the Acoustical Society of America*, Vol. 33, pp. 1174-1178.

Johnson, K. 2005. Speaker Normalization in speech perception' in Pisoni, D. B. & Remez, R. (eds.) *The Handbook of Speech Perception*. Oxford: Blackwell Publishers.

Johnson, K. 2010. *Speech Perception.* UC Berkeley Phonology Lab Report.

Jones, D. 1944. 'Chronemes and Tonemes' in *Acta Linguistica*, Vol. 4, pp. 1–10.

Jones, D. 1950. *The Phoneme: Its Nature and Use.* Cambridge: Heffer.

Jongman, A., Sereno, J. A., Raaijmakers, M. & Lahiri, A. 1992. 'The Phonological Representation of [VOICE] in Speech Perception' in *Language and Speech*, Vol. 35, pp. 137-152.

Jusczyk, P. W., Pisoni, D. B., Reed, M. A., Fernald, A. & Myers, M. 1983. 'Infants' discrimination of the duration of a rapid spectrum change in nonspeech signals' in *Science*, Vol. 222, pp. 175-177.

Keating, P. & Blumstein, S. 1978. 'Effects of transition length on the perception of stop consonants' in *Journal of the Acoustical Society of America*, Vol. 64, pp. 57-64.

Kennard, H. J. & Lahiri, A. 2019. 'Nonesuch phonemes in loanwords' in *Linguistics*. Berlin: De Gruyter Mouton.

Kimura, D. 1961. 'Cerebral dominance and perception of verbal stimuli' in *Canadian Journal of Psychology*, Vol. 15, pp. 166-171.

Kimura, D. 1964. 'Left-right differences in the perception of melodies' in *Quarterly Journal of Experimental Psychology*, Vol. 16, pp. 355-358.

Kimura, D.1967. Functional asymmetry of the brain in dichotic listening' in *Cortex*, Vol. 3, pp. 163-178.

Klatt, D. H. 1976. 'Linguistic uses of segmental duration in English: Acoustic and perceptual evidence' in *Journal of the Acoustical Society of America*, Vol. 59, pp. 1208-1221.

Klatt, D. 1979. ‘Speech perception: A model of acoustic- phonetic and lexical access’ in *Journal of Phonetics*, Vol. 62, pp. 1345-1366.

Klatt, D. 1980. ‘Speech perception: A model of acoustic- phonetic analysis and lexical access’ in Cole, R. (eds.) *Perception and production affluent speech*, pp. 243- 288. Hillsdale, NJ: Erlbaum.

Klatt, D. H. & Cooper, W. E. 1975. ‘Perception of segment duration in sentence contexts’ in *Journal of the Acoustical Society of America*, Vol. 57, pp. S47-S48.

Kuhl, P. K. & Miller, J. D. 1975a. ‘Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants’ in *Science*, Vol. 190, pp. 69-72.

Kuhl, P. K. & Miller, J. D. 1975b. ‘Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech sound categories’ in *Journal of the Acoustical Society of America*, Vol. 870, pp. 340-349.

Kuhl, P. K. & Miller, J. D. 1978. ‘Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli’ in *Journal of the Acoustical Society of America*, Vol. 63, pp. 905-917.

Kuhl, P. K. & Padden, D. M. 1983. ‘Enhanced discriminability at the phonetic boundaries for the place feature in macaques’ in *Journal of the Acoustical Society of America*, Vol. 73, pp. 1003-1010.

Lahiri, A. & Reetz, H. 2002. 'Underspecified recognition' in Gussenhoven, C., Werner, N. & Rietveld, T. (eds.) *Laboratory phonology 7*, pp. 637–676. Berlin: Mouton.

Lehtonen, J. 1970. *'Aspects of quantity in standard Finnish'*. Jyväskylä: Jyväskylä University Press.

Liberman, A. M. 1996. *Speech: A Special Code*. Cambridge: MIT Press.

Liberman, A. M., Harris, K. S., Hoffman, H. S. & Griffith, B. C. 1957. 'The discrimination of speech sounds within and across phoneme boundaries' in *Journal of Experimental Psychology*, Vol. 54, pp. 358-368.

Liberman, A. M., Harris, K., Eimas, P., Lisker, L. & Bastian, J. 1961a. 'An effect of learning on speech perception: the discrimination of durations of silence with and without phonemic significance' in *Language and Speech*, Vol. 4, pp. 175–95.

Liberman, A. M., Harris, K. S., Kinney, J. A. & Lane, H. 1961b. 'The discrimination of relative onset- time of the components of certain speech and nonspeech patterns' in *Journal of Experimental Psychology*, Vol. 61, pp. 379– 88.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. 1967. 'Perception of the speech code' in *Psychological review*, Vol. 74, pp. 431-461.

Liberman, A. M. & Mattingly, I. G. 1985. 'The motor theory of speech perception revised' in *Cognition*, Vol. 21, pp. 1-36.

Lotto, A. J. & Holt, L. L. 2006. 'Putting phonetic context effects into context: A commentary on Fowler (2006)' in *Perception & Psychophysics*, Vol. 68, pp. 178-183.

Luce, P. A. 1986. *Neighbourhoods of words in the mental lexicon*. Unpublished Doctoral dissertation. Bloomington: Indiana University.

Luce, P. A. & Pisoni, D. B. 1998. 'Recognizing spoken words: The neighborhood activation model' in *Ear and hearing*, Vol. 19, pp. 1-36.

Mack, M. 1982. 'Voicing-dependent vowel duration in English and French: monolingual and bilingual production' in *Journal of the Acoustical Society of America*, Vol. 71, pp. 173–178.

Malécot, A. 1970. 'The Lenis-Fortis Opposition: Its Physiological Parameters' in *Journal of the Acoustical Society of America*, Vol. 47, pp. 1588-1592.

Mann, V. A. 1980. 'Influence of preceding liquid on stop-consonant perception' in *Perception and Psychophysics*, Vol. 28, pp. 47-412.

Mann, V. A. & Liberman, A. M. 1983. 'Some differences between phonetic and auditory modes of perception' in *Cognition*, Vol. 14, pp. 211-235.

Marslen-Wilson, W. & Welsh, A. 1978. 'Processing interactions and lexical access during word recognition in continuous speech' in *Cognitive Psychology*, Vol. 10, pp. 29-63.

Marslen-Wilson, W. & Warren, P. 1994. 'Levels of perceptual representation and process in lexical access: Words, phonemes, and features' in *Psychological* Review, Vol. 101, pp. 653–675.

McClelland, J. L. & Elman, J. L. 1986. 'The TRACE model of speech perception' in *Cognitive psychology*, Vol. 18, pp. 1-86.

