THE EFFECT OF TALKER NATIVENESS ON THE PERCEPTION OF VOICING IN SYLLABLE-INITIAL PLOSIVES BY AMERICAN ENGLISH MONOLINGUALS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

LINGUISTICS

AUGUST 2019

By

Peter Chong Jian Li

Dissertation Committee:

Katie Drager, Chairperson
Kamil Deen
Bradley McDonnell
Rory Turnbull
Grayden Solman

Keywords: Non-native accent, Speech perception, Sociophonetics

Acknowledgement

This dissertation would not have been possible without the community around me. I truly stand on the shoulders of giants. First and foremost I would like to thank my dear advisor, Katie Drager. Without your advice and steady guidance I would not be able to get started on this work, let alone finish. I would also like to extend my gratitude to the rest of my committee: Kamil Deen, for your wise advice on all matters pertaining to academia; Brad McDonnell, for your input on an earlier version of this study and of course for the dissertation; Rory Turnbull, for your encyclopedic knowledge on phonology and R; Grayden Solman, for willing to help in my time of need. I would also like to thank the other faculty members of the Department of Linguistics, especially Amy Schafer, who provided valuable advice during the early stages of this study.

I would also not be able to cope with the rigors of graduate school and life away from home without the love and support of my friends. Amber, you are truly a gift, without your help and counsel, I would not be able to finish half my work; Yen-Ling, your sense of humour was very much appreciated; Chris, I appreciate your experience and camaraderie. Bonnie, Jonny, Wenyi, Nozomi, Sejung, Paulina, and many more, it was a blessing to know you all.

I would also like to extend my appreciation to the government of Malaysia and Universiti Sains Malaysia, without whose generous financial support I would not be in graduate school.

Last but not least, I would like to thank my parents and my brother and sister for their support and love from afar, without whose strength I would not have been able to endure this journey.

ABSTRACT

Talker nativeness has been shown to influence listeners' perception and comprehension of speech (Floccia, Butler, Goslin, & Ellis, 2009; Rubin, 1992). Previous research on the effect of talker nativeness has focused on larger linguistic units such as lexical items and sentences. The current dissertation investigates whether American English monolinguals' perception of syllable-initial bilabial plosives differed according to when they believe a syllable is produced by a native talker or a non-native talker. This question was explored through a series of three experiments all using the binary forced choice syllable identification task. Specifically, these experiments test whether the native listeners' perception of bilabial plosives is different when an identical syllable continuum is inserted into a frame sentence produced by a native speaker of American English versus a frame sentence produced by a non-native speaker.

In Experiment One, a syllable continuum constructed from natural /pa/ and /ba/ tokens produced by a female native speaker of American English was inserted into a frame sentence produced by the same native speaker (native talker condition) and a frame sentence produced by a Mexican woman who is a native speaker of Spanish (non-native talker condition); the talkers varied across condition but the continuum was produced by a single talker in order to control for acoustic differences inherent in syllables produced by different talkers and to ensure the VOT range of the syllable continuum is equal across both talker conditions. The analysis focuses on responses from 25 participants who were American English monolinguals. The results show that participants are more likely to perceive syllables in the non-native talker condition as /pa/ than those in the native talker condition.

The results of Experiment One provide evidence that whether a talker is a native speaker affects how their bilabial plosives are perceived. However, it is possible that the findings were a result of participants using voice onset time (VOT) cues present in the talkers' frame sentences. Experiment Two aimed to determine if talker nativeness had an effect even when the frame sentences for both talkers were modified to have identical VOT cues in word initial bilabial plosives. Responses from a new set of 25 monolingual participants were analysed. The results show that participants are still more likely to perceive syllables in the non-native talker condition as /pa/ than those in the native talker condition.

In Experiment Three, a native speaker of Mandarin from China produced the frame sentence in the non-native talker condition. This was done to test whether the talker nativeness effect observed in Experiments One and Two was present when a different non-native accent is used. Responses from 25 monolingual participants who did not participate in any of the previous experiments were analysed. The results demonstrate that participants responded differently to items in the two conditions and were more likely to perceive syllables in the native talker condition as /pa/ than those in the non-native talker condition.

In conclusion, this dissertation found that like other types of talker-related information such as age and gender, talker nativeness has an effect on listeners' perceptions of speech sounds. The findings of the dissertation are also consistent with experience-based models of speech perception.

Table of Contents

Acknowledgement....iii
ABSTRACT....iv
Table of Contents....vi
List of Tables....viii
List of Figures....ix
Chapter 1 Introduction....1
1.1 Introduction....1
1.2 Research Questions....2
1.3 Structure of the dissertation....2
Chapter 2 Background....4
2.1 Talker-Related Information, and Speech Perception....5
2.1.1 Listener Experience and Perception....7
2.1.2 Top-down Social Information Affecting the Perception of Sounds....9
2.1.3 Talker-Related Information Inherent in a Signal Affecting the Perception of Sounds....10
2.1.4 Accent....12
2.2 Acoustic Properties of English Word-Initial Plosives....17
2.3 The Hypotheses....20
Chapter 3 Experiment One....21
3.1 Participants....21
3.2 Materials and Procedure....22
3.3 Results....31
3.3.1 Participants who Reported Noticing Splicing in Stimuli....36
3.3.2 Participants who Perceived Non-native Talker as an L1 Spanish Speaker....38
3.4 Discussion....39
Chapter 4 Experiment Two....42
4.1 Participants....43
4.2 Materials and Procedure....44
4.2.1 Other Acoustic Cues....45
4.3 Results....46
4.3.1 Participants who Reported Noticing Splicing in Stimuli....50
4.3.2 Participants who Perceived the Non-native Talker as an L1 Spanish Speaker....51
4.4 Comparison of Experiments One and Two....53
4.5 Discussion....55
Chapter 5 Experiment Three....59
5.1 Participants....60
5.2 Materials and Procedure....61
5.2.1 Other Acoustic Cues....62
5.3 Results....63
5.3.1 Participants who Reported Noticing Splicing in Stimuli....67
5.3.2 Participants who Perceived the Non-native Talker as an L1 Mandarin Speaker....68
5.4 Comparison of Experiments Two and Three....69
5.5 Discussion....70

Chapter 6 General Discussion..74
6.1 Summary of Experiment Results..74
6.1.1 Effect of Block...77
6.2 Implications...77
6.3 Limitations..81
6.4 Overall Summary...81
Chapter 7 Conclusion...83
Appendix A: Language Background Questionnaire...85
Appendix B: Background of Monolingual Native Listeners.......................................95
References...98

List of Tables

Table 3.1: Breakdown of monolingual native speaker participants according to gender, age, and their response when identifying the non-native talker's L1..........................22
Table 3.2: VOT (shown in ms) of critical syllable steps...27
Table 3.3: Output of logistic mixed effects model on responses to both talkers' VOT steps in Experiment One. Model: glmer(Critical Responses ~ Talker + VOT Step + Block + Talker:Block + VOT Step:Block + (1|Participant), data = Native and Non-native Talkers, family = binomial)..36
Table 4.1: Breakdown of monolingual native speaker participants according to gender, age, and their response for the non-native talker's L1...43
Table 4.2: F0 is shown in Hertz (Hz) and Bark for the vowels in *please* and *pick* for the native and non-native talkers, and critical syllable (produced by native talker)....46
Table 4.3: Output of logistic mixed effects model on responses to both talkers' VOT steps in Experiment Two. Model: glmer(Critical Responses ~ Talker + VOT Step + Block + (1|Participant), data = Native and Non-native Talkers, family = binomial)…49
Table 5.1: Breakdown of monolingual native speaker participants' demographic information according to gender, age, and their response for the non-native talker's L1...61
Table 5.2: F0 is shown in Hertz (Hz) and Bark for the vowels in *please* and *pick* for the native and non-native talkers, and critical syllable (produced by native talker)....63
Table 5.3: Output of logistic mixed effects model on responses to both talkers' VOT steps in Experiment Three. Model: glmer(Critical Responses ~ Talker*VOT Step*Block + (VOT Step|Participant), data = Native and Non-native Talkers, family = binomial)..67
Table 6.1: Summary of results across experiments...74

List of Figures

Figure 3.1: Spectrogram and waveform from Praat showing the first phase of the resynthesis process in which a 40ms segment (i.e., step 8) was taken from the onset of the /pa/ sample....26
Figure 3.2: Spectrogram and waveform from Praat showing the second phase of the resynthesis process for step 8 (40ms) in which the selected segment was pasted onto the onset of the /ba/ sample....26
Figure 3.3: Spectrogram and waveform from Praat showing the third phase of the resynthesis process for step 8 (40ms) in which the segment with roughly the same duration as the inserted segment was selected and removed. Duration was controlled to the extent possible while restricting splicing to zero crossover points to maintain naturalness....27
Figure 3.4: Spectrogram and waveform from Praat showing the splicing process in which the critical syllable was spliced into the frame sentence with approximately 450ms of silence before and after the syllable....28
Figure 3.5: Picture showing the keys used to log responses to the audio stimuli. Participants pressed F if they perceive /ba/ and J if they perceive /pa/....30
Figure 3.6: Overall proportion of /pa/ responses for critical items in Experiment One, averaged across participants and blocks and shown separately for the talker who is a native speaker of English (native) and the one who is not (non-native). Voice onset time (VOT) from the relevant step of each continuum is shown in milliseconds (ms) on the x-axis. 95% confidence intervals for each participant's mean responses across blocks are shown....32
Figure 3.7: Proportion of /pa/ responses for critical items in Experiment One across the eight experimental blocks, averaged across participants....33
Figure 3.8: Proportion of /pa/ responses for critical items in Experiment One across participants, averaged across blocks....34
Figure 3.9: Proportion of /pa/ responses for critical items in Experiment One, averaged across participants and blocks. Proportions are shown separately for participants who reported noticing the splice after being asked explicitly (grey, N=6) and those who reported that they did not notice the splice (black, N=19), and for whether the talker was a native speaker of English (solid) or not (dotted)....37
Figure 3.10: Proportion of /pa/ responses for critical items in Experiment One, averaged across participants and blocks. Proportions are shown separately for participants who responded perceiving the non-native talker as having Spanish as L1 (grey, N=8) and those who responded that the non-native talker has an L1 other than Spanish (black, N=17), and for whether the talker was a native speaker of English (solid) or not (dotted)....39
Figure 4.1: The VOT for the native talker's *please* was equalised with the non-native talker's by removing a segment from the end of the VOT portion....45
Figure 4.2: Overall proportion of /pa/ responses for critical items in Experiment Two, averaged across participants and blocks and shown separately for the talker who is a native speaker of English (native) and the one who is not (non-native). Voice onset time (VOT) from the relevant step of each continuum is shown in milliseconds (ms) on the x-axis. 95% confidence intervals for each participant's mean responses across blocks are shown....47

Figure 4.3: Proportion of /pa/ responses for critical items in Experiment Two across the eight experimental blocks, averaged across participants.......................................48
Figure 4.4: Proportion of /pa/ responses for critical items in Experiment Two across participants, averaged across blocks..48
Figure 4.5: Proportion of /pa/ responses for critical items in Experiment Two, averaged across participants and blocks. Proportions are shown separately for participants who reported noticing the splice after being asked explicitly (grey, N=9) and those who reported that they did not notice the splice (black, N=16), and for whether the talker was a native speaker of English (solid) or not (dotted)..................51
Figure 4.6: Proportion of /pa/ responses for critical items in Experiment Two, averaged across participants and blocks. Proportions are shown separately for participants who responded perceiving the non-native talker as having Spanish as L1 (grey, N=8) and those who responded that the non-native talker has an L1 other than Spanish (black, N=17), and for whether the talker was a native speaker of English (solid) or not (dotted)...52
Figure 4.7: Overall proportion of /pa/ responses for critical items in Experiments One and Two, averaged across participants and blocks and shown separately across experiment, and for the talker who is a native speaker of English (native) and the one who is not (non-native). Voice onset time (VOT) from the relevant step of each continuum is shown in milliseconds (ms) on the x-axis. 95% confidence intervals for each participant's mean responses across blocks are shown.......................................54
Figure 5.1: Overall proportion of /pa/ responses for critical items in Experiment Three, averaged across participants and blocks and shown separately for the talker who is a native speaker of English (native) and the one who is not (non-native). Voice onset time (VOT) from the relevant step of each continuum is shown in milliseconds (ms) on the x-axis. 95% confidence intervals for each participant's mean responses across blocks are shown...63
Figure 5.2: Proportion of 'pa' responses for critical items in Experiment Three across the eight experimental blocks, averaged across participants.......................................64
Figure 5.3: Proportion of /pa/ responses for critical items in Experiment Three across participants, averaged across blocks..65
Figure 5.4: Proportion of /pa/ responses for critical items in Experiment Three, averaged across participants and blocks. Proportions are shown separately for participants who reported noticing the splice after being asked explicitly (grey, N=13) and those who reported that they did not notice the splice (black, N=12), and for whether the talker was a native speaker of English (solid) or not (dotted)..................68
Figure 5.5: Overall proportion of /pa/ responses for critical items in Experiments Two and Three, averaged across participants and blocks and shown separately according to Experiment Two (solid) and Experiment Three (dotted), and for the talker who is a native speaker of English (black), who is Spanish non-native (grey) and who is Chinese non-native(light blue). Voice onset time (VOT) from the relevant step of each continuum is shown in milliseconds (ms) on the x-axis. 95% confidence intervals for each participant's mean responses across blocks are shown.......................................70

Chapter 1 Introduction

1.1 Introduction

A growing body of research provides evidence that talker-related information influences a listener’s interpretation of speech. Perception can be affected by phonetic cues encountered earlier in an utterance (Ladefoged & Broadbent, 1957; Kakehi, 1992; Rubin, 1992; Rubin & Smith, 1990), by social information inferred from non-linguistic sources such as faces shown in photographs (Hay, Warren, & Drager, 2006; Johnson, Strand, & D’Imperio, 1999), and by socially-indexed phonetic cues present in the signal (Babel & McGuire, 2013; Munson, Jefferson, & McDonald, 2006). The present study investigates the effect of one particular type of talker-related information, talker[1] accent, on the perception of speech sounds.

In recent decades, research on non-native accents have found that these accents negatively affect a listener’s lexical access and general comprehension (Kang & Rubin, 2009; Lev-Ari, 2015b). While these findings show that listeners have relatively more difficulty understanding words and sentences produced by non-native speaker compared to native speakers, it is unclear what the mechanisms behind this difficulty are. Specifically, there is a lack of research investigating whether this difference between the perception of native and non-native speech is caused by a difference in perception of sub-lexical units, such as individual phones or syllables. On the other hand, studies using speech produced by native talkers have found that listeners do adjust their perception of speech sounds based on talker-specific cues (Kraljic, Samuel, & Brennan, 2007). These adjustments at the phonetic level have then been found to contribute towards lexical access and identification.

1 In this dissertation, I use talker to refer to someone whose voice was recorded for the perception experiment and speaker to refer to someone who speaks a particular variety more generally.

The objective of my dissertation is to address this gap in the non-native talker literature by adapting methods previously used in native talker research (Strand & Johnson, 1996). Specifically, this dissertation reports on results from a series of perception experiments that use a binary forced choice identification task to test whether American English listeners' perceptual boundary for bilabial plosives is affected by whether the talker is a native speaker of English or not.

1.2 Research Questions

The broad issue of whether a listener's perceptual boundary differs according to the nativeness of a talker can be divided into the following research questions which will be explored in this dissertation:

- What is a listener's perceptual boundary for non-native speech sounds in relation to that of native speech sounds?
- Is a listener's perceptual boundary for non-native speech sounds constructed based on a default internal representation of a non-native speaker or does a listener take into account acoustic cues present in the signal?
- Is a listener's perceptual boundary for non-native talkers specific to a particular type of non-native accent or does a listener use similar boundaries for non-native accents in general?

1.3 Structure of the dissertation

To address the questions outlined above, this dissertation is organised into seven chapters. The current chapter, Chapter 1, introduces the goals and rationale of the study, and outlines the structure of the dissertation. Chapter 2 is a literature review providing an overview of important research conducted on the role of talker-related information in the process of speech perception. Chapter 3 introduces the experimental methodology used in the study and reports the results of Experiment One which tests whether there are differences in how listeners perceive identical syllables which are spliced into frame sentences spoken by talkers with different accents. Chapter 4 reports the findings of Experiment Two which investigates whether listeners incorporate specific voice onset time (VOT) cues embedded in the frame sentence. Chapter 5 then reports the findings of Experiment Three which determines if effects found in Experiments One and Two can be observed with a different non-native accent. Chapter 6 consists of an overall discussion which compares the results of all the experiments and discusses their implications as a whole. Chapter 7 concludes the dissertation by summarising the findings of the study.