McClelland, J. L., Thomas, A., McCandliss, B. D. & Fiez, J. A. 1999. 'Understanding failures of learning: Hebbian learning, competition for representational space, and some preliminary experimental data' in Reggia, J., Ruppin, E. & Glanzman, D. (eds.) *Disorders of brain, behavior, and cognition: The neurocomputational perspective. Progress in Brain Research*, Vol. 121, pp. 75–80. Oxford, England: Elsevier.

McClelland, J. L., Mirman, D. & Holt. L. L. 2006. 'Are there interactive processes in speech perception?' in *TRENDS in Cognitive Sciences,* Vol. 10, pp. 363-369.

McQueen, J. M. 1991. 'The Influence of the Lexicon on Phonetic Categorization: Stimulus Quality in Word-Final Ambiguity' in *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 17, pp. 433-443.

McQueen, J. M., Norris, D. & Cutler, A. 1999. 'Lexical influence in phonetic decision making: Evidence from subcategorical mismatches' in *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 25, pp. 1363–1389.

Mermelstein, P. 1978. 'On the relationship between vowel and consonant identification when cued by the same acoustic information' in *Perception & Psychophysics,* Vol. 23, pp. 331-336.

Meyer, D. E. & Schvaneveldt, R. W. 1971. 'Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations' in *Journal of experimental psychology*, Vol. 90, pp. 227-234.

Miller, J. L. & Liberman, A. M. 1979. 'Some effects of later-occurring information on the perception of stop consonant and semi-vowel' in *Perception & Psycholinguistics*, Vol. 25, pp. 457-465.

Miller, J. L. & Eimas, P. D. 1983. 'Studies on the categorization of speech by infants' in *Cognition*, Vol. 13, pp. 135-165.

Miller, J. L. & Dexter, E. R. 1987. *Effects of speaking rate and lexical status on phonetic perception*. Manuscript submitted for publication.

Mirman, D., McClelland, J. L. & Holt, L. L. 2005. 'Computational and behavioral investigations of lexically induced delays in phoneme recognition' in *Journal of Memory and* Language, Vol. 52, pp. 424–443.

Morse, P.A. & Snowdon, C.T. 1975. 'An investigation of categorical speech discrimination by rhesus monkeys' in *Perception & Psychophysics*, Vol. 17, pp. 9–16.

Morton, J. 1969. 'Interaction of information in word perception' in *Psychological Review*, Vol. 76, pp. 165-178.

Morton, J. 1979. 'Word recognition' in Morton, J. & Marshall, J. C. (eds.) in *Psycholinguistics 2: Structure and processes*, pp. 107-156. London: Paul Elek.

Munro, M. J. 1990. 'Attention to spectral and temporal cues in vowel perception among native speakers of Arabic and English' in *Journal of the Acoustical Society of America*, Vol. 88, pp. S53.

Neely, J. H. 1976.' Semantic priming and retrieval from lexical memory: Evidence for facilitatory and inhibitory processes' in *Memory & Cognition*, Vol. 4, pp. 648-654.

Norlin, K. 1987. *'A phonetic study of emphasis and vowels in Egyptian Arabic'*. Department of Linguistics and Phonetics, Working Papers, 30. Lund: Bloms Lund.

Norris, D., McQueen, J. M. & Cutler, A. 2000. 'Merging information in speech recognition: Feedback is never necessary' in *Behavioral and Brain Sciences*, Vol. 23, pp. 299-325.

Ohala, M. 1983. *'Aspects of Hindi Phonology'*. New Delhi, India: Motilal Banarsidass Publishers Pvt. Ltd.

Öhman, S. 1967. 'Peripheral motor commands in labial articulation' in *Speech Transmission Laboratory Quarterly Progress and Status Reports,* Vol. 8, pp. 30–63.

Pallier, C., Colomé, A. & Sebastián-Gallés, N. 2001. 'The influence of native-language phonology on lexical access: Exemplar-Based Versus Abstract Lexical Entries' in *Psychological Science*, Vol. 12, pp. 445-449.

Pickett, E. R., Blumstein, S. E. & Burton, M. W. 1999. 'Effects of Speaking Rate on the Singleton/ Geminate Consonant Contrast in Italian' in *Phonetica*, Vol. 56, pp. 135–157.

Pisoni, D. B. 1973. 'Auditory and phonetic codes in the discrimination of consonants and vowels' in *Perception and Psychophysics*, Vol. 13, pp. 253-260.

Pisoni, D. B. & Lazarus, J. H. 1974. 'Categorical and noncategorical modes of speech perception along the voicing continuum' in *Journal of the Acoustical Society of America*, Vol. 55, pp. 328-333.

Pisoni, D. B., Carrell, T. D. & Gans, S, J. 1983. 'Perception of the duration of rapid spectrum changes in speech and non-speech signals' in *Perception and Psycholinguistics,* Vol. 34, pp. 314-322.

Pitt, M. A., Kim, W., Navarro, D. J. & Myung, J. I. 2006. 'Global model analysis by parameter space partitioning' in *Psychological Review*, Vol. 113, pp. 57–83.

Polivanov, E. 1932. 'La perception des sons d'une langue étrangère' in *Travaux du Cercle Linguistique de Prague*, Vol. 4, pp. 79-96.

Raphael, L. J. 1972. 'Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English' in *Journal of the Acoustical Society of America*, Vol. 51, pp. 1296-1303.

Raphael, L. J. 1975. 'The physiological control of durational differences between vowels preceding voiced and voiceless consonants in English' in *Journal of Phonetics*, Vol. 3, pp. 25-33.

Raphael, L., Dorman, M., Freeman, F. & Tobin, C. 1975. 'Vowel and nasal duration as cues to voicing in word-final stop consonants: Spectrographic and perceptual studies' in *Journal of Speech Language & Hearing Research*, Vol. 18, pp. 389–400.

Remez, R. E., Rubin, P. E., Pisoni, D. B. & Carrell, T. D. 1981. 'Speech perception without traditional speech cues' in *Science*, Vol. 212, pp. 947-950.

Repp, B. H. 1982. ‘Phonetic Trading Relations and Context Effects: New Experimental Evidence for a Speech Mode of Perception’ in *Psychological Bulletin*, Vol. 92, pp. 81-110.

Roberts, A. C., Wetterlin, A. & Lahiri, A. 2013. ‘Aligning mispronounced words to meaning: Evidence from ERP and reaction time studies’ in *The Mental Lexicon*, Vol.8, pp.140-163.

Rubin, P., Turvey, M. T. & Van Gelder P. 1976. ‘Initial phonemes are detected faster in spoken words than in spoken nonwords’ in *Perception and Psychophysics*, Vol. 19, pp. 394-398.

Samuel, A. G. 2010. ‘Speech Perception’ in *The Annual Review of Psychology 2011*, Vol. 62, pp. 16.1-16.24.

Shrotriya, N., Siva Sarma A. S., Verma, R. & Agrawal S. S. 1995. ‘Acoustic and perceptual characteristics of geminate Hindi stop consonants’ in *Proceedings of ICPhS*, pp. 132-135.

Summers, W. V. 1987. ‘Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses’ in *Journal of the Acoustical Society of America*, Vol. 82, pp. 847-863.

Summers, W. V. 1988. ‘F1 structure provides information for final-consonant voicing’ in *Journal of the Acoustical Society of America*, Vol. 84, pp. 485-492.

Trubetzkoy, N. 1969. *Principles of phonology* (C.A. Baltaxe, Trans.). Berkeley: University of California Press. (Original work published 1939).

Van Santen, J. P. H. 1992. ‘Contextual effects on vowel duration’ in *Speech Communication*, Vol. 11, pp. 513-546.

Van Santen, J. P. H. & Olive J. P. 1990. ‘The analysis of contextual effects on segmental duration’ in *Computer Speech and Language*, Vol. 4, pp. 359-390.

Walsh, T. & Parker, F. 1983. ‘Vowel length and vowel transition cues to [+/-voice] in post-vocalic stops’ in *Journal of Phonetics*, Vol. 11, pp. 407-412.

Wardrip-Fruin, C. 1982. ‘On the status of temporal cues to phonetic categories: Preceding vowel duration as a cue to voicing in final stop consonants’ in *Journal of the Acoustical Society of America*, Vol. 71, pp. 187-195.

Warren, R. M. 1970. ‘Perceptual restoration of missing speech sounds’ in *Science*, Vol. 167, pp. 392-393.

Warren, R. M. 1976. ‘Auditory illusions and perceptual processes’ in Lass, N. J. (eds.) *Contemporary issues in experimental phonetics*, pp. 389-418. New York: Academic Press.

Werker, J. F. & Logan, J. S. 1985. ‘Cross-language evidence for three factors in speech perception’ in *Perception & Psychophysics*, Vol. 37, pp. 35-44.

Whalen, D. H. 1991. ‘Subcategorical phonetic mismatches and lexical access’ in *Perception and Psycholinguistics*, Vol. 50, pp. 315-360.

Whalen, D. H. & Liberman, A. M. 1987. ‘Speech perception takes precedence over nonspeech perception’ in *Science*, Vol. 237, pp. 169-171.

Wolf, C. G. 1978. ‘Voicing cues in English final stops’ in *Journal of Phonetics,* Vol.6, pp. 299-309.

Wurm, L. H. & Samuel, A. G. 1997. ‘Lexical inhibition and attentional allocation during speech perception: Evidence from phoneme monitoring’ in *Journal of Memory and Language*, Vol. 36, pp. 165–187.

Yantis, S. & Pashler, H. 2002. *Stevens' Handbook of Experimental Psychology*, *Sensation and Perception.* New York: John Wiley and Sons.

APPENDIX A

CUREC approval document

SOCIAL SCIENCES & HUMANITIES
INTER-DIVISIONAL RESEARCH ETHICS COMMITTEE

Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD
Tel: +44(0)1865 616576 Fax: +44(0)1865 280467
ethics@socsci.ox.ac.uk

22 February 2017

Zoe Kay Fletcher
Faculty of Linguistics/Philology/Phonetics (LPP)

Dear Zoe,

Research Ethics Approval (CUREC 1A)
Ref No: R50069/RE001

Title: Vowels and Voicing: A perceptual study investigating the links between vowel length and word-final obstruent voicing in English

The above application has been considered on behalf of the Social Sciences and Humanities Inter-divisional Research Ethics Committee (IDREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to the IDREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly approval has been granted.

Should there be any subsequent changes to the project, which raise ethical issues not covered in the original application, you should submit details to the IDREC for consideration.

Yours sincerely,

Claudia Kozeny-Pelling
Research Ethics Manager and Secretary SSH IDREC

cc: Professor Aditi Lahiri
Dan Holloway

APPENDIX B

Experiment 2: information sheet and consent form

Faculty of Linguistics, Philology & Phonetics

UNIVERSITY OF OXFORD

Identification Task for Research into Language Perception
Information Sheet

Language & Brain Laboratory, University of Oxford
M.Phil student Zoe Fletcher (Supervisor: Prof. Aditi Lahiri)
Contact: zoe.fletcher@ling-phil.ox.ac.uk Phone: 07773507473

This study looks into language perception and can help us understand how one of the most fundamental characteristics of the human species – language- is perceived by the brain.

You are invited to take part in this experiment, as you are a native speaker of British English. The following instructions will give you a description of what this experiment involves and what you will be required to do. You will be able to ask questions before the experiment and before you decide whether you want to participate or not. Should you decide to take part, you will be asked to sign a consent form, but please bear in mind that you are free to withdraw from the experiment at any point without penalty and without specifying a reason for your decision.

At the end of the experiment, there will be a short debriefing session where you will be able to ask specific questions about the design of the experiment. I hope that you will find the opportunity to learn more about linguistics and language perception enriching and maybe even a beneficial experience for your own studies. There are no risks involved in the study. If you, nonetheless, would like to raise any concerns you may have, or would like to make a formal complaint at any point, please speak to the relevant researcher (07773507473) [or their supervisor ([01865865243]), where applicable,] who will do her best to answer your query. The researcher should acknowledge your concern within 10 working days and give you an indication of how she intends to deal with it. If you remain unhappy or wish to make a formal complaint, please contact the chair of the Research Ethics Committee at the University of Oxford (Chair, Social Sciences & Humanities Inter-Divisional Research Ethics Committee; Email: ethics@socsci.ox.ac.uk; Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD). The chair will seek to resolve the matter in a reasonably expeditious manner.

This study has been reviewed by the University of Oxford Central University Research Ethics Committee and has received ethical approval. The data obtained in this study will be anonymised and used for research purposes only. Anonymity means that you will receive a unique participant identification number and that this participant identification number will not be associated with your first or last name. The data obtained will be available to the student researcher, her supervisor and the examiners of this M.Phil thesis. The data will be stored on the local hard drive of the PC in the perception lab at the Oxford Language & Brain laboratory, as well as on a USB stick to which only the student researcher will have access. At the end of the

project, the results of this MPhil thesis may be published in the Oxford University Research Archive. This is an online archive of research materials that the University of Oxford has established to make its research available for the benefit of society and the economy. It includes student theses that were successfully submitted as part of a postgraduate programme. Should you agree to participate in this study, the data will be used to write up a thesis, which, upon its successful completion and submission, will be available in print and online in the University archives.

Instructions

Please read the following instructions carefully before the task.

In this experimental study, you will be asked to put on a pair of headphones, look at a computer screen and press a button.

Procedure:

You will have a chance to ask questions before start of the experiment. There will also be a short practice session.

The experiment is expected to take 30minutes from start to finish, including setting up the task and a short debriefing period after you have completed the task in which you will have the opportunity to ask any other questions you may have about the experiment.