Chapter 2 Background

Speech perception is a seemingly automatic skill for competent language users; it usually requires little effort and happens almost instantaneously. Nevertheless, how speech perception works is not straightforward. In the course of speech perception, listeners decode the incoming signal, all the while taking into account non-linguistic contextual information, such as a talker's identity (Foulkes, 2010). The degree to which their perception is affected by a given context - and the nature of any effect - depends at least partially on their prior exposure to the talker (Derwing & Munro, 1997; McGowan, 2015). While listeners have been observed to perceive non-native accented speech differently in comparison with native accented speech at the level of larger linguistic units (Clarke & Garrett, 2004; Lev-Ari, 2015b; Porretta, Tucker, & Järvikivi, 2016), it is unclear how the perception of non-native accents compares to that of native accents for sub-lexical units.

In this chapter, I present previous work that informs the research presented in this dissertation. Section 2.1 covers how information about talker-based information affects how their speech is perceived. This section first begins with a brief overview of how a listener's experience can influence their perception (2.1.1). This is then followed by a discussion of studies that demonstrate that talker-related information can be derived from top-down sources (2.1.3) and extracted from the speech signal (2.1.4). Section 2.1.4 describes how non-native accents affect perception. Section 2.2 describes the acoustic properties of plosives in word-initial position. Finally, Section 2.3 outlines the hypotheses.

2.1 Talker-Related Information, and Speech Perception

A listener takes into account a broad range of talker-related information[2] in the process of speech perception. Talker-related information can be in the form of a talker's personal attributes, such as gender (Babel & McGuire, 2013; Johnson et al., 1999) and age (Drager, 2011; Kim, 2016), their idiosyncracies (Kraljic et al., 2007), and their membership in social categories (Eckert, 2008). This information can be inherent to the acoustic signal or inferred via non-linguistic means, such as photographs, video, and labels (Hay et al., 2006; Johnson, 2005; Johnson et al., 1999; Niedzielski, 1999; Rubin, 1992).

The effect of talker-related information is observed on various linguistic units, ranging from an entire discourse (Rubin, 1992), to sentences (Casasanto, 2008), to words (Kim, 2016), and down to sublexical units such as syllables and sound segments (Kraljic, Brennan, & Samuel, 2008; Podesva, Reynolds, Callier, & Baptiste, 2015). Rubin (1992) found that a recording of a lecture produced by a native speaker was perceived as less comprehensible, more foreign and more accented when it was presented with a more foreign looking face. Comprehension scores were lower for participants exposed to an Asian face compared to those exposed to a Caucasian face. Similarly, listeners have different sentence completion expectations when presented with faces of different ethnicities (Casasanto, 2008). Participants reacted faster to written sentence endings which fit a consonant cluster reduction interpretation of an earlier sentence portion when that portion is presented with a male African American face, while they reacted faster to endings which fit a non-consonant cluster reduction interpretation when the sentence portion is associated with a male Anglo American

2 Different scholars have used various terms to describe non-linguistic factors attributable to the talker, such as social information (Campbell-Kibler, 2009), indexical information (Foulkes, 2010), socioindexical information (McGowan, 2011). Talker-related information was chosen here as a more general term which can refer to more than the talker's social grouping and includes the talker's idiosyncracies.

face. This is in line with consonant cluster reduction being a feature that is more closely associated with African American men than Anglo American men. At the word level, listeners showed better word recognition accuracy and reaction times when talker voice age matched with the speaker age range a word is associated with (Kim, 2016). When talkers exhibit unexpected variation in their pronunciation of individual speech sounds, listeners are able to identify which of those talkers' pronunciations are driven by idiosyncratic characteristics and which are driven by other reasons, such as dialectal differences (Kraljic et al., 2008, 2007). In conditions where listeners were provided with visual cues that a talker's unusual pronunciations of sibilant fricatives were due to the talker holding a pen in her mouth, listeners did not adjust their perception of that talker's sibilants (Kraljic et al., 2007). In contrast, listeners who were exposed to visual cues which support the notion that the talker's unusual pronunciation is characteristic of the talker adjusted their perception to accomodate the talker's idiosyncracy. This adjustment was reflected in listeners in the characteristic visual cue condition perceiving more tokens in an /s/-/ʃ/ continuum as /ʃ/ in a forced choice category identification task. Accomodation was also found to occur for dialect-driven variation (Kraljic et al., 2008). Listeners who were exposed to variant pronunciation of sibilants in constrained phonological contexts consistent with Philadelphia English did not differ in their perception of sibilants with the control group. However, those who were exposed to variation in unconstrained contexts differed from the control group in their identification of items in the sibilant continuum. Even when the phonemic identity of a speech sound is not in doubt, there is evidence to show that talker identity matters in the social meaning listeners interpret from allophones. In a matched guise experiment, listeners were found to differ in the social meaning assigned to released and unreleased /t/ according to which American politician said them (Podesva et al., 2015). The results from all these studies indicate that listeners have expectations based on different groups and talker identities, and these expectations operate at all linguistic units.

Next, I review the factors which affect speech perception, ranging from factors internal to the listener, to social information assumed from non-linguistic sources, to social infomation inferred based on linguistic sources. I begin in Section 2.1.1 by outlining work that demonstrates that a listener's background - including their previous experience with languages, dialects and speakers - affects their perception. Section 2.1.2 discusses top-down information which affects the perception of sounds. Then in Section 2.1.3, I turn to a discussion of how listeners infer social information about a talker based on linguistic cues inherent in the signal.

2.1.1 Listener Experience and Perception

While the discussion up to this point has focused on talker-related information, listener characteristics have an important role to play in speech perception as well (Hanulíková, 2018). Perception is influenced by a listener's experience, which shapes their categorical representations (Tees & Werker, 1984) and affects their interpretation of talker-related information (Hanulíková, 2018; Kang & Rubin, 2009). While infants quickly lose the ability to discriminate sound contrasts which are not present in their native language (Werker & Tees, 1984), even brief exposure to another language early in childhood has been shown to give a listener an advantage in discriminating speech sounds in that foreign language years later (Oh, Au, & Jun, 2010; Oh, Jun, Knightly, & Au, 2003; Tees & Werker, 1984). College students who were only exposed to a heritage language for the first couple of years of life with no continued use after that were able to perform better than their novice peers in discriminating the speech sounds of their heritage language (Oh et al., 2003; Tees & Werker, 1984), in some cases even approaching the level of native speakers (Tees & Werker, 1984). There is an effect of early childhood exposure even for learners with less than a year's worth of exposure and subsequently raised outside of their heritage language community. Oh et al. (2010) found that Korean adoptees who were adopted by non-Korean-speaking families reliably outperformed novice Korean learners in phoneme identification. Further, the adoptees were adopted at a mean age of five months and had minimal exposure to Korean since then.

Further, studies (Bent & Bradlow, 2003; Hayes-Harb, Smith, Bent, & Bradlow, 2008) have found that non-native listeners have an advantage compared to native listeners in terms of identifying words spoken by non-native talkers. Specifically, native Mandarin listeners were better at identifying English words spoken by a native Mandarin talker compared to native English listeners (Hayes-Harb et al., 2008).

Likewise, an effect of experience is evidenced by research involving the environment a listener was raised in (Hanulíková, 2018; Hay, Drager, & Gibson, 2018). Listeners raised in a multilingual society such as the Netherlands did not perceive a difference in comprehensibility between foreign and native looking faces (Hanulíková, 2018) while listeners raised in a relatively monolingual society such as the United States of America do (Rubin, 1992). In a study on the perception of r-sandhi (Hay et al., 2018), listeners who were raised in non-rhotic speaking areas in New Zealand were more likely to perceive intrusive-r in a phoneme monitoring task than their rhotic counterparts. Other effects of experience include higher transcription accuracy when transcribing non-native accented speech (McGowan, 2015). Listeners who were rated as having more experience with Chinese and Chinese native speakers were able to more accurately transcribe English sentences spoken by a Chinese native speaker compared to their less experienced counterparts.

2.1.2 Top-down Social Information Affecting the Perception of Sounds

Listeners' perception of sounds can be influenced by exposure to social primes external to the talker's voice, such as pictures, videos, written labels, and stuffed toys (Hay & Drager, 2010; Hay et al., 2006; Johnson et al., 1999; Niedzielski, 1999; Rubin, 1992; Rubin & Smith, 1990; Strand & Johnson, 1996). For example, Strand & Johnson (1996) found that fricatives produced by non-prototypical male and female voices were perceived differently when those voices were paired with videos containing faces of different sexes. Specifically, in a binary forced choice paradigm task where synthetic fricatives were spliced onto natural coda produced by talkers judged to be prototypical and non-prototypical male and female talkers, participants' perceptual boundaries for /s/-/ʃ/ were found to shift depending on the gender[3] attributed to the talker. The saliency of visual stimuli was made even more apparent in Johnson et al. (1999) when faces of different genders were found to have an effect even on prototypical male and female voices. The gender effect is present even when talker-related information is inferred from less salient sources. Further, mere imagination was sufficient to provide listeners with the necessary talker gender; in addition to using videos depicting people of different genders, Johnson et al. (1999) also report results from a perception experiment where participants were instructed to visualise a male or female talker while a gender-ambiguous recording was being

3 While 'sex' is used to refer to 'male' and 'female' and 'gender' refers to 'men' and 'women', many previous works studying speech perception according to 'sex' or 'gender' (e.g. Johnson, 2005; Johnson, Strand, & D'Imperio, 1999; Strand, 1999) have conflated those categories and the levels associated with them. Here I use 'gender' when discussing the category as this appears to be the term favoured by many scholars in recent years (e.g. Drager, 2011; Johnson, 2005). When reporting on individual studies however, I report the terms according to how they were used in those studies.

played. Their results show that listeners perceive a gender-ambiguous recording as being a different vowel depending on whether they visualise the talker to be male or female. In a binary forced choice identification task using a *hood-hud* continuum, listeners were more likely to label an ambiguous vowel as *hood* when told that the talker was female. There is even evidence that top-down information might sometimes override acoustic evidence (Niedzielski, 1999). Niedzielski (1999) found that listeners from Detroit responded differently to identical stimuli spoken by a talker from Detroit based on labels claiming that the talker originated from Michigan or Canada; listeners in the Canada condition perceived the talker as producing shifted vowels even when speakers from Detroit also produce such shifted vowels. Further, the responses from listeners in the Michigan condition indicated that they associated the regional dialect with Standard American English.

2.1.3 Talker-Related Information Inherent in a Signal Affecting the Perception of Sounds

Besides top-down sources, listeners also take into account a talker's prior phonetic realizations (Kakehi, 1992; Kraljic et al., 2007; Norris, James M., & Cutler, 2003). Listeners adjust their perception depending on phonetic realizations previously encountered in the signal (Bradlow & Bent, 2008; Ladefoged & Broadbent, 1957). When listeners are exposed to introductory sentences prior to critical stimuli, their judgments on sound segments differ compared to conditions with only critical stimuli, providing evidence that acoustic information presented earlier in the signal affects perception (Ladefoged & Broadbent, 1957). While early research in this area focused on fully synthesized stimuli approximating native speech, recent studies have shown that native listeners are also able to make similar adjustments when perceiving natural non-native accented speech. Specifically, listeners can incorporate acoustic information from a single talker to make perceptual adaptations specific to that talker's speech, although multiple talkers are required for talker independent adaptation to occur (Bradlow & Bent, 2008). Cues on a talker's phonetic realizations can also be presented through a context-based training phase. In this case the ambiguous segment is incorporated into lexical items, which then signals to the listener that a talker has idiosyncratic pronunciation for this particular segment (Kraljic & Samuel, 2006; Kraljic et al., 2007; Norris et al., 2003). When a training phase is not available, listeners make an initial 'first impression' judgment for an unfamiliar speaker with potentially ambiguous pronunciations. This first impression is based on contextual information, such as lexical identity, which can then be subsequently adjusted (Kraljic et al., 2007). For example, an ambiguous sound is more likely to be perceived as the phone which fits an actual real word as opposed to a nonsense word. If the ambiguity of the sound is perceived as being an idiosyncratic characteristic specific to the talker, this interpretation will persist. However, if there is evidence showing the ambiguous pronunciation is incidental, such as the result of holding a pen held in the talker's mouth, the listener will adjust their perception accordingly.

Social information about a talker can also be extracted from their speech signal. Categories such as gender and sexuality can be inferred purely from the speech signal (Munson et al., 2006; Strand, 1999; Strand & Johnson, 1996). Using only auditory stimuli, listeners were more likely to perceive a token from a fricative continuum as /ʃ/ for a prototypical male sounding voice and /s/ for a prototypical female sounding voice (Strand, 1999). While gender might be a prominent social category, it is also possible to elicit shifts in perceptual boundaries between sound segments from more fine-grained social categories. For example, within gender categories, shifts in perceptual boundaries can be linked with how normatively masculine or feminine the talker is perceived to be or even the sexuality of the talker. Ambiguous male and female voices were perceived as having category boundaries between the prototypical male and female voices in Strand (1999). Using a binary forced choice word identification task in which participants were instructed to identify whether stimuli were words beginning with /s/ or /ʃ/, Munson et al. (2006) found that listeners perceived bisexual- or lesbian-sounding women as producing more words beginning with /s/ than their heterosexual-sounding counterparts. Overall, this means that information extracted from the speech signal is sufficient for a listener to obtain information related to the talker's attributes, idiosyncrasies or social groupings.

2.1.4 Accent

One particular type of talker-related information that is known to influence speech perception is accentedness. Notably, foreign or non-native accents have a negative impact on comprehensibility and listening effort (Floccia, Butler, Goslin, & Ellis, 2009; Kang & Rubin, 2009; Rubin & Smith, 1990; Van Engen & Peelle, 2014), delaying word identification and requiring more executive resources for successful comprehension. Listeners are generally good at identifying whether a voice has a non-native accent despite the difficulty of identifying what specific acoustic properties make an utterance sound more or less accented (Derwing & Munro, 2009. However, listeners are less accurate at identifying specific accents (Atagi & Bent, 2013; Derwing & Munro, 1997; McCullough & Clopper, 2016; Rubin, 1992; Vieru & Boula, 2011). This means that, even when listeners are unable to pinpoint a specific accent, their comprehension can be affected. This raises the question of whether non-native accents are perceived as a single entity or as different individual accents. In other words: do listeners treat all non-native speakers as the same group regardless of their accent?

There is some existing evidence that supports the notion that non-native accents are treated as broader entities, at least in terms of general comprehensibility and in certain circumstances. For example, the impact of accentedness on reported comprehensibility is observed even when the auditory signal is held constant but participants are shown photographs of faces that are identified as more or less foreign looking (Rubin, 1992). In an experiment where listeners were tested on their ability to recall material presented by talkers of different assumed ethnicities, Rubin (1992) paired different faces with a recording of a lecture. Listeners judged the recording to be less comprehensible when shown an accented speaker-associated face than when that same recording was paired with a native speaker-associated face even though the recording was made by a native speaker; conversely, comprehensibility of a recording made by a non-native speaker was ameliorated when paired with a native speaker-associated face. In that study, the accented speaker-associated face and native speaker-associated face were only represented by one person each.

In studies where listeners are exposed to multiple non-native accents, listeners were able to categorise non-native accents separately (Atagi & Bent, 2013; McCullough & Clopper, 2016). While this categorisation is not necessarily accurate, for example Mandarin native speakers and Korean native speakers were more likely to be grouped together than with the other non-native languages used in the study such as Spanish and Hindi (McCullough & Clopper, 2016), this is evidence that non-native accents are not perceived as a single entity when multiple non-native accents are present in a situation.

Most studies exploring how listeners perceive non-native accents have looked at the perceptions of words, semantic meaning, and memory. In terms of linguistic processing, non-native accents have been found to affect word identification (Clarke & Garrett, 2004; Floccia et al., 2009; Porretta et al., 2016), semantic integration (Lev-Ari & Keysar, 2012; Romero-Rivas, 2016), and even a listener’s memory of their own speech (Lev-Ari, Ho, & Keysar, 2018). Listeners exhibited slower reaction times in lexical decision tasks both when items are spoken by non-native talkers (Clarke & Garrett, 2004; Floccia et al., 2009) and when written items are primed by non-native stimuli (Poretta et al., 2016). This delay is found both with accents which tend to be more familiar, such as Spanish, and those that tend to be less familiar, such as Chinese (Clarke & Garrett, 2004). While they did not directly test and compare the magnitude of the delay between the different accents, perception of both accents were similar in the sense that listeners were able to reduce their reaction time in both accent conditions within a few utterances. However, within the same type of accent there is a gradient effect of accentedness on listener perception; reaction times during lexical access increase the more strongly accented the talker is (Poretta et al., 2016). In sum, while native accents are perceived differently from non-native accents, it is unclear how the type of non-native accent affects perception. However, the degree of accentedness within the same accent does have an effect.

Results from studies that test the effect of accentedness on detecting word changes in stories provide evidence that listeners have more difficulty in integrating semantic meaning because they tend to remember fewer lexical details in non-native utterances (Lev-Ari, 2015b, 2015a; Lev-Ari & Keysar, 2012). Besides recalling fewer lexical details in non-native speech, listeners are less able to accurately recall their own responses when interacting with a non-native talker (Lev-Ari et al., 2018). These findings are supported by studies which used neuroimaging methodologies to observe that listeners undergo higher cognitive perturbation when listening to non-native speech (Porretta, Tremblay, & Bolger, 2017; Romero-Rivas, 2016). Taken together, this body of work provides further evidence that there is a difference in how listeners perceive non-native speech compared to native speech.

Despite the seemingly negative effect non-native speech has on native listeners' perception, this perception can be improved if listeners have exposure to sufficient examples. The number of examples needed ranges from as few as two sentence-length utterances for visual probe matching tasks (Clarke & Garrett, 2004) to sixteen utterances for sentence transcription tasks (Bradlow & Bent, 2008). However, the rate of adaptation towards non-native speech does not improve linearly with the amount of exposure. Bradlow & Bent (2008) compared keyword recognition accuracy rates of non-native speech across different quartiles. They found that although there was a general improvement between scores in the first and fourth quartiles, those scores were not always the lowest and highest, respectively, for individual listeners. Further, there is evidence that the increase in accuracy observed occurs only to a certain extent. For example, Floccia et al. (2009) present evidence that initial perturbation in lexical decision reaction times can be reduced with exposure but that the shorter delay does not then improve with further exposure during the course of the experiment.

While there is ample work on the effects of non-native accents on the comprehension of entire utterances or words, there is little work exploring how non-native accents affect the perception of sounds. However, it is apparent from work examining other kinds of talker-related information, including those described in Section 2.1, that information attributed to a talker can shift the perceptual boundary between phones (Drager, 2011; Johnson et al., 1999; Kraljic & Samuel, 2006; Kraljic et al., 2007; Strand & Johnson, 1996).

Closely related to the current study is Sumner (2011), which also used a binary forced choice syllable identification task to determine whether listeners' category boundaries of non-native bilabial plosives are affected by the amount of variability the participant was exposed to in previous stimuli. In that study, listeners were exposed to stimuli consisting of individual English words produced by a French native speaker in an exposure phase before being presented with randomly ordered /ba/-/pa/ syllables produced by the same talker. Sumner (2011) found that the top-down training method with highly variable stimuli shifted the perceptual category boundary of the non-native accent, providing evidence that listeners integrated talker-related information about VOT that is inferred from the word stimuli. However, one crucial difference between the experiment in Sumner (2011) and those presented herein is that the non-native talker stimuli was not presented alongside native talker stimuli in Sumner's experiment since the goal of that study was to explore the effects of different types of non-native speech exposure. In contrast, the current study explores the extent to which a listener's perception of sounds is affected by whether or not the talker is a native speaker of English or not.

Specifically, my dissertation investigates the effect of talker nativeness at a sub-lexical level, examining whether talker nativeness affects a shift in the perceptual boundary between two phones. The phonetic realizations of the phones selected for this study – /b/ and /p/ – are known to differ across languages (Lisker & Abramson, 1964) and across native and non-native varieties of English (Arslan & Hansen, 1997).

2.2 Acoustic Properties of English Word-Initial Plosives

American English plosives have an aspiration distinction in word-initial position; voiceless plosives have a long lag voice onset time (VOT) and voiced plosives have a short lag VOT (Lisker & Baer, 1984), and the voicing boundary for bilabial plosives is reported to be around 25ms (Abramson & Lisker, 1973). In contrast, non-native speakers of English often have a different voicing boundary than native speakers. For example, Spanish native speakers produce plosives which have a crossover point of about 14ms (Abramson & Lisker, 1973); English plosives produced by native speakers of Spanish are more [b]-like (Flege & Eefting, 1988). In contrast, Mandarin speakers produce English voiceless plosives regardless of place with slightly longer VOT than native speakers (Chen, Chao, & Peng, 2007). Malaysian English speakers are often considered to fully voice initial voiced plosives and do not consistently aspirate voiceless ones, with a VOT as low as 7 ms for some tokens of word initial [p] (Yamaguchi & Pétursson, 2012). In other words, the VOT of a voiceless plosive produced by a non-native English speaker might have the same duration as the VOT of a voiced plosive produced by a native speaker. Even among speakers of languages with word initial aspiration distinctions, English learners might not have native-like VOT.

Listeners' perception of a plosive as voiced or not is closely linked with its VOT. For example, if listeners hear a native talker who produces a voiceless plosive with a VOT that is similar to its voiced counterpart, they are more willing to interpret it as the voiced plosive (McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2008;

McMurray, Clayards, Tanenhaus, & Aslin, 2008; McMurray, Tanenhaus, & Aslin, 2002). In an experiment in which listeners had to do a lexical decision task after being primed by a different word, words with altered VOT were less able to prime semantically related targets, especially when the voicing counterpart is an actual word (Andruski, Blumstein & Burton, 1994). That VOT affects the perception of voicing is further evidenced by eye-tracking studies, which demonstrate that when a /p/ onset word is produced with a lower than average VOT (making it closer to that of a /b/ onset word), listeners show an increase in gazes to images representing the voiced item (McMurray et al., 2002).

Despite the prominence of VOT as a marker for voicing in plosives, F0 of the following vowel has also been shown to influence voicing perception. Specifically, higher F0 is associated with voiceless plosives while lower F0 is associated with voiced plosives. While F0 only has an effect on the perception of plosives with *ambiguous* VOT in unspeeded tasks, Whalen, Abramson, Lisker, & Mody (1993) found that it even has an effect on plosives with *unambiguous* VOT in speeded tasks. F0 even supersedes VOT as a prominent cue for voicing when syllables are presented in masking noise or when the signal has undergone low pass filtering (Winn, Chatterjee, & Idsardi, 2013).

Chong (2018) provides some preliminary evidence that listeners' perception of VOT can be affected by whether a talker is a native or non-native speaker of English. For that study, I conducted a binary forced choice syllable identification task testing how /ba/-/pa/ syllables with varying VOT produced by American and Malaysian talkers were perceived. The results from that study provide evidence that listeners do indeed have a different perceptual boundary for plosives produced by a non-native talker compared to a native talker. The perceptual boundary for the non-native plosives was shallower compared to that for the native plosives, indicating that there is more variation in perceiving non-native plosives. However, in contrast to expectations, listeners were less likely to perceive non-native plosives as voiceless, even though the non-native talker has lower VOT for similar plosives in the frame sentence. There was an increase in voiceless responses over the course of the experiment, but there did not appear to be a linear increase across successive blocks.

While this preliminary study suggested that listeners can perceive plosives produced by native and non-native talkers differently, certain aspects of the experiment's design as well as the unexpected pattern of voiceless responses left the mechanism behind this difference in perception unclear. Specifically, the VOT ranges of the syllable continua used were not equal, with most tokens from the non-native talker occupying a VOT range lower than the generally cited voicing boundary for bilabial plosives of 25ms (Abramson & Lisker, 1973). This could potentially have primed listeners to expect fewer voiceless syllables from the non-native talker. Secondly, the critical syllables were constructed from stimuli produced by each talker, ie. frames were paired with syllables produced by the same talker. This means that there could be other acoustic features in the syllables which influenced the listeners' responses. Therefore, the experimental design was altered for the experiments presented in this dissertation in the following ways: all three experiments use a single /ba/-/pa/ continuum based on the /pa/ and /ba/ realizations of a single talker, and steps from the continuum were then spliced into frame sentences produced by two different talkers who differed in whether they were a native or a non-native speaker of English. This ensures a large and equal VOT range for the syllable continuum for both talkers and eliminates unforeseen idiosyncrasies arising from using different critical syllables.

2.3 The Hypotheses

Taken together, the work outlined in Sections 2.1-2.2 suggests that listeners' perception of sounds will be influenced by whether or not the talker is a native speaker of the language, and it raises questions such as: does a listener's perceptual boundary for non-native speech differ from that for native speech? Do listeners take into account acoustic cues for non-native talkers, or do listeners default to pre-existing representations instead? Is their perception based on the specific non-native accent they are exposed to during the experiment or do listeners use similar boundaries for non-native accents in general? In this dissertation, I present three different experiments that explore these hypotheses. Specifically, I hypothesized that:

(1) Listeners will have different perceptual boundaries between word initial /pa/ and /ba/ when listening to different talkers; listeners will be more likely to perceive a token with an ambiguous VOT as voiceless for talkers who produce a shorter VOT for similar sounds in the frame sentence.

(2) In the absence of prior experience with an accent and cues clearly indicating differing VOT, listeners will have similar perceptual boundaries for non-native accents in general, especially when they are not familiar with the accents and only one type of non-native accent is present.

These hypotheses will be tested in the Experiments reported in the following chapters.

Chapter 3 Experiment One

Experiment One is a binary forced choice identification task that uses a within subjects design. It was designed to test for an effect of a talker's accent as native or non-native on the perceptual boundary between /ba/ and /pa/. To avoid confounds that can arise when conditions involve different /pa/-/ba/ continua (c.f. Chong 2018), a single continuum produced by a single speaker was spliced into frame sentences that were produced by a native and a non-native talker. By controlling the continuum across condition - including controlling the talker who produced the continuum - I was able to ensure that any difference in responses across conditions stemmed from phonetic information inherent in the frame sentence. I hypothesized that there would be a difference in the listener's perceptual boundaries for the native and non-native talkers; listeners were expected to respond differently depending on whether the talker who produced the sentence was a native speaker of English or not. Additionally I hypothesized that exposure to the non-native talker's frame sentence would elicit more /pa/ responses overall than when that same participant was exposed to the native talker's frame sentence.

This chapter begins with a brief overview of the participants of the experiment (Section 3.1). The experiment materials and procedure is then explained in detail in Section 3.2. Section 3.3 analyzes the results of the experiment. This chapter is concluded by a discussion of the results in Section 3.4.

3.1 Participants

53 participants were recruited from the University of Hawaiʻi at Mānoa Linguistics Department Linguistics Beyond the Classroom program and via word of mouth, and they received partial class credit or a gift card for taking part. The analysis

primarily focuses on 25 participants (14 female, 11 male) who were monolingual native speakers of American English. All participants were at least 18 years of age at the time of the study and gave their written consent prior to taking part. After giving consent, they then participated in the experiment. After finishing the experiment portion, they underwent a short interview which recorded their comments on the experiment and their background information. Each participant took approximately 30 minutes to complete the study.

Based on responses during a short interview following the experiment, 25 participants were identified as belonging to a monolingual native speaker group. The criteria for inclusion were that participants had normal hearing, were native speakers of American English, acquired English as their first language, were children of native speakers, and spoke only English at home and with their parents. The data were restricted in this way because even limited exposure to a language early in childhood has an effect on a listener's perception of speech sounds (Section 2.1.1). All participants also grew up in neighborhoods where English was the main language used. After data from participants who did not meet these criteria were removed, responses from 25 participants (14 female, 11 male) were analysed. None of the participants whose data were analysed had studied a second language in school before the age of 6.

Table 3.1: Breakdown of monolingual native speaker participants according to gender, age, and their response when identifying the non-native talker's L1.

	Non-native talker L1 = Spanish	Non-native talker L1 = Non-Spanish	Age: min, median, max
Female	6	8	18, 22.5, 30
Male	2	9	19, 21, 35

3.2 Materials and Procedure

Stimuli were produced by one American English speaker and one Spanish speaker from Mexico, both of whom were women in their 20s. The Spanish speaker was born and raised in Mexico, but had been living in Hawaiʻi for 5 years at the time of recording. While fluent in English, she has a noticeable non-native accent, being frequently referred to as the "accented" talker by participants during post-experiment interviews. Both talkers were selected because they have a similar voice quality and pitch range, making it possible to splice a single item (i.e., a syllable resynthesized using the native talker's voice) into each of their frame sentences without it sounding unnatural. The native talker had an average F0 of 174Hz (1.72 Bark) for her frame sentence, while the non-native talker had an F0 of 165Hz (1.63 Bark), with a difference of less than .1 Bark between the mean F0 across the two talkers' frame sentences. During post-experiment interviews, no participant reported noticing the splice and syllable being from the same talker prior to being asked specifically about it. After being asked explicitly whether they noticed syllables from the same talker had been spliced into the frame sentence, 6 responded affirmatively. An analysis which compares responses from these 6 participants to the other 19 participants is presented in Section 3.3.1.

Each of the two talkers were recorded reading the frame sentence *Please listen to the syllable __, and pick what you heard.* The talkers were instructed to read in a clear but normal talking rate. The frame sentence was designed to contain word initial /p/ through the inclusion of the words *please* and *pick*. This was done so that listeners could have some level of expectation regarding each talker's word initial /p/; that way, I can make a comparison of responses with those from Experiment Two (Chapter 4) in which the VOT is controlled across talkers. This comparison will allow me to test whether any difference in responses is a result of listeners using the cues present in the talkers' frame sentences or is caused by an accentedness effect. As listeners are said to allocate less attention to and take less information from a non-native talker's linguistic cues (Section 2.1.4), this will shed some light on whether this claim holds true for phonetic cues. In the case of the non-native talker, the VOT of *please* was 39ms and the VOT of *pick* was 14ms. The VOT of both /p/-initial words produced by the native talker were longer than those produced by the non-native talker, at 81ms (*please*) and 47ms (*pick*). Consequently, the non-native talker in the present study produced a voiceless plosive with a VOT around the typical categorical boundary for bilabial plosives spoken by American English speakers (Section 2.2). Other than VOT, the frame sentence was designed so that none of the words contained the vowel that was in the target syllable. This was done to reduce the likelihood that listeners would detect a potential mismatch between the spliced syllables and the frame sentence.

Items from both talkers were matched for intensity at 60dB. The native talker spoke at a speech rate of 3.8 words per second while the non-native talker was slightly slower at 3.3 words per second. The effect this difference in speech rate may have had on responses is discussed in Section 3.4. After the syllables were spliced into the frame sentences, the complete stimuli were filtered to remove some high frequency background noise. The same filter settings were used for both talkers.

In addition to recording the frame sentence, the native talker also recorded the natural /ba/ and /pa/ syllables used to create the /ba/-/pa/ continuum. The critical stimuli were constructed based on the method outlined on McMurray and Aslin's website[4], using Praat (Boersma & Weenink, 2019). This method was selected as it

4 http://www2.psychology.uiowa.edu/faculty/mcmurray/publications/mcmurray_aslin_supplement/ (accessed November/1/2018)

retained the burst portion of /pa/, while containing the vowel qualities of /ba/. As the F0 of a following vowel is a possible marker for voicing, merely removing portions from /pa/ might result in syllables which are still biased towards /pa/. Additionally McMurray and Aslin's method results in syllables which do not vary in duration across steps. Figure 3.1, Figure 3.2, and Figure 3.3 below illustrate the stimuli construction process using the 40ms step as an example.

First, a segment corresponding to a VOT step (i.e., at increments of 5ms) was selected from the onset of the natural /pa/ syllable (Figure 3.1). This segment was then pasted onto the onset of the natural /ba/ sample (Figure 3.2). A segment with roughly the same duration was then removed from /ba/ right after the pasted segment (Figure 3.3). Thus, the VOT portion was taken from /pa/ while the rest of the syllable came from /ba/. As the segment removed from the /ba/ sample includes part of the vowel, it was necessary to remove segments from the nearest zero crossing to reduce the amount of artifacts arising from the splice. This in turn resulted in slight discrepancies of less than 1ms in the duration of the segment removed. This process was repeated for all the steps outlined in Table 3.2 below. Note that none of the steps of the continuum were created using the vowel from /pa/.

As a result of the resynthesis, there were 10 steps of the continuum. These steps range from a VOT of 5ms (most /b/-like) to 50ms (most /p/-like), with 5ms intervals, as shown in Table 3.2. The 5 – 50ms continuum range was chosen because the endpoints are roughly equidistant from the commonly cited perceptual boundary of bilabial plosives, 25ms (Abramson & Lisker, 1973). The 5ms step interval was chosen as it was large enough to not produce too many steps in the continuum while still being small enough to reveal the perceptual boundary between /pa/ and /ba/.