After putting on a pair of headphones and sitting in front of the screen, I will ask if you are ready. If you are, I will press a button on my computer to start the experiment. The screen will be blank for a few moments, and then you will hear a word/non-word and see on the screen a choice of two final sounds. You must decide which sound you think you have heard and press the corresponding button. After you press a button, the screen will go blank until the next item appears. You will only be played one item at a time. Try to be as accurate as possible, but don't ponder a word/non-word for too long or worry about having made a mistake. I hope you enjoy the task!

Thank you very much for your participation.

Faculty of Linguistics, Philology & Phonetics

UNIVERSITY OF OXFORD

Identification Task for Research into Language Perception

Consent Form

Language & Brain Laboratory, University of Oxford
M.Phil student Zoe Fletcher (Supervisor: Prof. Aditi Lahiri)
Contact: zoe.fletcher@ling-phil.ox.ac.uk Phone: 07773507473

Consent

I, _____, hereby declare that

- I have read the information sheet on this experiment and therefore agree to participate in this study.
- I was given the opportunity to ask questions about the study and all my questions and queries were answered to my full satisfaction.
- I know that I can withdraw from this study any time without penalty and without giving a reason for my decision.
- I understand that this project received ethics clearance by the University of Oxford Central University Research Ethics Committee.
- I am aware that the anonymised results of the task will be used for research purposes and will thus be available to the student researcher, her supervisor and the examiners. The results of this study may be stored and published in the Oxford University Research Archive and I do not object to my data being used.
- I have also been given information on how to raise a concern and make a complaint.

Date: _____ Signature: _____

I, the researcher Zoe Fletcher, hereby declare that I confirm and witness the giving of consent from the participant.

Date: _____ Signature: _____

Task Number: _____ File Name: _____

Participant Number: _____

APPENDIX C

Experiment 3: information sheet and consent form

FACULTY OF LINGUISTICS, PHILOLOGY AND PHONETICS

Miss Zoe Fletcher
Centre for Linguistics and Philology
Walton Street, Oxford OX1 2HG
Tel: 07714981147
zoe.fletcher@ling-phil.ox.ac.uk

Identification Task for Research into Language Perception
Information Sheet

Language and Brain Laboratory
DPhil student Zoe Fletcher (Supervisor: Prof. Aditi Lahiri)
Contact: zoe.fletcher@ling-phil.ox.ac.uk

This study looks into speech perception and can help us understand how one of the most fundamental characteristics of the human species – language- is perceived by the brain.

You are invited to take part in this experiment as you are a native speaker of British English who is over the age of 18, and have no known hearing, language, or learning difficulties. The following instructions will give you a description of what this experiment involves and what you will be required to do. We collect data by recording your responses as well as response times. The procedure is completely non-invasive and involves no risk. There is no direct benefit to you for taking part, but the results you provide for this study will contribute to our understanding of speech processing.You will be able to ask questions before the experiment and before you decide whether you want to participate or not. Should you decide to take part, you will be asked to sign a consent form, but please bear in mind that you are free to withdraw from the experiment at any point without penalty, and without specifying a reason for your decision.

If you are happy to take part in the study you will be asked to put on a pair of headphones and sit in front of a computer screen, attached to which will be a handset with a choice of two buttons. Once you are ready, I will then press a button on my computer to start the experiment. The screen in front of you will be blank for a few moments, you will hear a series of beeps, and then you will hear part of a word and see on the screen a choice of two word final sounds. You must decide which of the two sounds you think the word should end with and press the corresponding button. You will hear a beep between items. You will only be played one item at a time, and will hear each item only once. Try to be as accurate as possible, but don't ponder an item for too long or worry about having made a mistake, once you have pressed a button you cannot change your decision. You will be compensated for your time at the rate of £5 per 30 minutes. The entire experiment is expected to last for 30 minutes, including set up, and the opportunity to ask questions both before and after the task has been completed. The entire experiment will take place in the Language and Brain Laboratory (37, Wellington Square, OX1 2JF), and there will not be any follow-up visits.

I hope that you will find the opportunity to learn more about linguistics and language perception enriching. If you would like to raise any concerns that you may have, or would like to make a complaint at any point, please speak to myself (07714981147) or my supervisor Professor Lahiri (+44-1865-280401) and we will do our best to answer your query. We should acknowledge your concern within 10 working days and give you an indication of how we intend to deal with it. If you remain unhappy, you will be able to contact the chair of the Research Ethics Committee at the University of Oxford (Chair, Social Sciences & Humanities Inter-Divisional Research Ethics Committee; Email: ethics@socsci.ox.ac.uk; Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD). The chair will seek to resolve the matter in a reasonably expeditious manner.

All research data and records will be stored for a minimum retention period of 3 years after the publication or public release of the research. This study has been reviewed by, and received, ethics clearance from the University of Oxford Central University Research Ethics Committee. The University of Oxford is committed to the dissemination of this research for the benefit of society and the economy and, in support of this commitment, has established an online archive of research materials. This archive includes digital copies of student theses successfully submitted as part of a University of Oxford postgraduate degree programme. Holding the archive online gives easy access for researchers to the full text of freely available theses, thereby increasing the likely impact and use of that research. If you agree to participate in this study, the research will be written up as a thesis. On successful submission of the thesis, it will be deposited both in print and online to the University archives, to facilitate its use in future research. This thesis will be published open access. These data obtained in this study will be anonymised and used for research purposes only. Anonymity means that you will receive a unique participant identification number and that this participant identification number will not be associated with your first or last name. These data obtained will be available to myself, my supervisor, and the examiners of the DPhil thesis. These data will be stored on the secure hard drive of the PC in the Oxford Language & Brain Laboratory, the secure University of Oxford server, as well as on my own personal password protected computer.

If you have any questions, please ask! I hope that you enjoy the task!

FACULTY OF LINGUISTICS, PHILOLOGY AND PHONETICS

Miss Zoe Fletcher
Centre for Linguistics and Philology
Walton Street, Oxford OX1 2HG
Tel: 07714981147
zoe.fletcher@ling-phil.ox.ac.uk

Identification Task for Research into Language Perception

Consent Form

Language and Brain Laboratory
DPhil student Zoe Fletcher (Supervisor: Prof. Aditi Lahiri)
Contact: zoe.fletcher@ling-phil.ox.ac.uk

I, _____, hereby declare that (please initial each statement on the line provided)

- I have read the information sheet on this experiment and therefore agree to participate in this study. ____
- I understand that this consent is being given subject to normal legal requirements. ____
- I was given the opportunity to ask questions about the study and all of my questions and queries were answered to my full satisfaction. ____
- I know that I can withdraw from this study any time without penalty and without giving a reason for my decision. ____
- I understand that this project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee. ____
- I am aware that the anonymised results of the task will be used for research purposes and will thus be available to the student researcher, her supervisor and the examiners according to these data Protection Act. The results of this study may be stored and published in the Oxford University Research Archive. ____
- I have also been given information on how to raise a concern and make a complaint.____

- I agree to take part in the above study. ____

Optional (please initial each statement on the line provided, should you wish to do so):

- I do not object to my data being used in future publications in scientific journals, presentations or conferences. ____
- I agree for research data collected in this study to be given to researchers, including those working outside of the EU, to be used in other research studies. I understand that any data that leaves the research group will be fully anonymised so that I cannot be identified. ____
- I agree for my personal data to be kept in a secure database for the purpose of contacting me about future studies. ____

Date: _____ Signature: _____

I, the researcher Zoe Fletcher, hereby declare that I confirm and witness the giving of consent from the participant.