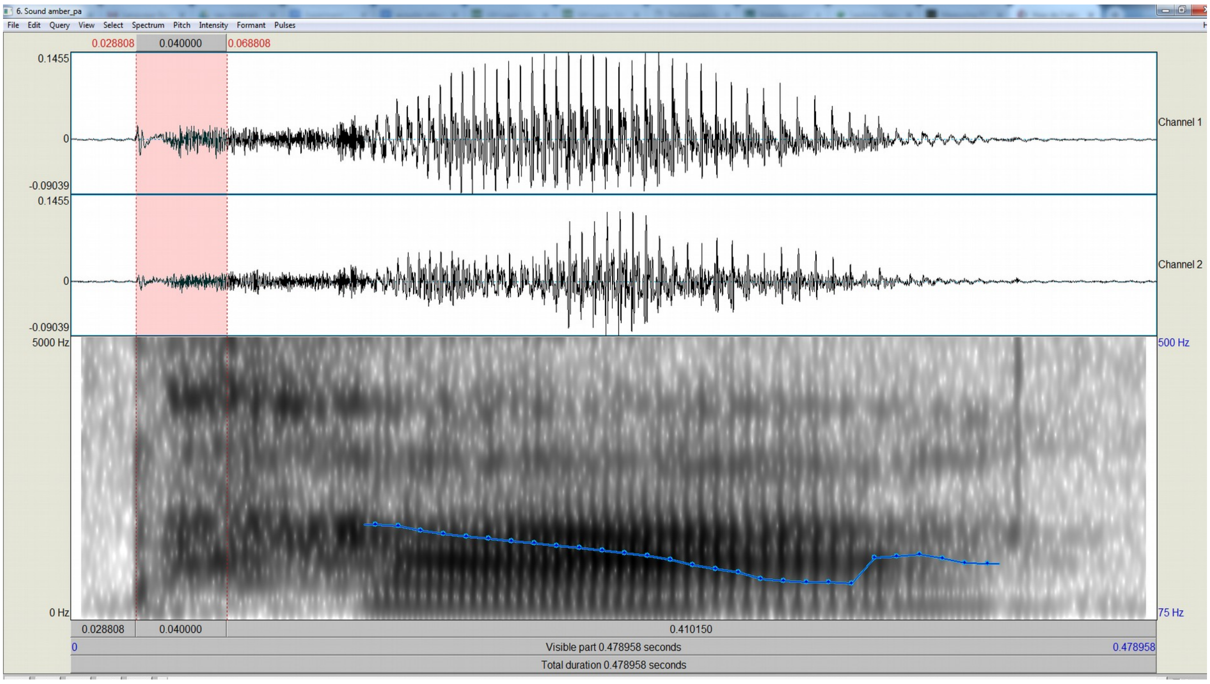

Figure 3.1: Spectrogram and waveform from Praat showing the first phase of the resynthesis process in which a 40ms segment (i.e., step 8) was taken from the onset of the /pa/ sample.

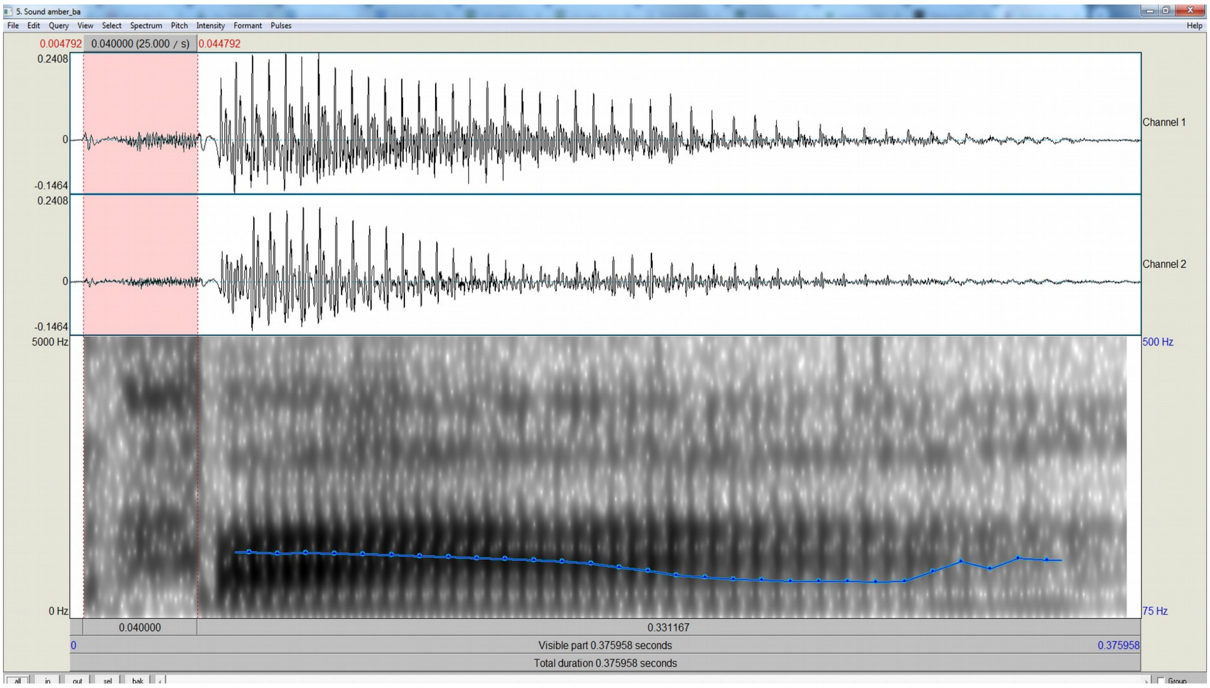

Figure 3.2: Spectrogram and waveform from Praat showing the second phase of the resynthesis process for step 8 (40ms) in which the selected segment was pasted onto the onset of the /ba/ sample.

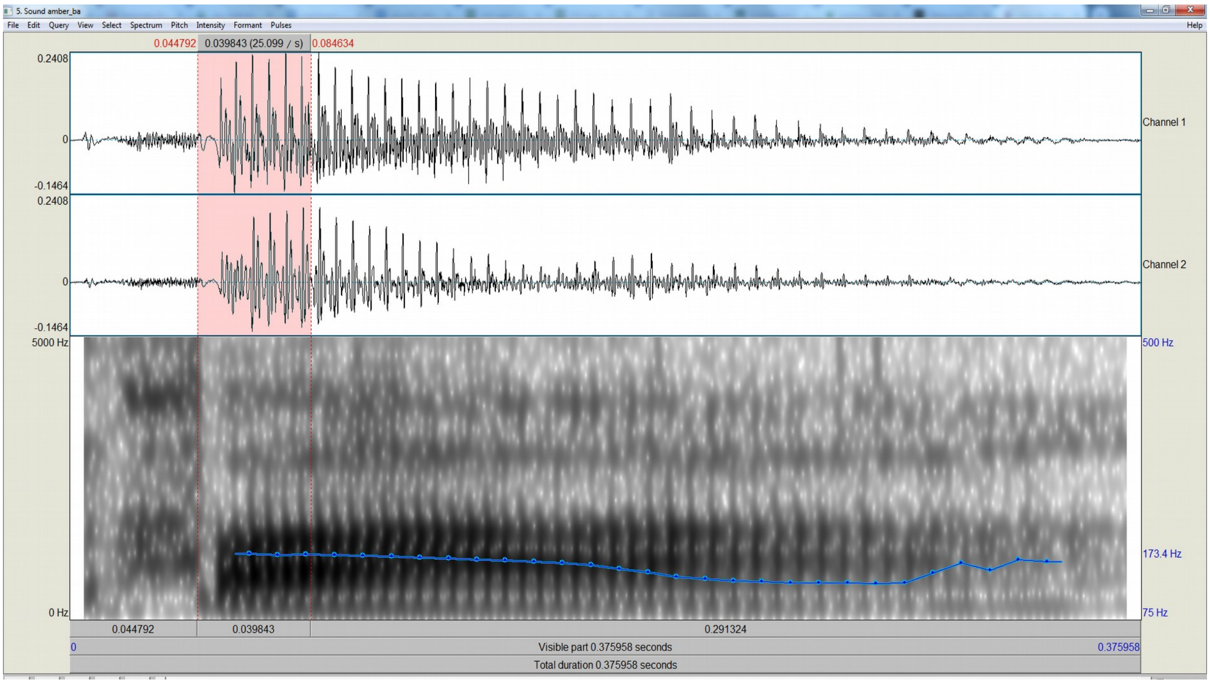

Figure 3.3: Spectrogram and waveform from Praat showing the third phase of the resynthesis process for step 8 (40ms) in which the segment with roughly the same duration as the inserted segment was selected and removed. Duration was controlled to the extent possible while restricting splicing to zero crossover points to maintain naturalness.

Table 3.2: VOT (shown in ms) of critical syllable steps.

Step	1	2	3	4	5	6	7	8	9	10
VOT (ms)	5	10	15	20	25	30	35	40	45	50

The 10 continuum steps were spliced into the two frame sentences, creating a total of 20 stimuli. All steps were spliced into the same spot for each talker’s frame sentence, with approximately 450ms of silence before and after the critical syllables. This amount of silence was included to create a natural sounding intonation which emphasized the critical syllables. This process is illustrated in Figure 3 .4.

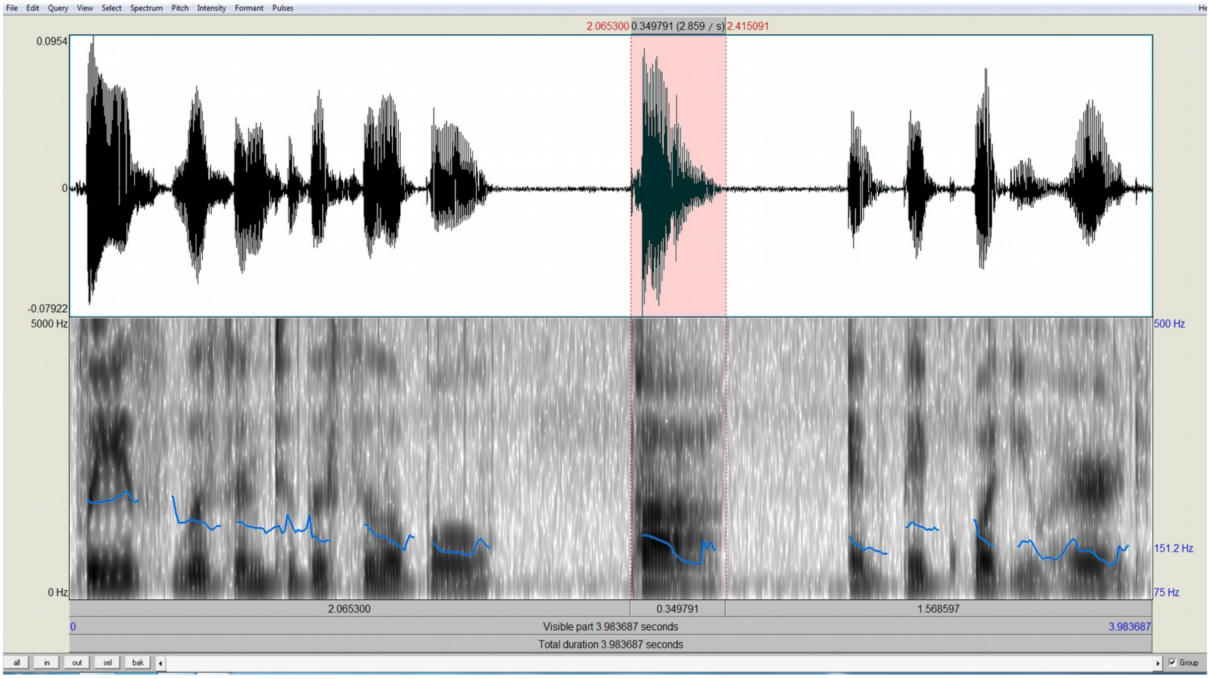

Figure 3.4: Spectrogram and waveform from Praat showing the splicing process in which the critical syllable was spliced into the frame sentence with approximately 450ms of silence before and after the syllable.

After splicing, the stimuli were converted to mono channel but presented binaurally. This was to ensure a consistent signal from both sides of the headphones and to reduce the likelihood of confounds arising from an imbalanced signal. They were then organised into 8 critical blocks, with no fillers. Fillers were not used because post-experiment feedback from participants who took part in Chong's (2018) experiment suggested that the inclusion of fillers made the experiment too long and tedious for the participants. Similar to previous studies on the perception of non-native accents (Magen, 1998; Sumner, 2011), fillers were not included so that the experiment would not last too long and cause fatigue.

While item progression was automatic within blocks, participants had to click through to start the next block. This was done so that they could take small breaks if needed. For each talker, each step in the critical continuum appears only once per block, meaning there was a total of 20 items per block. The order of items in each block was fully randomised. Prior to starting the main experiment, participants responded to two practice items containing /ra/ and /la/ which were produced by a different non-native talker; this was done in order to familiarise participants with the task.

The experiment took place in the General Lab, one of the Language Analysis and Experimentation labs at the University of Hawaiʻi at Mānoa. During the experiment, participants listened to the recordings through a pair of Sennheiser HD202 headphones and were asked to select keys on a regular computer keyboard corresponding to whichever syllable they heard. The *F* key was used for the /ba/ choice while the *J* key was used for the /pa/ choice. These keys were chosen because they are adequately spaced apart as shown in Figure 3.5, and are less likely to be mistakenly selected. The key choices were consistent for the whole experiment for all participants. Participants were not specifically instructed as to how they should press the keys as reaction time is not taken into consideration in the analysis. Although reaction times were not a factor, participants were instructed to answer as quickly as possible and not overthink their responses. The experiment took 15 – 20 minutes to complete.

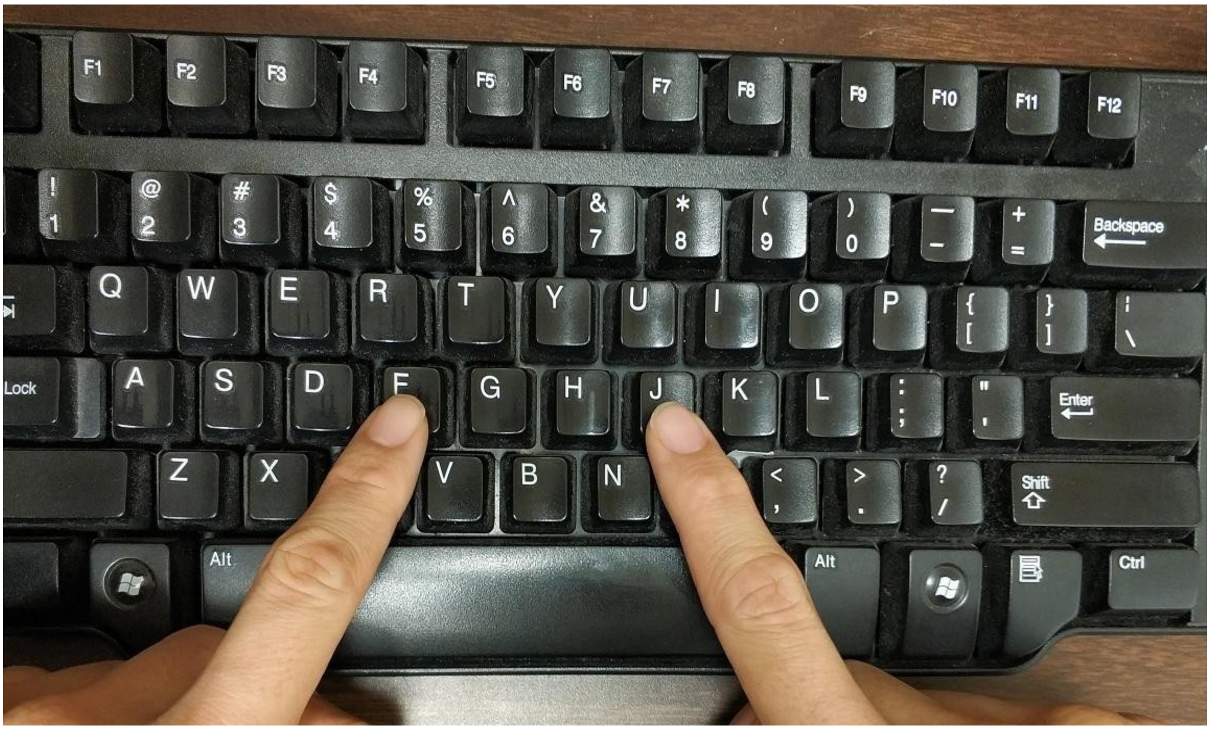

Figure 3.5: Picture showing the keys used to log responses to the audio stimuli. Participants pressed F if they perceive /ba/ and J if they perceive /pa/.

After completing the main task, participants underwent a short post-experiment interview (Appendix A). In addition to providing feedback and language background information, they were asked to identify talker attributes such as the talkers' likely native language, place of origin, and age. The feedback portion was designed to determine whether participants were aware of the talkers' accents and to check how well they were able to identify the accents accurately. In addition, participants were told at the end that syllables from one talker was spliced into both frame sentences and they asked if they noticed the syllables were produced by the same talker. Information obtained from the language background information portion was used to filter participants for analysis based on their language exposure during early childhood.

3.3 Results

Figure 3.6 presents the overall proportion of /pa/ responses to the 10 continuum steps, averaged across all blocks in the experiment. Error bars representing 95% confidence intervals were included to show the amount of variation between the participants' mean responses for the whole experiment. As is evident in Figure 3.6, the VOT range where talker nativeness affects the perception of /pa/ and /ba/ appears to be 15 – 25ms; there was less inter-talker difference for VOTs outside this range, with people categorically responding either /pa/ or /ba/. This is expected since 25ms is commonly cited as the perceptual boundary in bilabial plosives in previous literature (Abramson & Lisker, 1973) and any variation in the participants' responses is likely to occur around this part of the continuum.