Date: _____ Signature: _____

APPENDIX D

Experiment 4: information sheet and consent form

FACULTY OF LINGUISTICS, PHILOLOGY AND PHONETICS

Ms Zoe Fletcher
Centre for Linguistics and Philology Walton Street,
Oxford,
OX1 2HG
Tel: 07714981147
zoe.fletcher@ling-phil.ox.ac.uk

Identification Task for Research into Language Perception

Information Sheet

DPhil student Zoe Fletcher (Supervisor: Prof. Aditi Lahiri)
Contact: zoe.fletcher@ling-phil.ox.ac.uk

This study looks into speech perception and can help us understand how one of the most fundamental characteristics of the human species – language- is perceived by the brain.

You are invited to take part in this experiment as you are a native speaker of Southern British English or native German who is over the age of 18, and have no known hearing, language, or learning difficulties. The following instructions will give you a description of what this experiment involves and what you will be required to do. You will be able to ask questions before the experiment. Should you decide to take part, you will be asked to sign a consent form, but please bear in mind that you are free to withdraw from the experiment at any point should you wish to do so.

If you are happy to take part in the study you will be asked to put on a pair of headphones and sit in front of a computer screen, attached to which will be a handset with a choice of two buttons. Once you are ready, the screen in front of you will count down and you will hear a series of beeps, and then you will hear a word (or nonsense word) in English and see on the screen a choice of two word final sounds. You must decide which of the two sounds you think that the word ended with and press the corresponding button. You will hear a beep between items. You will only be played one item at a time, and will hear each item only once. Try to be as accurate as possible, but don't ponder an item for too long or worry about having made a mistake, once you have pressed a button you cannot change your decision. We collect data by recording your responses as well as response times. The entire experiment is expected to last for 20 minutes, including set up, and the opportunity to ask questions both before and after the task has been completed. The experiment will take place in the Language and Brain Laboratory (37, Wellington Square, OX1 2JF).

I hope that you will find the opportunity to learn more about linguistics and language perception enriching. If you would like to raise any concerns that you may have, or would like to make a complaint at any point, please speak to myself (07714981147) or my supervisor Professor Lahiri

(+44-1865-280401) and we will do our best to answer your query. We should acknowledge your concern within 10 working days and give you an indication of how we intend to deal with it. If you remain unhappy, you will be able to contact the chair of the Research Ethics Committee at the University of Oxford (Chair, Social Sciences & Humanities Inter-Divisional Research Ethics Committee; Email: ethics@socsci.ox.ac.uk; Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD). The chair will seek to resolve the matter in a reasonably expeditious manner.

All research data and records will be stored for a minimum retention period of 3 years after the publication or public release of the research. This study has been reviewed by, and received, ethics clearance from the University of Oxford Central University Research Ethics Committee. The University of Oxford is committed to the dissemination of this research for the benefit of society and the economy and, in support of this commitment, has established an online archive of research materials. This archive includes digital copies of student theses successfully submitted as part of a University of Oxford postgraduate degree programme. Holding the archive online gives easy access for researchers to the full text of freely available theses, thereby increasing the likely impact and use of that research. If you agree to participate in this study, the research will be written up as a thesis. On successful submission of the thesis, it will be deposited both in print and online to the University archives, to facilitate its use in future research. This thesis will be published open access. The data obtained in this study will be anonymised and used for research purposes only. Anonymity means that you will receive a unique participant identification number and that this participant identification number will not be associated with your first or last name. The data obtained will be available to myself, my supervisor, and the examiners of the DPhil thesis. The data will be stored on the secure hard drive of the PC in the Oxford Language & Brain Laboratory, the secure University of Oxford server, as well as on my own personal password protected computer.

If you have any questions, please ask. I hope that you enjoy the task!

FACULTY OF LINGUISTICS, PHILOLOGY AND PHONETICS

Miss Zoe Fletcher
Centre for Linguistics and Philology
Walton Street, Oxford OX1 2HG
Tel: 07714981147
zoe.fletcher@ling-phil.ox.ac.uk

Identification Task for Research into Language Perception

Consent Form

DPhil student Zoe Fletcher (Supervisor: Prof. Aditi Lahiri)
Contact: zoe.fletcher@ling-phil.ox.ac.uk

I, _____, hereby declare that (please initial each statement on the line provided)

- I have read the information sheet on this experiment and therefore agree to participate in this study. ____
- I understand that this consent is being given subject to normal legal requirements. ____
- I was given the opportunity to ask questions about the study and all of my questions and queries were answered to my full satisfaction. ____
- I know that I can withdraw from this study any time without penalty and without giving a reason for my decision. ____
- I understand that this project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee. ____
- I am aware that the anonymised results of the task will be used for research purposes and will thus be available to the student researcher, her supervisor and the examiners according to the Data Protection Act. The results of this study may be stored and published in the Oxford University Research Archive. ____
- I have also been given information on how to raise a concern and make a complaint. ____
- I agree to take part in the above study. ____

Optional (please initial each statement on the line provided, should you wish to do so):

- I do not object to my data being used in future publications in scientific journals, presentations or conferences. ____

- I agree for research data collected in this study to be given to researchers, including those working outside of the EU, to be used in other research studies. I understand that any data that leaves the research group will be fully anonymised so that I cannot be identified. ____
- I agree for my personal data to be kept in a secure database for the purpose of contacting me about future studies. ____

Date: _____ Signature: _____

Age: _____ Gender:_____

Native Language(s): _____

Level of English (basic/independent/proficient) _____

Any other languages (and level to which you speak them): _____

I, the researcher Zoe Fletcher, hereby declare that I confirm and witness the giving of consent from the participant.

Date: _____ Signature: _____

APPENDIX E

Experiment 5: information sheet and consent form

FACULTY OF LINGUISTICS, PHILOLOGY AND PHONETICS

Ms Zoe Fletcher
Centre for Linguistics and Philology Walton Street,
Oxford,
OX1 2HG
Tel: 07714981147
zoe.fletcher@ling-phil.ox.ac.uk

Lexical Decision Task for Research into Language Perception

Information Sheet

DPhil student Zoe Fletcher (Supervisor: Prof. Aditi Lahiri)
Contact: zoe.fletcher@ling-phil.ox.ac.uk

This study looks into speech perception and can help us understand how one of the most fundamental characteristics of the human species – language- is perceived by the brain.