Overall, the slope of the non-native talker's perceptual boundary appears to be as steep as that of the native talker, with a larger area under the non-native talker's line. The crossover point[5] (i.e., where listeners perceive /ba/ versus /pa/) of the non-native talker appears to be when the VOT is between 15 – 20ms, while that of the native talker is between 20 – 25ms. In fact, the 95% confidence intervals for both talkers do not overlap for the 15ms and 20ms steps, and marginally overlap for the 25ms step, providing evidence that listeners perceive the 15 – 25ms step syllables differently when they are placed in the differently accented frame sentences. The range of the 95% confidence intervals of the steps in the 20 – 30ms VOT range are much larger compared to the other steps, indicating that there is much more variation in responses to those stimuli. The range of the 95% confidence interval at the 20ms step is similar across talkers, which indicates that the 20ms step elicited similar

5 The 50% proportion mark is treated as the crossover point.

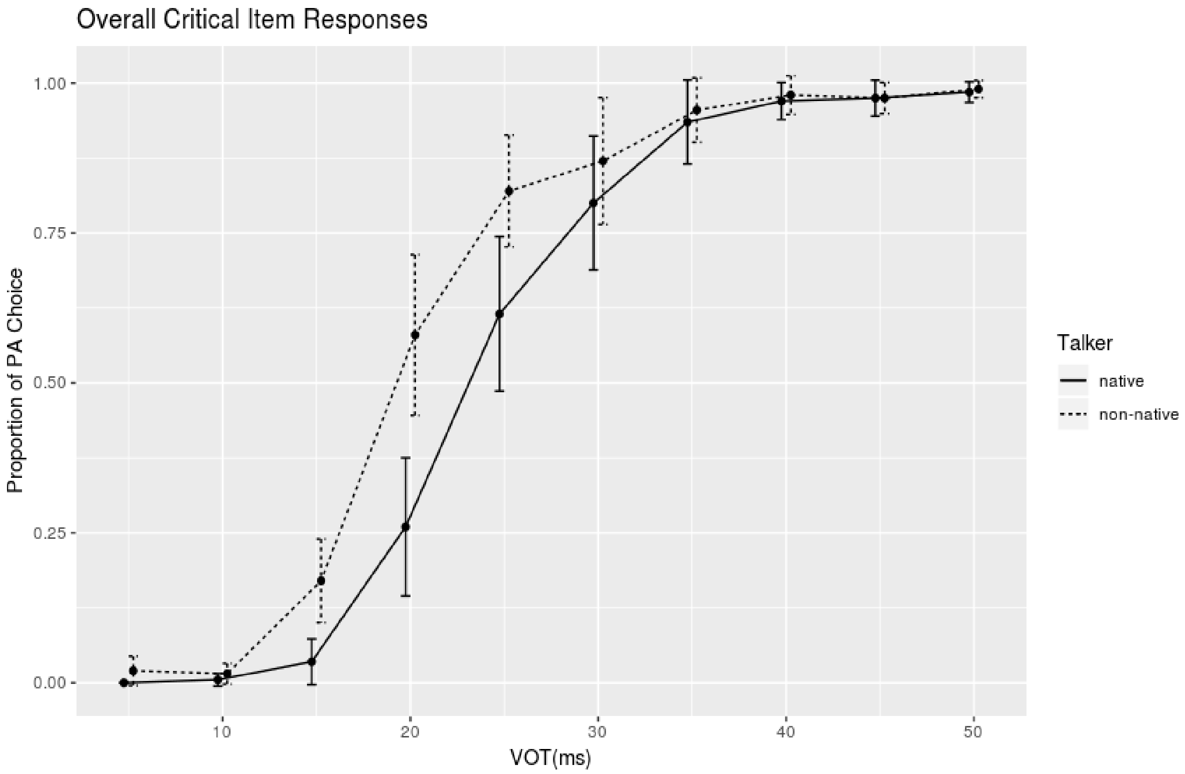

Figure 3.6: Overall proportion of /pa/ responses for critical items in Experiment One, averaged across participants and blocks and shown separately for the talker who is a native speaker of English (native) and the one who is not (non-native). Voice onset time (VOT) from the relevant step of each continuum is shown in milliseconds (ms) on the x-axis. 95% confidence intervals for each participant's mean responses across blocks are shown.

amounts of variation in participants' responses regardless of which frame sentence the step is spliced in.

Figure 3.7 presents the proportion of /pa/ responses for both talkers according to block. There does not appear to be a noticeable pattern of change in /pa/ responses across block; across all blocks, listeners' perceptual boundary between /pa/ and /ba/ has a smaller VOT when the frame sentence is produced by the non-native talker compared to when it is produced by the native talker. The shape of the slope for both talkers also does not appear to change in relation to each other.

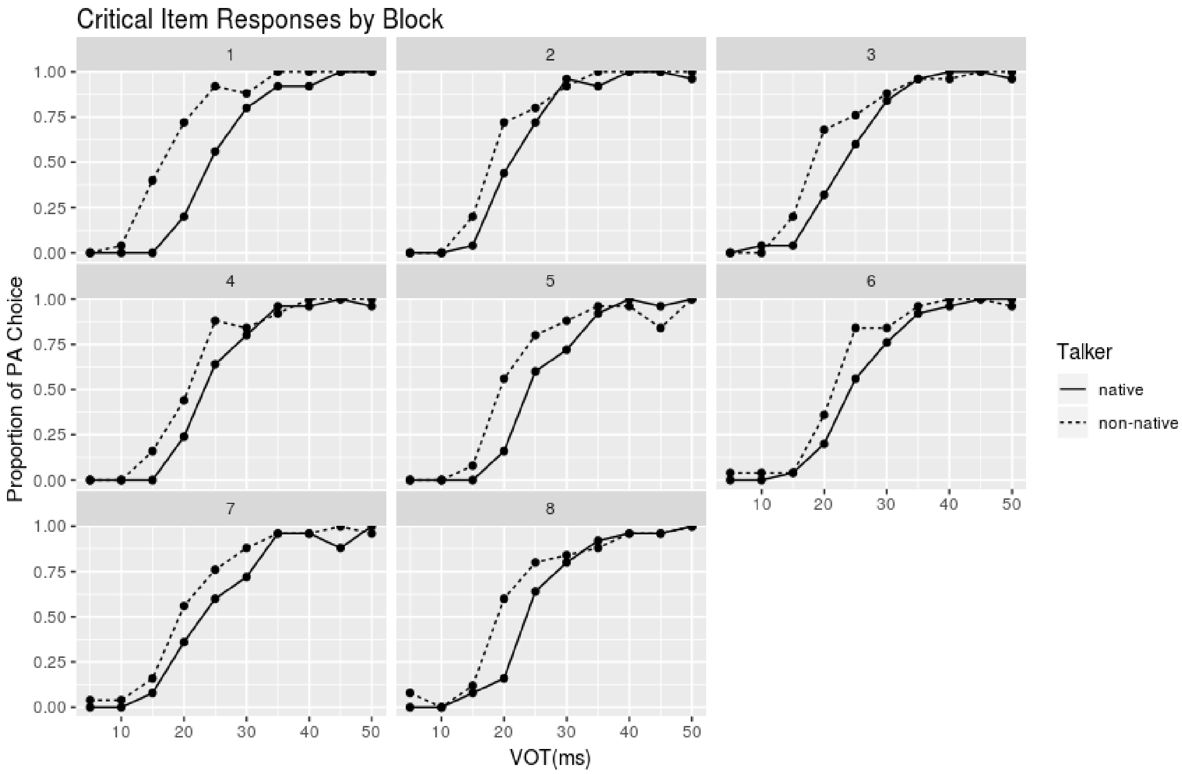

Figure 3.7: Proportion of /pa/ responses for critical items in Experiment One across the eight experimental blocks, averaged across participants.

Figure 3.8 shows each participant's overall responses averaged across all the blocks. In general, participants perceived more items as /pa/ for the non-native talker, except for participant 43, who appeared to perceive both talkers similarly. There is no information from the interview or feedback responses for participant 43 that can provide an explanation for their divergent behavior; any explanations would be mere speculation.

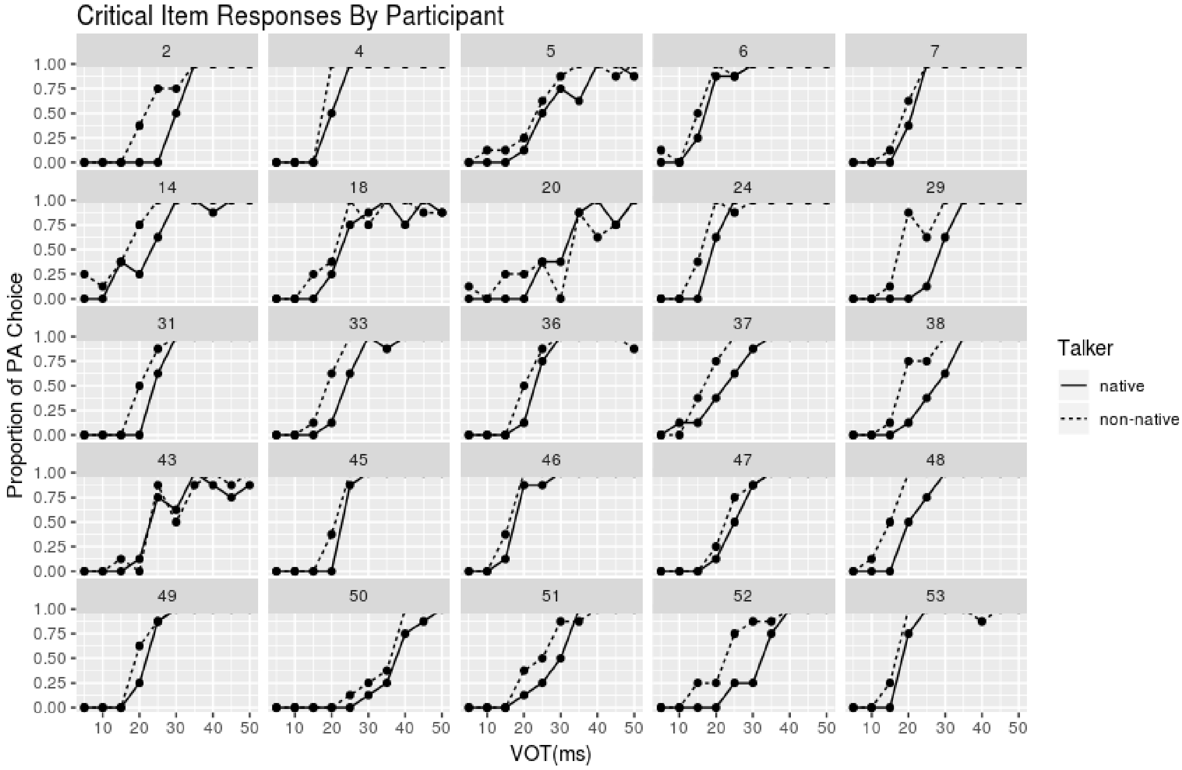

Figure 3.8: Proportion of /pa/ responses for critical items in Experiment One across participants, averaged across blocks.

To determine whether the difference in responses across condition was significant, a logistic mixed effects regression was fit to the responses to the critical syllables. The most parsimonious model which converged is reported here[6]. It included fixed effects of talker (native and non-native), VOT step (10 steps), and block (8 blocks), two-way interactions of talker and VOT step, talker and block, VOT step and block, and a by-participant random intercept. Although this experiment does not test the time course of shifts in perceptual boundary, block is included as a fixed effect in order to test the consistency of the participants' responses across the

6 The model reported in Table 3.3 was determined through a series of ANOVA model comparisons to be as well-fit as a more complex model containing a non-significant three-way interaction of Talker, VOT step and block, and a non-significant two-way interaction of VOT step and block.

experiment and for each talker. All predictors are centered. The reference level for talker type is the native English speaking talker, while VOT step and block are numerically ordered. The output of this model is reported in Table 3.3. Not included in the model reported here are the hypothesized fixed effects of response to the previous item and VOT step of the previous item. To test whether responses were influenced by the immediately previous item, the reported model was modified with the addition of fixed effects of response to previous item and VOT step of previous item. However, this modified model did not converge. When tested in a simpler model that converged, these factors did not reach significance and did not improve model fit, as tested using an ANOVA comparison of the models.

All fixed effects as well as the interactions between talker type and block and VOT step and block were revealed to be significant. Consistent with previous work (e.g., Lisker & Baer, 1984), the results indicate that there were fewer /pa/ responses when VOT was shorter ($p<.0001$); listeners were more likely to respond ‘ba’ the more /ba/-like the VOT. Block reaches significance as a simple main effect ($p<.0001$), indicating that there were more /pa/ responses toward the beginning of the experiment, with the number of /pa/ responses decreasing over the course of the experiment. The model also confirmed the graphical observation that there were significantly more /pa/ responses for the non-native talker than the native talker ($p<.0001$). The significant interaction of talker and block ($p<.01$) indicates that there is a change in the difference between responses to each talker across the blocks, with fewer /pa/ responses to the non-native talker as the experiment went on, indicating that listeners might be less affected by talker accent as the experiment progressed. The significant VOT step and block interaction ($p<.01$) indicates a significant difference in responses to different VOT across the blocks, with participants less likely to respond /pa/ to certain VOT steps as the experiment progressed.

Table 3.3: Output of logistic mixed effects model on responses to both talkers' VOT steps in Experiment One. Model: glmer(Critical Responses ~ Talker + VOT Step + Block + Talker:Block + VOT Step:Block + (1|Participant), data = Native and Non-native Talkers, family = binomial).

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.5351	0.2857	5.373	< .0001
Talker	1.2534	0.1325	9.458	< .0001
VOT Step	0.3094	0.0111	27.993	< .0001
Block	-0.1527	0.0331	-4.615	< .0001
Talker:Block	-0.1398	0.0564	-2.477	0.0133
VOT Step:Block	-0.0114	0.0037	-3.031	0.0024

The trends reported in Table 3 .3 confirm the hypotheses that the listeners differ in their perceptual boundaries for the native and non-native talkers and syllables spliced into the non-native talker's frame sentence elicit more /pa/ responses. However, questions remain: Does a participant's ability to notice the splice affect their sensitivity to the prime? Does correct identification of the non-native talker's native language affect a participant's perceptual boundary for the non-native talker?

3.3.1 Participants who Reported Noticing Splicing in Stimuli

Of the 25 participants who were analysed, 6 reported that they noticed that syllables from both sets of talkers were produced by the same talker. While they did not claim to notice the splice until after they were prompted, it is still possible that these participants behaved differently on the task. While the number of participants who responded in this way is too small for a "noticing factor" to be included in the model reported in Table 3 .3, a possible trend is explored here to determine if the frame sentences still have an effect on these participants' perceptual boundaries.

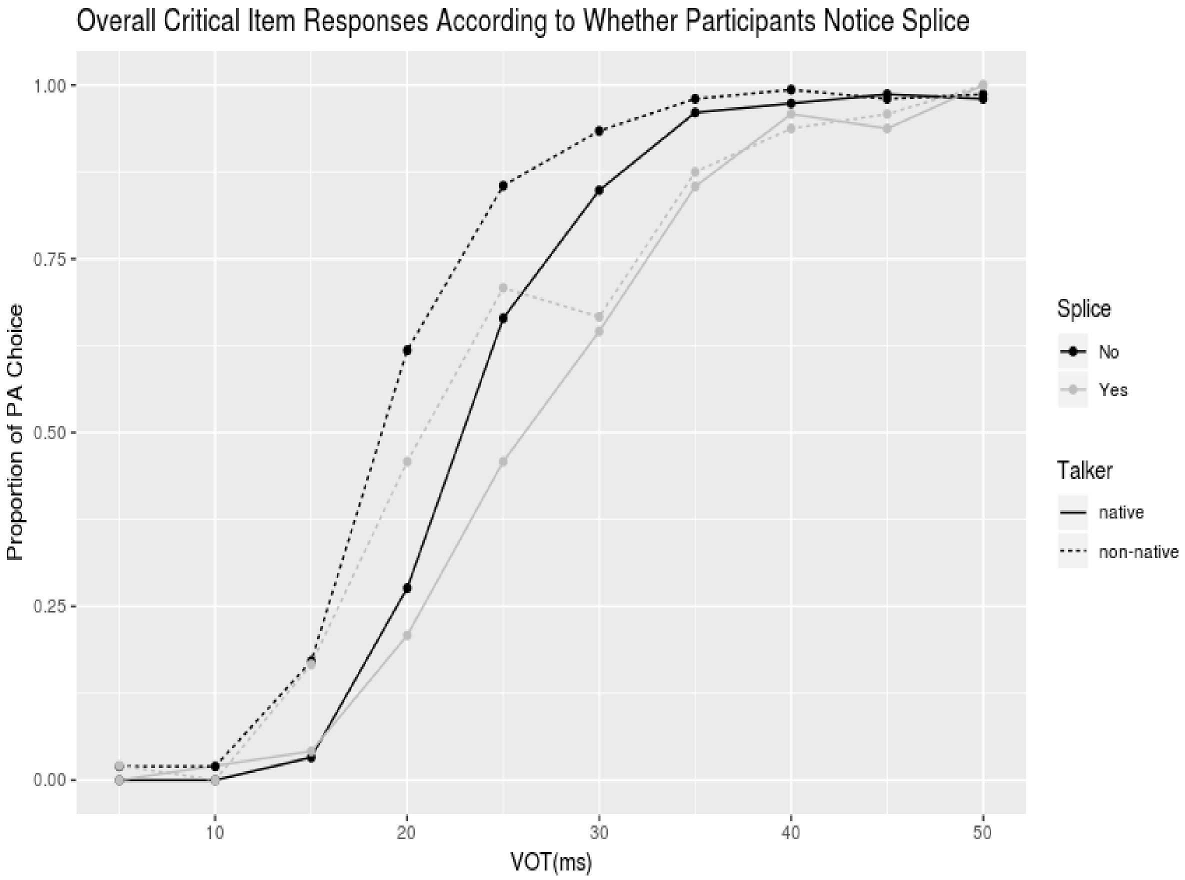

Figure 3.9: Proportion of /pa/ responses for critical items in Experiment One, averaged across participants and blocks. Proportions are shown separately for participants who reported noticing the splice after being asked explicitly (grey, N=6) and those who reported that they did not notice the splice (black, N=19), and for whether the talker was a native speaker of English (solid) or not (dotted).