You are invited to take part in this experiment as you are a native speaker of Southern British English or native German who is over the age of 18, and have no known hearing, language, or learning difficulties. The following instructions will give you a description of what this experiment involves and what you will be required to do. You will be able to ask questions before the experiment. Should you decide to take part, you will be asked to sign a consent form, but please bear in mind that you are free to withdraw from the experiment at any point should you wish to do so.

If you are happy to take part in the study you will be asked to put on a pair of headphones and sit in front of a computer screen, attached to which will be a handset with a choice of two buttons, one reading 'YES' and one reading 'NO'. You will hear a series of beeps before hearing some auditory output, you will then see a word or non-word on your screen. You must press 'YES' if you think that what you see on the screen is a word in English, and conversely press 'NO' if you do not think that what you see on the screen is a word in English. We will have a practice session first so that you can familiarise yourself with the format of this experiment. We collect data by recording your responses as well as response times. The entire experiment is expected to last for 30 minutes, including set up, and the opportunity to ask questions both before and after the task has been completed. The experiment will take place in the Language and Brain Laboratory (37, Wellington Square, OX1 2JF).

I hope that you will find the opportunity to learn more about linguistics and language perception enriching. If you would like to raise any concerns that you may have, or would like to make a complaint at any point, please speak to myself (07714981147) or my supervisor Professor Lahiri (+44-1865-280401) and we will do our best to answer your query. We should acknowledge your concern within 10 working days and give you an indication of how we intend to deal with it. If

you remain unhappy, you will be able to contact the chair of the Research Ethics Committee at the University of Oxford (Chair, Social Sciences & Humanities Inter-Divisional Research Ethics Committee; Email: ethics@socsci.ox.ac.uk; Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD). The chair will seek to resolve the matter in a reasonably expeditious manner.

All research data and records will be stored for a minimum retention period of 3 years after the publication or public release of the research. This study has been reviewed by, and received, ethics clearance from the University of Oxford Central University Research Ethics Committee. The University of Oxford is committed to the dissemination of this research for the benefit of society and the economy and, in support of this commitment, has established an online archive of research materials. This archive includes digital copies of student theses successfully submitted as part of a University of Oxford postgraduate degree programme. Holding the archive online gives easy access for researchers to the full text of freely available theses, thereby increasing the likely impact and use of that research. If you agree to participate in this study, the research will be written up as a thesis. On successful submission of the thesis, it will be deposited both in print and online to the University archives, to facilitate its use in future research. This thesis will be published open access. The data obtained in this study will be anonymised and used for research purposes only. Anonymity means that you will receive a unique participant identification number and that this participant identification number will not be associated with your first or last name. The data obtained will be available to myself, my supervisor, and the examiners of the DPhil thesis. The data will be stored on the secure hard drive of the PC in the Oxford Language & Brain Laboratory, the secure University of Oxford server, as well as on my own personal password protected computer.

If you have any questions, please ask. I hope that you enjoy the task!

FACULTY OF LINGUISTICS, PHILOLOGY AND PHONETICS

Miss Zoe Fletcher
Centre for Linguistics and Philology
Walton Street, Oxford OX1 2HG
Tel: 07714981147
zoe.fletcher@ling-phil.ox.ac.uk

Lexical Decision Task for Research into Language Perception

Consent Form

DPhil student Zoe Fletcher (Supervisor: Prof. Aditi Lahiri)
Contact: zoe.fletcher@ling-phil.ox.ac.uk

I, _____, hereby declare that (please initial each statement on the line provided)

- I have read the information sheet on this experiment and therefore agree to participate in this study. ____
- I understand that this consent is being given subject to normal legal requirements. ____
- I was given the opportunity to ask questions about the study and all of my questions and queries were answered to my full satisfaction. ____
- I know that I can withdraw from this study any time without penalty and without giving a reason for my decision. ____
- I understand that this project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee. ____
- I am aware that the anonymised results of the task will be used for research purposes and will thus be available to the student researcher, her supervisor and the examiners according to the Data Protection Act. The results of this study may be stored and published in the Oxford University Research Archive. ____
- I have also been given information on how to raise a concern and make a complaint. ____
- I agree to take part in the above study. ____

Optional (please initial each statement on the line provided, should you wish to do so):

- I do not object to my data being used in future publications in scientific journals, presentations or conferences. ____

- I agree for research data collected in this study to be given to researchers, including those working outside of the EU, to be used in other research studies. I understand that any data that leaves the research group will be fully anonymised so that I cannot be identified. ____
- I agree for my personal data to be kept in a secure database for the purpose of contacting me about future studies. ____

Date: _____ Signature: _____

Age: _____ Gender:_____

Native Language(s): _____

Level of English (basic/independent/proficient) _____

Any other languages (and level to which you speak them): _____

I, the researcher Zoe Fletcher, hereby declare that I confirm and witness the giving of consent from the participant.

Date: _____ Signature: _____

APPENDIX F

Experiment 5: full list of primes and targets[26]

	V	**VL**	**V**	**VL**
Orthographic Prime	**CHIDE**	*CHITE	JAZZ	*JASS
Auditory Prime	**[ʧaɪ:]**	*[ʧaɪ]	[ʤæ:]	*[ʤæ]
Underlying Prime	**/ʧaɪ:d/**	*/ʧaɪt/	/ʤæ:z/	*/ʤæs/
Target	SCOLD	SCOLD	SCOLD	SCOLD
Orthographic Prime	ENGAGE	*ENGAICH	STOVE	*STOAF
Auditory Prime	[ɪngeɪ:]	*[ɪngeɪ]	[stəʊ:]	*[stəʊ]
Underlying Prime	/ɪngeɪ:ʤ/	*/ɪngeɪʧ/	/stəʊ:v/	*/stəʊf/
Target	*TROYP	*TROYP	*TROYP	*TROYP
	VL	**V**	**VL**	**V**
Orthographic Prime	**VOTE**	*VOAD	KNIFE	*KNIVE
Auditory Prime	**[vəʊ]**	*[vəʊ:]	[naɪ]	*[naɪ:]
Underlying Prime	**/vəʊ/**	*/vəʊ:d/	/naɪf/	*/naɪ:v/
Target	POLL	POLL	POLL	POLL
Orthographic Prime	FLASH	*FLAZH	CAPE	*CAIB
Auditory Prime	[flæ]	*[flæ:]	[keɪ]	*[keɪ:]
Underlying Prime	/flæʃ/	*/flæ:ʒ/	/keɪp/	*/keɪ:b/
Target	*RISP	*RISP	*RISP	*RISP
	V	**VL**	**V**	**VL**
Orthographic Prime	**TAG**	*TAK	BABE	*BAIP
Auditory Prime	**[tæ:]**	*[tæ]	[beɪ:]	*[beɪ]
Underlying Prime	**/tæ:g/**	*/tæk/	/beɪ:b/	*/beɪp/
Target	LABEL	LABEL	LABEL	LABEL
Orthographic Prime	ROSE	*ROAS	GLIDE	*GLITE
Auditory Prime	[rəʊ:]	*[rəʊ]	[glaɪ:]	*[glaɪ]
Underlying Prime	/rəʊ:z/	*/rəʊs/	/glaɪ:d/	*/glaɪt/
Target	*DALGO	*DALGO	*DALGO	*DALGO
	VL	**V**	**VL**	**V**
Orthographic Prime	**CLOAK**	*CLOAG	MATCH	*MADGE
Auditory Prime	**[kləʊ]**	[kləʊ:]	[mæ]	*[mæ:]
Underlying Prime	**/kləʊk/**	/kləʊ:g/	/mæʧ/	*/mæ:ʤ/
Target	ROBE	ROBE	ROBE	ROBE
Orthographic Prime	SHAPE	*SHAIB	KITE	*KIDE
Auditory Prime	[ʃeɪ]	*[ʃeɪ:]	[kaɪ]	*[kaɪ:]
Underlying Prime	/ʃeɪp/	*/ʃeɪ:b/	/kaɪt/	*/kaɪ:d/
Target	*VOTH	*VOTH	*VOTH	*VOTH

[26] Words found in bold are the related real word primes

	V	**VL**	**V**	**VL**
Orthographic Prime	**CRAVE**	*CRAIF	STROBE	*STROAP
Auditory Prime	**[kreɪ:]**	*[kreɪ]	[strəʊ:]	*[strəʊ]
Underlying Prime	**/kreɪ:v/**	*/kreɪf/	/strəʊ:b/	*/strəʊp/
Target	YEARN	YEARN	YEARN	YEARN
Orthographic Prime	GAG	*GAK	PRIDE	*PRITE
Auditory Prime	[gæ:]	*[gæ]	[praɪ:]	*[praɪ]
Underlying Prime	/gæ:g/	*/gæk/	/praɪ:d/	*/praɪt/
Target	*BLAFT	*BLAFT	*BLAFT	*BLAFT
	VL	**V**	**VL**	**V**
Orthographic Prime	**CASH**	*CAZH	FIGHT	*FIDE
Auditory Prime	**[kæ]**	*[kæ:]	[faɪ]	*[faɪ:]
Underlying Prime	**/kæʃ/**	*/kæ:ʒ/	/faɪt/	*/faɪ:d/
Target	MONEY	MONEY	MONEY	MONEY
Orthographic Prime	SNAKE	*SNAIG	COPE	*COAB
Auditory Prime	[sneɪ]	*[sneɪ:]	[kəʊ]	*[kəʊ:]
Underlying Prime	/sneɪk/	*/sneɪ:g/	/kəʊp/	*/kəʊ:b/
Target	*JUNDE	*JUNDE	*JUNDE	*JUNDE
	V	**VL**	**V**	**VL**
Orthographic Prime	**COVE**	*COAF	ARISE	*ARICE
Auditory Prime	**[kəʊ:]**	*[kəʊ]	[araɪ:]	*[araɪ]
Underlying Prime	**/kəʊ:v/**	*/kəʊf/	/araɪ:z/	*/araɪs/
Target	BAY	BAY	BAY	BAY
Orthographic Prime	BEIGE	*BAISH	STAB	*STAP
Auditory Prime	[beɪ:]	*[beɪ]	[stæ:]	*[stæ]
Underlying Prime	/beɪ:ʒ/	*/beɪʃ/	/stæ:b/	*/stæp/
Target	*FAW	*FAW	*FAW	*FAW
	VL	**V**	**VL**	**V**
Orthographic Prime	**STRIKE**	*STRIGE	TAPE	*TAIB
Auditory Prime	**[straɪ]**	*[straɪ:]	[teɪ]	*[teɪ:]
Underlying Prime	**/straɪk/**	*/straɪ:g/	/teɪp/	*/teɪ:b/
Target	HIT	HIT	HIT	HIT
Orthographic Prime	MASS	*MAZZ	DOTE	*DOAD
Auditory Prime	[mæ]	*[mæ:]	[dəʊ]	*[dəʊ:]
Underlying Prime	/mæs/	*/mæ:z/	/dəʊt/	*/dəʊ:d/
Target	*WIR	*WIR	*WIR	*WIR

	V	**VL**	**V**	**VL**
Orthographic Prime	**DAD**	*DAT	POSE	*POAS
Auditory Prime	**[dæ:]**	*[dæ]	[pəʊ:]	*[pəʊ]
Underlying Prime	**/dæ:d/**	*/dæt/	/pəʊ:z/	*/pəʊs/
Target	FATHER	FATHER	FATHER	FATHER
Orthographic Prime	SHAVE	*SHAIF	IMBIBE	*IMBIPE
Auditory Prime	[sheɪ:]	*[sheɪ]	[ɪmbaɪ:]	*[ɪmbaɪ]
Underlying Prime	/sheɪ:v/	*/sheɪf/	/ɪmbaɪ:b/	*/ɪmbaɪp/
Target	*WIPLOL	*WIPLOL	*WIPLOL	*WIPLOL
	VL	**V**	**VL**	**V**
Orthographic Prime	**CRATE**	*CRAID	YAK	*YAG
Auditory Prime	**[kreɪ]**	*[kreɪ:]	[jæ]	*[jæ:]
Underlying Prime	**/kreɪt/**	*/kreɪ:d/	/jæk/	*/jæ:g/
Target	BOX	BOX	BOX	BOX
Orthographic Prime	VICE	*VIZE	MOPE	*MOAB
Auditory Prime	[vaɪ]	*[vaɪ:]	[məʊ]	*[məʊ:]
Underlying Prime	/vaɪs/	*/vaɪ:z/	/məʊp/	*/məʊ:b/
Target	*YIP	*YIP	*YIP	*YIP
	V	**VL**	**V**	**VL**
Orthographic Prime	**HAVE**	*HAF	CAGE	*CAICH
Auditory Prime	**[hæ:]**	*[hæ]	[keɪ:]	*[keɪ]
Underlying Prime	**/hæ:/**	*/hæ/	/keɪ:/	*/keɪ/
Target	POSSESS	POSSESS	POSSESS	POSSESS
Orthographic Prime	SCRIBE	*SCRIPE	ERODE	*EROAT
Auditory Prime	[skraɪ:]	*[skraɪ]	[ɪrəʊ:]	*[ɪrəʊ]
Underlying Prime	/skraɪ:b/	*/skraɪp/	/ɪrəʊ:d/	*/ɪrəʊt/
Target	*LISTRAI	*LISTRAI	*LISTRAI	*LISTRAI
	VL	**V**	**VL**	**V**
Orthographic Prime	**COACH**	*COAGE	BAKE	*BAIG
Auditory Prime	**[kəʊ]**	*[kəʊ:]	[beɪ]	*[beɪ:]
Underlying Prime	**/kəʊʧ/**	*/kəʊ:ʤ/	/beɪk/	*/beɪ:g/
Target	BUS	BUS	BUS	BUS
Orthographic Prime	GRIPE	*GRIBE	VAT	*VAD
Auditory Prime	[graɪ]	*[graɪ:]	[væ]	*[væ:]
Underlying Prime	/graɪp/	*/graɪ:b/	/væt/	*/væ:d/
Target	*TIJ	*TIJ	*TIJ	*TIJ