Figure 3.9 compares the responses of these participants with the responses from the participants who reported being unaware of the splice. Participants who said they noticed the splice are referred to as the "Yes" group, while those who did not are referred to as the "No" group. No additional statistical models were fitted as the sample size was too small.

As shown in Figure 3.9, participants who noticed the splice had fewer /pa/ responses overall than those who did not, as both 'Yes' lines were lower than the corresponding 'No' lines. However, both groups of participants responded with a higher proportion of /pa/ responses when listening to the non-native talker. For participants who noticed the splice, their responses to both talkers are very similar for steps above 30ms. Despite their reported awareness of the splice, the frame sentences still elicited different responses for the 15 – 25ms steps. Given the small sample size, it is possible that the differences between the groups is merely noise, but future research testing explicitly the possibility of an awareness-based difference in responses would be worthwhile.

3.3.2 Participants who Perceived Non-native Talker as an L1 Spanish Speaker

When queried about the non-native talker's native language, 8 of the participants responded with Spanish, while the remaining 17 answered with non-Spanish languages. The ability of the participants to recognise the non-native talker's native language might be indicative of those participants' more extensive experience with this particular non-native accent, which is in line with previous studies which found that listeners who have more experience with a regional dialect are able to more accurately identify it (Clopper & Pisoni, 2004, 2006). This experience might then in turn affect their perception of the talker (Section 2.1.1). The possibility of an effect is explored here to determine if the perceptual boundaries of the group who were able to to correctly identify the non-native talker's native language differs from that of the group who were unable to do so.

As shown in Figure 3.10, both groups of participants appear to trend similarly in terms of their responses to each talker. Responses to both talkers are very similar across groups, especially for the 15ms and 20ms steps. The group which were able to correctly identify the non-native talker's native language appears to have a larger distance between their perceptual boundaries for both talkers. Their crossover point for the non-native talker is at a shorter VOT step while their crossover point for the native talker is at a longer VOT step.

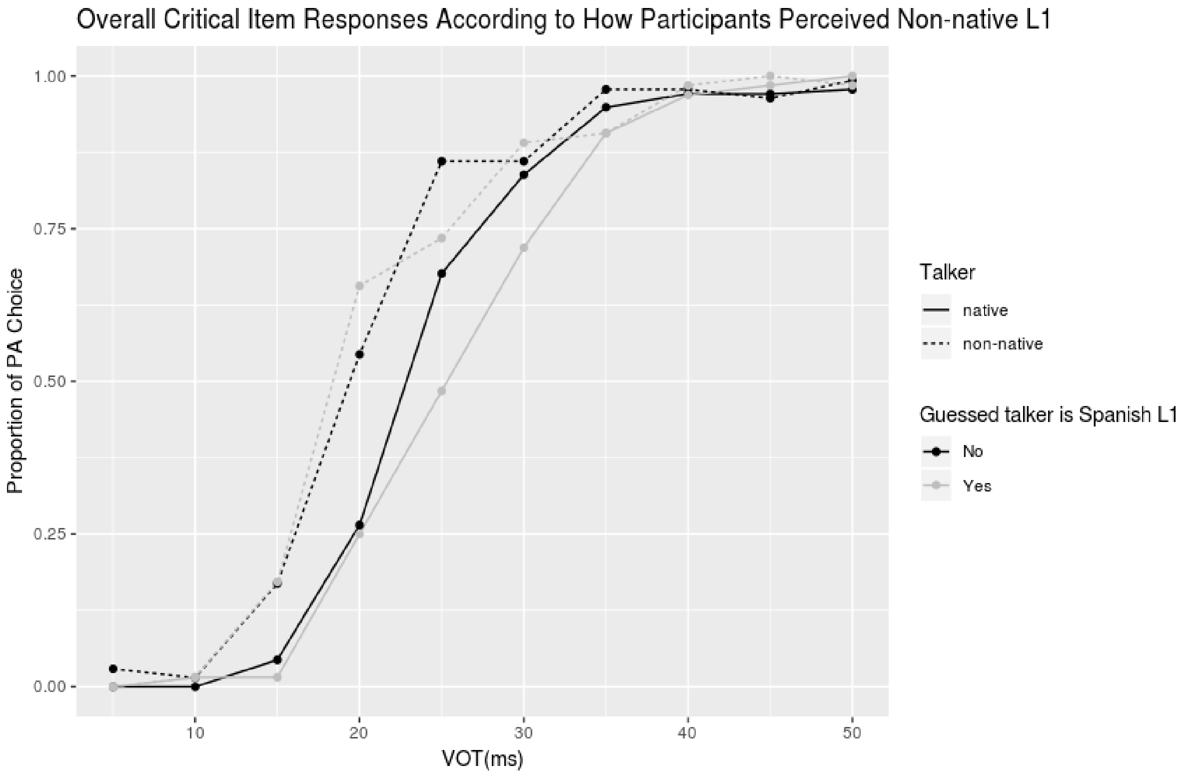

Figure 3.10: Proportion of /pa/ responses for critical items in Experiment One, averaged across participants and blocks. Proportions are shown separately for participants who responded perceiving the non-native talker as having Spanish as L1 (grey, N=8) and those who responded that the non-native talker has an L1 other than Spanish (black, N=17), and for whether the talker was a native speaker of English (solid) or not (dotted).

3.4 Discussion

The results from Experiment One provide evidence that listeners have different perceptual boundaries for the native and non-native talkers; listeners responded differently to the two talkers. Specifically, the results indicate that listeners are more likely to hear /pa/ when listening to the non-native talker. These findings are consistent with the prediction (Section 2.3) that listeners will have different perceptual boundaries between word initial bilabial plosives when listening to talkers with differing accentedness conveyed through the talkers' frame sentences and they will be more likely to perceive an item with an ambiguous VOT as voiceless for talkers who produce frame sentence cues with shorter VOT. The difference in responses across condition is especially apparent for the steps of the continuum with a VOT that lies between 15ms and 25ms. Based on the 95% confidence intervals, responses to these steps were also highly variable for both talkers. This indicates that listeners resort to talker-related information when perceiving acoustically ambiguous steps. Further, this was true even for listeners who reported noticing that syllables were produced by the same talker.

While the higher amount of /pa/ responses overall for the non-native talker supports the notion that listeners' perception of /pa/ and /ba/ was influenced by the talkers' accent, it is unclear whether the effect is due to the difference between the talkers' VOT values in the frame sentence words *please* and *pick*, or whether the listeners were influenced by their expectations stemming from the talkers' accentedness more generally. For example, listeners may have encountered L2 accented speech that is similar to the non-native talker and therefore expect this talker to produce word initial voiceless plosives with a shorter VOT than that which they expect from a native talker. As a consequence, these expectations may have influenced their perception of /p/ during the experiment. Alternatively, they may have been based entirely on exposure to /p/ in the frame sentences.

Another potential source of an effect are differences in pitch and speech rate. As discussed in Section 2.2, the pitch of the following vowel is correlated with the perceived voicing of a plosive. In this experiment, the role of the pitch in the syllable continuum is controlled by splicing the same continuum in both talker frame sentences. While there was a difference in speech rate between the talkers with the non-native talker's speech rate being 0.5 words per second slower, this difference also does not appear to be able to account for the difference in effect. Slower speech rates are generally associated with perceptual boundaries at a longer VOT value (Miller, O'Rourke, & Volaitis, 1997). However, the direction of the effect of speech rate is the opposite of that seen here. It is possible that the mismatch between the non-native talker's short VOT and low speech rate emphasised the talker's lack of aspiration, leading to listeners having a lower perceptual boundary for her. The relationship between speech rate and response patterns will be discussed in more detail in conjunction with Experiments Two and Three (Section 5.5).

Altogether, the results from Experiment One provide evidence that there is an effect of whether a talker is a native or non-native speaker on a listener's perceptual boundary between /ba/ and /pa/. However, it is unclear whether the result stems from listeners using VOT cues available in the frame sentences or from an accentedness effect. Therefore, Experiment Two aims to investigate whether there is a difference in perceptual boundary across conditions when the VOT of /p/ in the words *please* and *pick* from the frame sentences is held constant across the talkers.

Chapter 4 Experiment Two

The results from Experiment One demonstrate that listeners have different perceptual boundaries for syllables in the native and non-native talker conditions. In Section 3.4, I proposed two possible explanations for the difference between conditions: 1) listeners construct a perceptual boundary based on VOT cues available in the frame sentences 2) the perceptual boundary is caused by an accentedness effect. Experiment Two was designed to test which of these explanations is the most accurate.

Like Experiment One, Experiment Two is a binary forced choice identification task that uses a within subjects design. However, the frame sentences used in Experiment Two are modified from those used in Experiment One. Specifically, the VOT in *pick* and *please* across the two talkers are controlled so that the voice onset times in the native talker's two word initial plosives have the same duration as those of the non-native talker. Controlling the VOT across talkers was done to test if the difference in perceptual boundaries seen in Experiment One is still observed. I hypothesized that a difference in talker accent alone is enough to elicit a shift in perceptual boundary, even when the talkers are associated with plosives with similar VOT durations. This hypothesis would be supported if a difference in perceptual boundaries is observed despite both talkers having similar VOT in their frame sentence. Conversely, if listeners merely incorporate the cues present in the frame sentence, they should have similar perceptual boundaries for both talkers in Experiment Two.

This chapter first describes the subjects who participated in the study (Section 4.1), and then explains how the materials were modified from those used in

Experiment One (Section 4.2). The results are then presented in Section 4.3. Section 4.4 compares the results of Experiments One and Two. Chapter 4 is then concluded by a discussion of the findings (Section 4.5).

4.1 Participants

Similar to Experiment One, participants were recruited through the University of Hawaiʻi at Mānoa Linguistics Beyond the Classroom program and via word of mouth, and they received partial class credit or a gift card for taking part. 51 participants who did not take part in Experiment One took part in the present experiment. The analysis was conducted on responses from monolingual native speakers of American English who fit the criteria described in Experiment One (Section 3.1). Data from participants who did not meet these criteria were removed. In addition, there were 3 participants who fit the criteria but also mentioned noticing that the syllables were produced by the same talker without being prompted. These participants were also excluded. After these participants were excluded, responses from 25 participants (14 female, 11 male) were analysed. All participants underwent experimental procedures identical to the participants of Experiment One, including giving written consent prior to the experiment and undergoing the post-experiment interview. A breakdown of background of the monolingual native speaker participants is provided in Table 4.1.

Table 4.1: Breakdown of monolingual native speaker participants according to gender, age, and their response for the non-native talker's L1.

	Non-native talker L1 = Spanish	Non-native talker L1 = Non-Spanish	Age: min, median, max
Female	3	11	19, 20, 23
Male	5	6	18, 20, 30

4.2 Materials and Procedure

The materials were adapted from those used in Experiment One, specifically the native talker's frame sentence was modified using Praat (Boersma & Weenink, 2019) so that the VOTs for *please* and *pick* were shortened to match those of the non-native talker's VOT. Thus, the VOT in the frame sentences for both talkers in Experiment Two was 39ms for *please* and 14ms for *pick*. The VOT of the native talker's tokens was modified rather than the non-native talker's tokens because reducing VOT resulted in more natural sounding tokens than increasing VOT. In particular the non-native talker's token of *pick* has very little aspiration to draw from and splice. Other than that, the procedure for Experiment Two was identical to that of Experiment One. When prompted, 9 of the 25 participants said they noticed the syllables were produced by the same talker and had been spliced into the frame sentence. Additional analysis on these participants were run and presented in Section 4.3.1.

The native talker's VOT was modified by removing a segment from the end of the VOT portion so that the initial burst of the plosive is not affected. This is illustrated in Figure 4.1, which shows how the native talker's please was modified. For this particular example, 42ms of the end of the VOT portion was selected and deleted so that the native talker's please has a similar VOT to the non-native talker's (39ms).

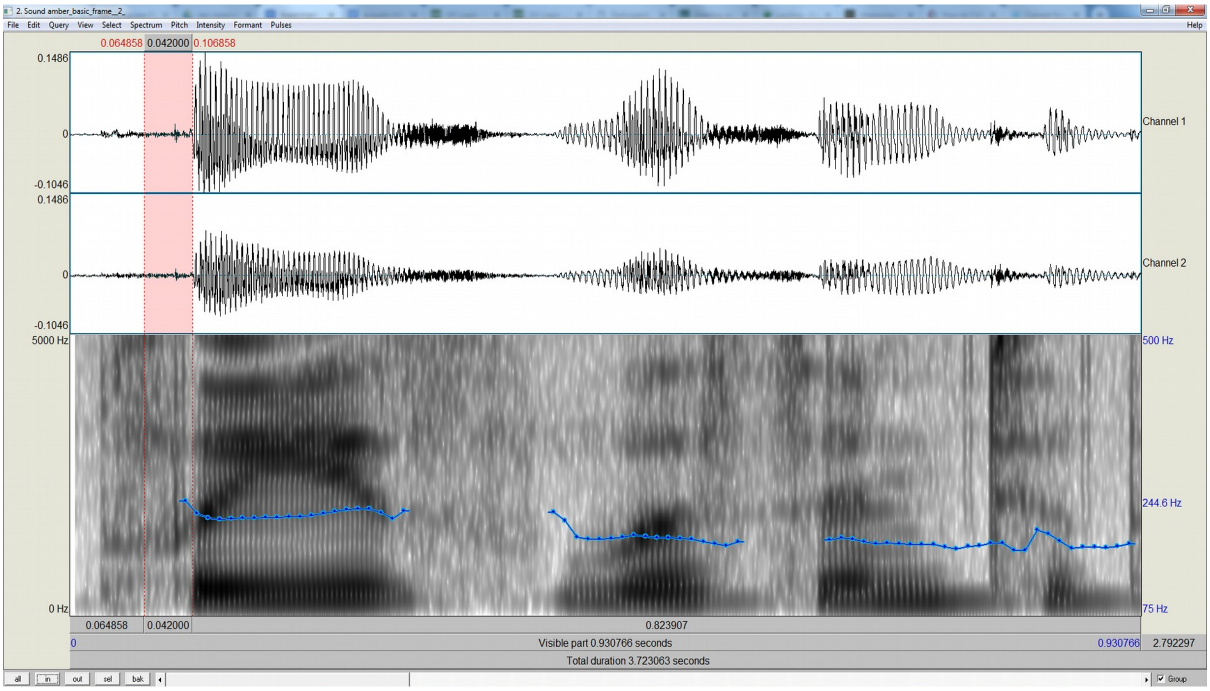

Figure 4.1: The VOT for the native talker's *please* was equalised with the non-native talker's by removing a segment from the end of the VOT portion.

4.2.1 Other Acoustic Cues

While VOT is regarded as the main marker for voicing in plosives (Lisker & Abramson, 1964), there is some evidence that listeners use F0 as a cue for voicing (Whalen et al., 1993; Winn et al., 2013). Although the average F0 of both talkers were similar as reported in Section 3.2, a post-hoc analysis revealed that there were some differences in the F0 of the vowels in *please* and *pick* between the talkers, which might have been used by the listeners to form a representation for a talker's plosives. Specifically, the native talker's /p/ in *please* and *pick* might still sound more aspirated even though they have the same VOT as the non-native talker's /p/ due to higher pitch in the native talker's vowels. The F0 and pitch of those words for both talkers are presented in Table 4.2. The F0 and pitch of the critical syllable is included for comparison.

Table 4.2: F0 is shown in Hertz (Hz) and Bark for the vowels in *please* and *pick* for the native and non-native talkers, and critical syllable (produced by native talker).

	please	*pick*	critical syllable
Native	228Hz, 2.23 Bark	185Hz, 1.82 Bark	153Hz, 1.51 Bark
Non-native	172Hz, 1.70 Bark	146Hz, 1.45 Bark	

The /b/ in *syllable* in the talkers' frame sentences were not explicitly controlled as that was not a word onset plosive. Additionally it occurred in an unstressed syllable. Hence, the VOT of this plosive was only measured after the data collection was completed. The native talker had a VOT of 20ms for /b/ while the non-native talker fully voiced her /b/. Therefore it is a possibility that this would lead listeners to form a perceptual boundary between /b/ and /p/ for the non-native talker at a shorter VOT.