	V	**VL**	**V**	**VL**
Orthographic Prime	**BLAZE**	*BLAIS	ABIDE	*ABITE
Auditory Prime	**[bleɪ:]**	*[bleɪ]	[əbaɪ:]	*[əbaɪ]
Underlying Prime	**/bleɪ:z/**	*/bleɪs/	/əbaɪ:d/	*/əbaɪt/
Target	FIRE	FIRE	FIRE	FIRE
Orthographic Prime	JAB	*JAP	CLOVE	*CLOAF
Auditory Prime	[ʤæ:]	*[ʤæ]	[kləʊ:]	*[kləʊ]
Underlying Prime	/ʤæ:b/	*/ʤæp/	/kləʊ:v/	*/kləʊf/
Target	*SHRO	*SHRO	*SHRO	*SHRO
	VL	**V**	**VL**	**V**
Orthographic Prime	**FLAT**	*FLAD	POPE	*POAB
Auditory Prime	**[flæ]**	*[flæ:]	[pəʊ]	*[pəʊ:]
Underlying Prime	**/flæt/**	*/flæ:d/	/pəʊp/	*/pəʊ:b/
Target	LEVEL	LEVEL	LEVEL	LEVEL
Orthographic Prime	BASE	*BAIZ	SPIKE	*SPIGE
Auditory Prime	[beɪ]	*[beɪ:]	[spaɪ]	*[spaɪ:]
Underlying Prime	/beɪs/	*/beɪ:z/	/spaɪk/	*/spaɪ:g/
Target	*PLODI	*PLODI	*PLODI	*PLODI
	V	**VL**	**V**	**VL**
Orthographic Prime	**CLOSE**	*CLOAS	GLADE	*GLAIT
Auditory Prime	**[kləʊ:]**	*[kləʊ]	*[gleɪ:]	*[gleɪ]
Underlying Prime	**/kləʊ:z/**	*/kləʊs/	*/gleɪ:d/	*/gleɪt/
Target	SHUT	SHUT	SHUT	SHUT
Orthographic Prime	BRAG	*BRAK	CHIVE	*CHIFE
Auditory Prime	[bræ:]	*[bræ]	[ʧaɪ:]	*[ʧaɪ]
Underlying Prime	/bræ:g/	*/bræk/	/ʧaɪ:v/	*/ʧaɪf/
Target	*RIPU	*RIPU	*RIPU	*RIPU
	VL	**V**	**VL**	**V**
Orthographic Prime	**ALIKE**	*ALIGE	MAP	*MAB
Auditory Prime	**[əlaɪ]**	*[əlaɪ:]	[mæ]	*[mæ:]
Underlying Prime	**/əlaɪk/**	*/əlaɪ:g/	/mæp/	*/mæ:b/
Target	SAME	SAME	SAME	SAME
Orthographic Prime	BAIT	*BAID	POACH	*POAGE
Auditory Prime	[beɪ]	*[beɪ:]	[pəʊ]	*[pəʊ:]
Underlying Prime	/beɪt/	*/beɪ:d/	/pəʊʧ/	*/pəʊ:ʤ/
Target	*PLIC	*PLIC	*PLIC	*PLIC

	V	**VL**	**V**	**VL**
Orthographic Prime	**ABODE**	*ABOAT	JIVE	*JIFE
Auditory Prime	**[əbəʊ:]**	*[əbəʊ]	[ʤaɪ:]	*[ʤaɪ]
Underlying Prime	**/əbəʊ:d/**	*/əbəʊt/	/ʤaɪ:v/	*/ʤaɪf/
Target	HOUSE	HOUSE	HOUSE	HOUSE
Orthographic Prime	GRAB	*GRAP	VAGUE	*VAIK
Auditory Prime	[græ:]	*[græ]	[veɪ:]	*[veɪ]
Underlying Prime	/græ:b/	*/græp/	/veɪ:g/	*/veɪk/
Target	*RIPAN	*RIPAN	*RIPAN	*RIPAN
	VL	**V**	**VL**	**V**
Orthographic Prime	**NICE**	*NIZE	CHAFE	*CHAIV
Auditory Prime	**[naɪ]**	*[naɪ:]	[ʧeɪ]	*[ʧeɪ:]
Underlying Prime	**/naɪs/**	*/naɪ:z/	/ʧeɪf/	*/ʧeɪ:v/
Target	GOOD	GOOD	GOOD	GOOD
Orthographic Prime	RAP	*RAB	QUOTE	*QUOAD
Auditory Prime	[ræ]	*[ræ:]	[kwəʊ]	*[kwəʊ:]
Underlying Prime	/ræp/	*/ræ:b/	/kwəʊt/	*/kwəʊ:d/
Target	*SREZ	*SREZ	*SREZ	*SREZ
	V	**VL**	**V**	**VL**
Orthographic Prime	**LAD**	*LAT	BRIBE	*BRIPE
Auditory Prime	**[læ:]**	*[læ]	[braɪ:]	*[braɪ]
Underlying Prime	**/læ:d/**	*/læt/	/braɪ:b/	*/braɪp/
Target	BOY	BOY	BOY	BOY
Orthographic Prime	VOGUE	*VOAK	CRAZE	*CRAIS
Auditory Prime	[vəʊ:]	*[vəʊ]	[kreɪ:]	*[kreɪ]
Underlying Prime	/vəʊ:g/	*/vəʊk/	/kreɪ:z/	*/kreɪs/
Target	*NIF	*NIF	*NIF	*NIF
	VL	**V**	**VL**	**V**
Orthographic Prime	**JOKE**	*JOAG	SPACE	*SPAIZ
Auditory Prime	**[ʤəʊ]**	*[ʤəʊ:]	[speɪ]	*[speɪ:]
Underlying Prime	**/ʤəʊk/**	*/ʤəʊ:g/	/speɪs/	*/speɪ:z/
Target	FUNNY	FUNNY	FUNNY	FUNNY
Orthographic Prime	SWIPE	*SWIBE	CATCH	*CADGE
Auditory Prime	[swaɪ]	*[swaɪ:]	[kæ]	*[kæ:]
Underlying Prime	/swaɪp/	*/swaɪ:b/	/kæʧ/	*/kæ:ʤ/
Target	*IPSOP	*IPSOP	*IPSOP	*IPSOP