4.3 Results

Figure 4.2 presents the overall proportion of /pa/ responses to these steps, averaged across all blocks in the experiment with error bars representing 95% confidence intervals. Similar to Experiment One (Figure 3.6), the slope of the non-native talker's perceptual boundary appears to be as steep as that of the native talker, with a larger area under the non-native talker's line. In addition, the VOT range where nativeness affects perception appears to be 15 – 25ms. The pattern here again appears similar to that observed in Experiment One, the crossover point between /pa/ and /ba/ is at a lower VOT when the frame sentence is produced by the non-native talker compared to when it is produced by the native talker. However, unlike Experiment One where the crossover point for the non-native talker occurred between 15 – 20ms, the crossover points for both talkers are now between 20 – 25ms. Compared to Experiment One, there is less distinction between the confidence intervals of the talkers. Figure 4.3 presents the proportion of /pa/ responses for both talkers across

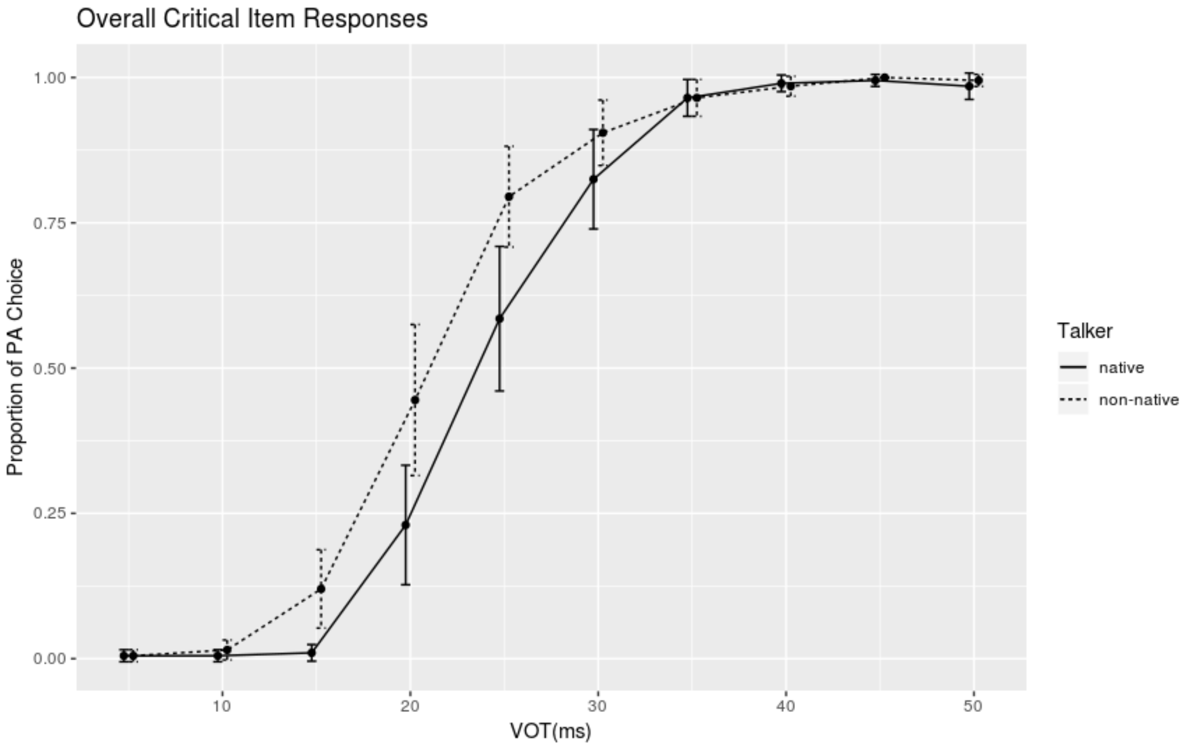

Figure 4.2: Overall proportion of /pa/ responses for critical items in Experiment Two, averaged across participants and blocks and shown separately for the talker who is a native speaker of English (native) and the one who is not (non-native). Voice onset time (VOT) from the relevant step of each continuum is shown in milliseconds (ms) on the x-axis. 95% confidence intervals for each participant's mean responses across blocks are shown.

block; there is no obvious pattern of change in /pa/ responses across all blocks.

Figure 4.4 shows each participant's overall responses averaged across all the blocks. Almost all participants perceived more items as /pa/ for the non-native talker. However, participants 5 and 26 appeared to perceive both talkers similarly. Participant 35 perceived more non-native talker items under 30ms as /pa/ while more native talker items above 30ms are perceived as /pa/. Nevertheless, these three participants were retained for analyses as there is no information from their background information and feedback responses that can provide an explanation for their divergent behavior. Compared to Experiment One, the distance between the perceptual boundaries of both talkers appear to be smaller for most participants.

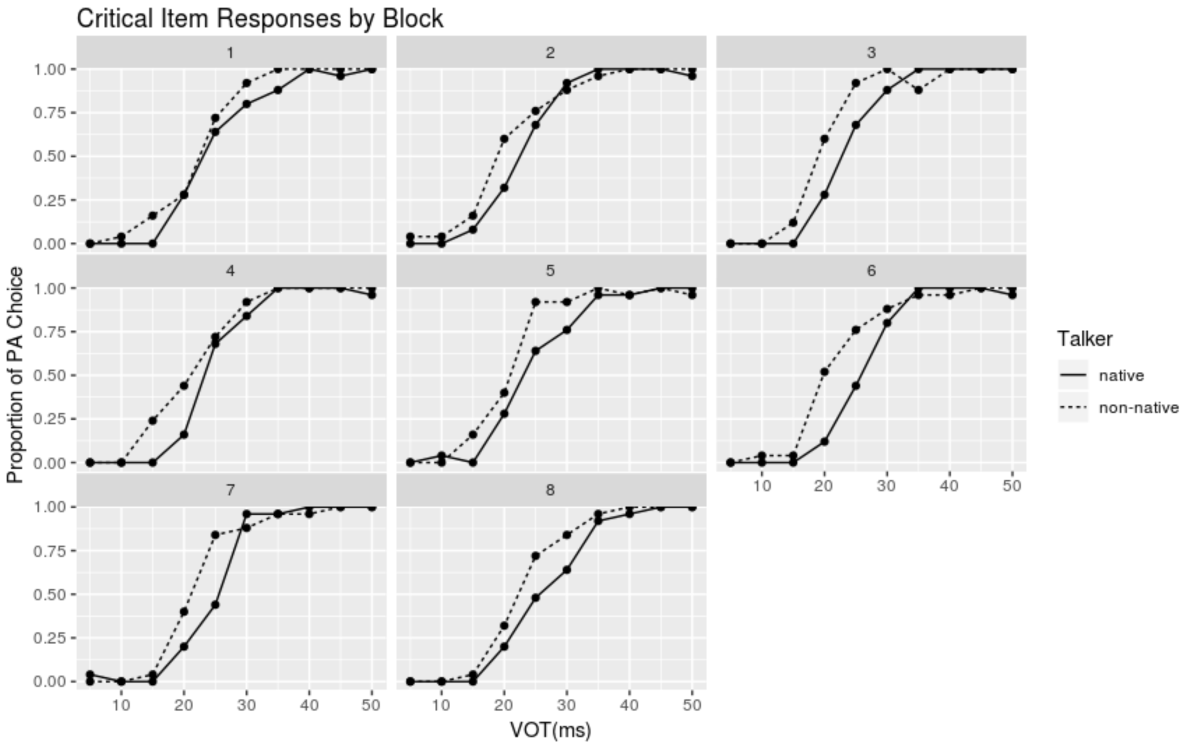

Figure 4.3: Proportion of /pa/ responses for critical items in Experiment Two across the eight experimental blocks, averaged across participants.

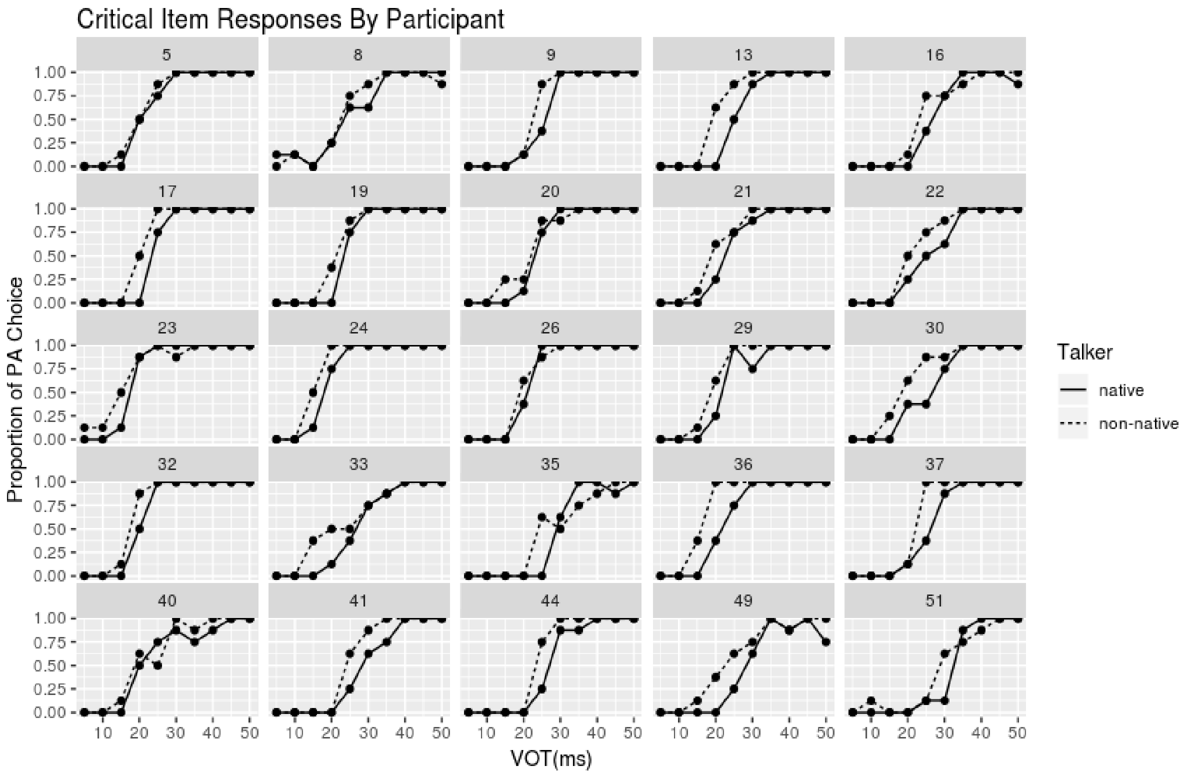

Figure 4.4: Proportion of /pa/ responses for critical items in Experiment Two across participants, averaged across blocks.

A logistic mixed effects regression model was fit to the responses to the critical syllables to determine whether the difference in responses across the talkers was significant. The most parsimonious statistical model which converged[7] is reported here. It included fixed effects of talker (native and non-native), VOT step (10 steps), and block (8 blocks), and by-participant random slopes. All predictors are centered. The reference level for talker type is the native talker, while VOT step and block are numerically ordered. The output of this model is reported in Table 4.3. All three fixed effects of talker type, VOT Step and block were revealed to be significant. Consistent with Experiment One, the results indicate that there were fewer /pa/ responses when VOT was shorter ($p<.0001$). There were also fewer /pa/ responses as the experiment proceeded through the blocks ($p<.001$). There were also more /pa/ responses for the non-native talker ($p<.0001$), although the magnitude of difference is smaller than observed in Experiment One.

Table 4.3: Output of logistic mixed effects model on responses to both talkers' VOT steps in Experiment Two. Model: glmer(Critical Responses ~ Talker + VOT Step + Block + (1|Participant), data = Native and Non-native Talkers, family = binomial)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.6011	0.2523	6.346	< .0001
Talker	1.1148	0.1388	8.034	< .0001
VOT Step	0.3506	0.0131	26.823	< .0001
Block	-0.1091	0.0292	-3.737	0.0002

The trends reported in Table 4.3 confirm the hypothesis that listeners do

7 The model reported in Table 4.3 was determined through a series of ANOVA model comparisons to be as well-fit as a more complex model containing non-significant two-way interactions of talker and VOT step, and talker and block.

perceive a talker with a non-native accent differently despite being exposed to frame sentence cues indicating both native and non-native talkers have similar VOT. While not included in the mixed effects model due to the small sample size, there is a possibility that participants being aware of the splice could affect sensitivity to the prime. Besides that, a participant's ability to recognise the non-native talker's native language might also affect their perception of the syllables. These issues will be explored in the following two sections.

4.3.1 Participants who Reported Noticing Splicing in Stimuli

Of the 25 participants who were analysed, 9 reported after being prompted that they noticed that syllables from both sets of talkers were produced by the same talker and spliced into the frame sentences. Figure 4.5 compares the responses of these participants with the other participants who reported being unaware of the splice. While they did not claim to notice the splice until after they were prompted, it is still possible that these participants behaved differently on the task. While the number of participants who responded in this way is too small for a "noticing factor" to be included in the model reported in Table 4.3, possible trends that stem from noticing are explored here. This is to determine firstly, whether their perception is similar to participants who did not notice the splice; secondly, whether there is a difference in their perception of both talkers.

Similar to Experiment One, participants who reported noticing the splice ('Yes' group) responded more often with /pa/ than participants who did not report noticing the splice ('No' group). This is likely merely noise that would not be observed with a larger sample size. Despite reporting that they noticed the modification after they were prompted, responses from participants who noticed the splice varied according to talker type, and the slopes of the lines resemble those from the group who did not notice the splice. This suggests that participants are still cued by the frame sentences regardless of whether they notice that the critical syllable was produced by the same talker.

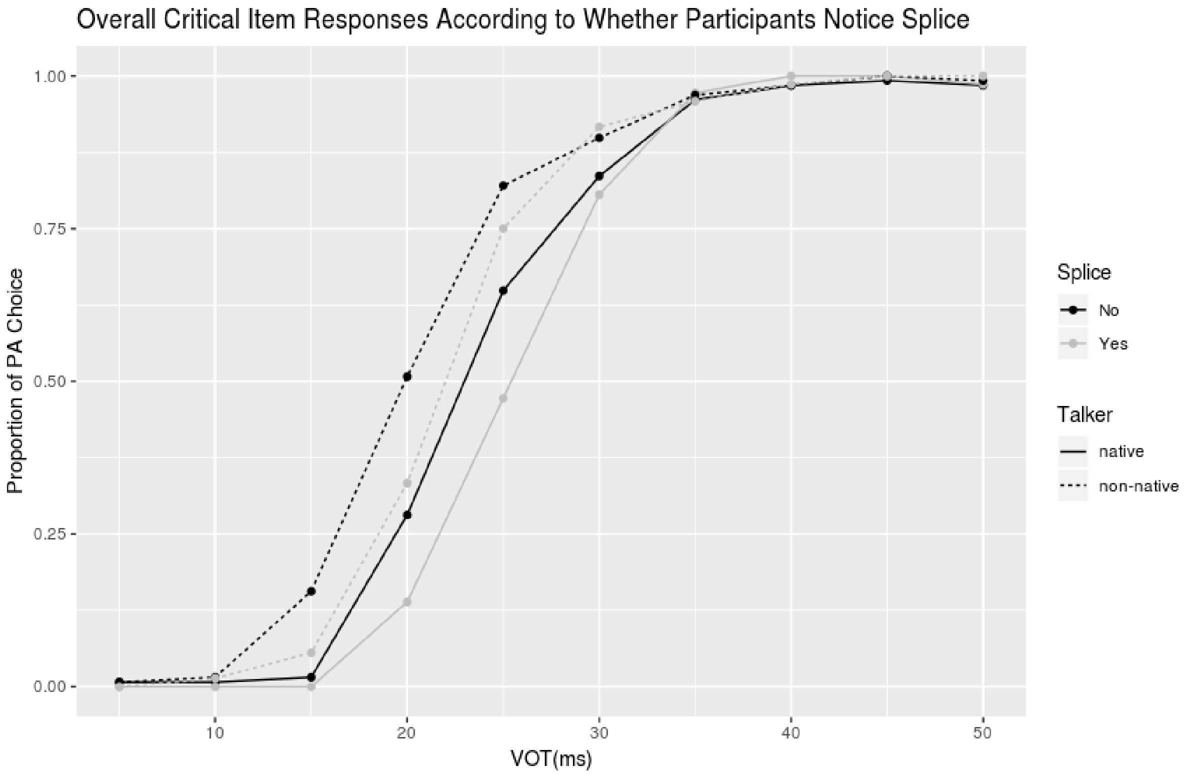

Figure 4.5: Proportion of /pa/ responses for critical items in Experiment Two, averaged across participants and blocks. Proportions are shown separately for participants who reported noticing the splice after being asked explicitly (grey, N=9) and those who reported that they did not notice the splice (black, N=16), and for whether the talker was a native speaker of English (solid) or not (dotted).

4.3.2 Participants who Perceived the Non-native Talker as an L1 Spanish Speaker

When queried about the non-native talker's native language, 8 of the participants responded with Spanish, while the remaining 17 answered with languages other than Spanish. It is possible that the group of participants who were able to correctly identify the non-native talker's language might have prior experience with non-native talkers who have similar accents to the non-native talker in Experiment Two and that this could affect their perception of sounds produced by the non-native talker during the task. Figure 4 .16 compares the responses between the group who were able to correctly identify the talker's native language and the group who were unable to do so. This is to determine if there was any trend of responses differing between both groups.

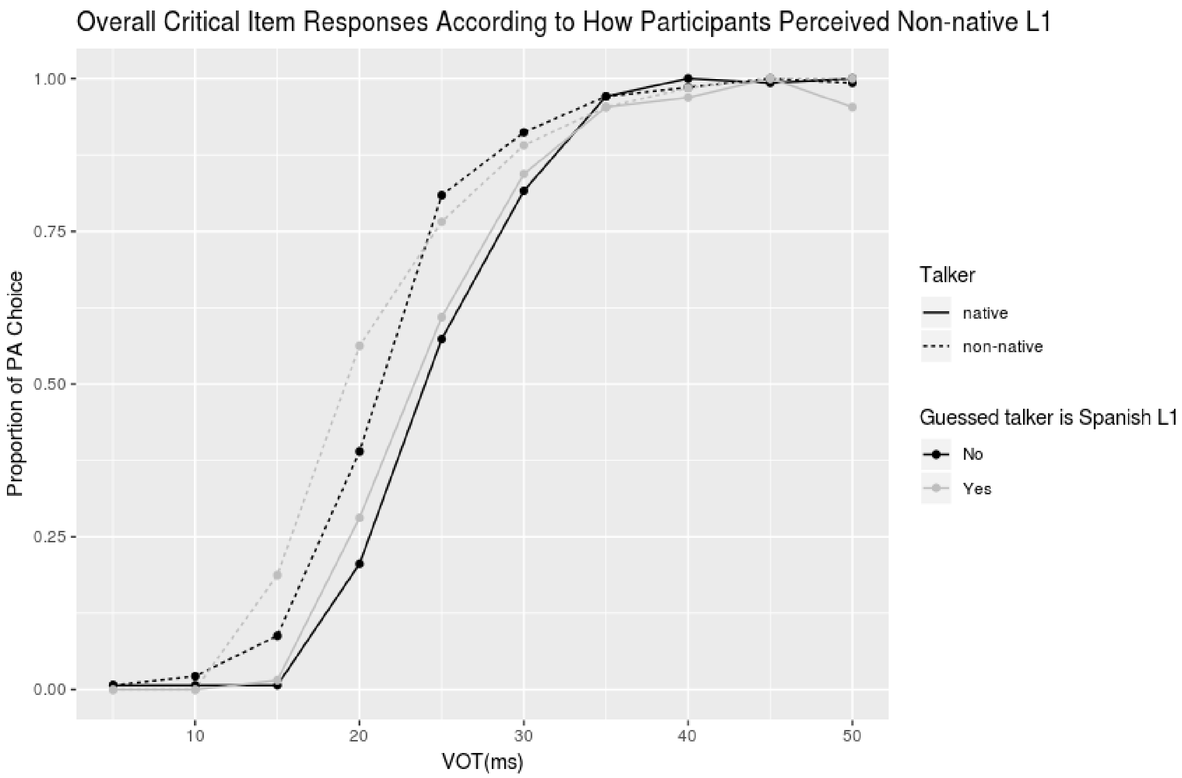

Figure 4.6: Proportion of /pa/ responses for critical items in Experiment Two, averaged across participants and blocks. Proportions are shown separately for participants who responded perceiving the non-native talker as having Spanish as L1 (grey, N=8) and those who responded that the non-native talker has an L1 other than Spanish (black, N=17), and for whether the talker was a native speaker of English (solid) or not (dotted).

Both groups of participants appear to trend similarly in terms of their responses to the native talker. However, there appears to be a difference in responses to the non-native talker, particularly at the 15ms and 20ms steps. Specifically, the participants who correctly identified the non-native talker as a native speaker of Spanish (the 'yes' line) had a higher proportion of /p/ responses for those steps compared to those participants who did not accurately identify the talker's native language. Overall, the the participants who correctly identified the non-native talker's native language appears to have slightly more /p/ responses for both talkers. Note that this contrasts from Experiment One (Section 3.3.2), in which the group who correctly identified the non-native talker's native language had slightly fewer /p/ responses for both talkers. The factor of whether a participant perceived the non-native talkers L1 as Spanish or otherwise was added to the model reported in Table 4.3, both as a fixed effect and in an interaction with talker. However, both the fixed effect and interaction were not found to be significant. Given the small sample size, it is possible that the differences between the groups is merely noise, especially since the trend is in the opposite direction as in Experiment One was non-significant for both experiments.

4.4 Comparison of Experiments One and Two

The modification of the native frame sentence from Experiment One to Experiment Two brought about a reduction in the difference between the perceptual boundaries for both talkers, which is supported both graphically (compare Figures 3.6 and 4.2) and statistically (compare Table 3 .3 and Table 4 .6, fixed effect of Talker). However it is not as readily apparent how this reduction was reached. It is possible that listeners' perceptual boundaries for either one or both talkers have 'shifted'[8]. If

8As the comparisons being made here and later in Section 6.3 are between results obtained from different groups of participants, there is no real shift in listeners' perception of the non-native talker here and of the native talker in Section 6.3. I use 'shift' as a general term to describe the changes observed in responses to similar the perceptual boundary for the native talker shifted, it would indicate that listeners incorporate the shorter VOT cues of native talker's frame sentence. If it was the perceptual boundary for the non-native talker that shifted instead, it would indicate that the perceptual boundary for a native talker is relatively stable regardless of any idiosyncratic cues and that it is the relative difference (or lack of one) in VOT duration between both talkers that has an effect on perception of non-native speech. A shift in perceptual boundaries for both talkers could mean that listeners have a less stable perceptual boundary for native talkers that is amenable to revision based on

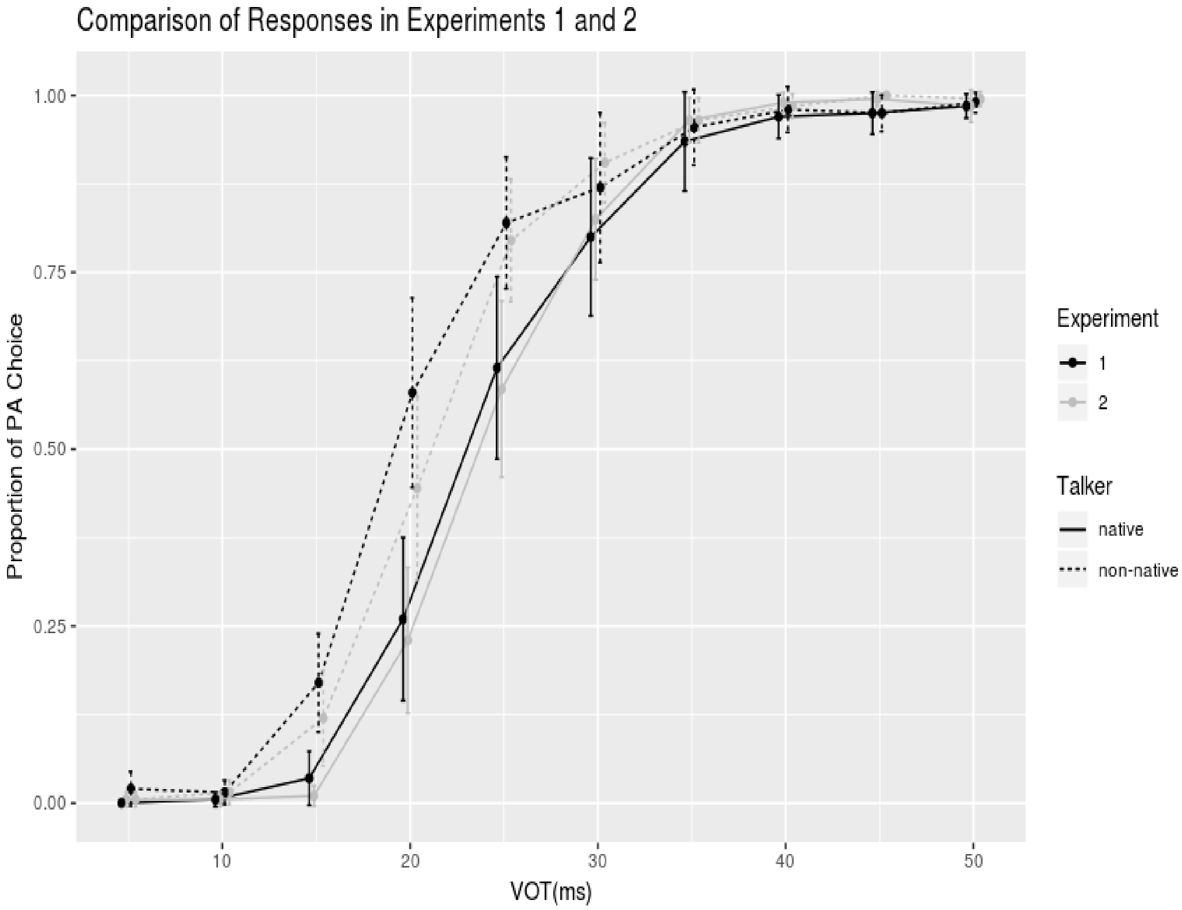

Figure 4.7: Overall proportion of /pa/ responses for critical items in Experiments One and Two, averaged across participants and blocks and shown separately across experiment, and for the talker who is a native speaker of English (native) and the one who is not (non-native). Voice onset time (VOT) from the relevant step of each continuum is shown in milliseconds (ms) on the x-axis. 95% confidence intervals for each participant's mean responses across blocks are shown.

available cues while also using the relative difference in VOT duration between both talkers to aid their perception of non-native speech.

While a comparison between the results of Experiments One and Two is not ideal in that the VOT cue modification factor is not tested on a within participant basis, both sets of participants are similar in terms of their demographics and also were exposed to the same non-native stimuli. Figure 4.7 compares the overall proportion of /pa/ responses to the 10 VOT steps for both talkers between Experiments One and Two, averaged across all blocks in those experiments. Error bars representing 95% confidence intervals were included to show the amount of variation between the participants' mean responses.

Based on Figure 4.7, the responses to the native talker appears to be similar between both experiments. The proportion of /pa/ responses appear to be slightly lower in Experiment Two for the VOT steps of 15-30ms, but this appears to be reversed for the 35ms step and longer. However, there appears to be a larger difference between the experiments for responses to the non-native talker. The perceptual boundary for the non-native talker appears to have shifted to a higher VOT in Experiment Two. A relatively smaller proportion of items at the 15-25ms steps were perceived as /pa/ in Experiment Two.

4.5 Discussion

The results of Experiment Two are largely consistent with the findings from Experiment One: listeners have different but parallel perceptual boundaries for the native and non-native talkers, and they have more /pa/ responses when listening to the non-native talker. This provides evidence that listeners' perception of word initial plosives are influenced by the accentedness of a talker in a way that is likely consistent with their prior experience with similar talkers; they needn't be exposed to differing VOT values of word initial voiceless plosives in the talker's frame sentences in order for their perception of word initial plosives to be affected by exposure to those talkers. However, the magnitude of the effect is smaller in Experiment Two compared with Experiment One. This suggests that any VOT cues which signal differences in duration have a cumulative effect with a talker's accentedness, which in turn influences listeners' perceptual boundaries. In other words, even when VOT cues are similar listeners rely on talker accentedness to form perceptual boundaries. A closer look at individual participant responses (Figure 4.4) suggests that there may be individual level variation. Two participants (Participants 5 and 26) have virtually identical perceptual boundaries for both talkers, indicating that they might rely more on using VOT cues than talker accentedness. This might be taken as an indication that different listeners use different strategies or combinations of strategies for perceiving non-native speech, although future research that explicitly tests perception strategies would be worthwhile.

A closer examination of the talkers' frame sentences suggest that other acoustic cues might have a role to play in the participants' responses. Although the /b/ in *syllable* appears in a less prominent phonetic environment, different to the focus of the dissertation, it is possible that it served as a cue for the participants. They might have extrapolated the shorter burst of the non-native talker /b/ to her having a category boundary between /b/ and /p/ at a shorter VOT duration.

In addition to VOT, F0 has been cited as having an effect on the voicing perception of plosives. Previous literature has claimed that a lower F0 is more likely to cause a segment to be perceived as a voiced plosive (Whalen et al., 1993).

Although the non-native talker has similar F0 for the overall frame sentence, her F0 is lower than the native talker's for the vowels in the /p/ cue words *please* and *pick*. The contrast of a lower F0 (generally associated with voiced plosives) in words beginning with a voiceless plosive might still cause listeners to perceive the non-native talker as having more /b/-like /p/. This in turn might cause them to adjust their category boundary between the non-native talker's /p/ and /b/ towards a shorter VOT duration. Nevertheless, any adjustments in category boundary arising from other non-VOT cues would still be interesting in that the differences in perceptual boundaries are manifested in critical syllables which only differ in terms of VOT. This would support a more experience based model of perception. Instead of building a representation from existing cues, listeners use whichever cues they have access to to activate pre-existing representations in their memories. These representations contain information on other cues, which can then be matched to the acoustic signal.

The comparison between the results of Experiments One and Two suggest that listeners may have a more stable perceptual boundary for a native talker that is more resistant to revision despite the presence of idiosyncratic cues. Conversely, it is possible that, when listening to a non-native talker, the perceptual boundary between /pa/ and /ba/ might be affected by phonetic realizations produced by a second talker who is a native speaker. A possible interpretation is that listeners have an available frame of reference or representation for native talkers and are less likely to make use of idiosyncratic cues for that talker; on the other hand, listeners may have less prominent representations for a non-native talker and therefore may then need to make use of any phonetic cues present, even that of another talker. An alternative explanation would be the longer VOT duration of the native talker /p/ in Experiment

One highlights the short VOT duration of the non-native talker /p/, leading the listeners to perceive more of the non-native items as /pa/ at a shorter VOT. As the VOT of the native talker /p/ is shortened in Experiment Two, the short VOT of the non-native talker /p/ is less prominent, leading the listeners to perceive fewer non-native items as /pa/ compared to Experiment One.

Altogether, the results from Experiment Two provide evidence that there is an effect of whether a talker is a native or non-native speaker even when VOT in /p/ is held constant across the talkers. Next, Experiment Three aims to investigate whether this accentedness effect is still observed with a non-native talker who has a different native language.

Chapter 5 Experiment Three

The results from Experiment Two provides evidence that listeners' perception of /pa/ and /ba/ can be influenced by differences in talker accent rather than solely influenced by VOT cues from *please* and *pick* in the talkers' frame sentences. However, the difference in responses between conditions was stronger in Experiment One than in Experiment Two, providing evidence that VOT cues might work in conjunction with talker accent in forming a listener's perceptual boundary.

Talker accent, here, is broadly defined, referring to whether a talker is a native speaker of American English or not. There are a number of different cues that listeners might have been influenced by, such as speech rate and F0. While the results from Experiments One and Two demonstrate that listeners' responses were influenced by talker accent, it is unclear whether similar effects would be observed with a different non-native accent. For instance, listeners' perception of word initial bilabial plosives might be influenced by talker- or accent-specific expectations, or they might be influenced by non-accented vs. accented talkers more generally. Experiment Three addresses the question: Do listeners have similar expectations for all talkers they identify as accented, or is the effect accent-specific? In other words, do listeners perceive an accented talker as having a general non-native accent, or do they associate the talker with specific non-native accents? Experiment Three tests this explicitly by using a frame sentence produced by a different non-native talker whose accent has different VOT patterns than that of the non-native talker from Experiments One and Two. Compared to the Spanish L1 accent used in the previous experiments which tends to have unaspirated (short VOT) voiceless plosives, Experiment Three uses a talker with a Chinese (Mandarin L1) accent which tends to aspirate (long VOT) voiceless plosives (Chen et al., 2007). This means that the non-native talker accent this time would pattern similarly to the native talker accent as far as plosives are concerned. While the Chinese talker's /p/ in *pick* and *please* are naturally aspirated, the VOT of those tokens are controlled so that the voice onset times in the native talker's two word initial plosives have the same duration as those of the non-native talker in the previous experiments. The native talker frame sentence was retained from Experiment Two. Controlling the VOT across all talkers was done to test if the difference in perceptual boundaries seen in Experiment Two is still observed across both non-native talker accents. I hypothesized that there would still be a difference in the listeners' perceptual boundaries for both talkers due to the accentedness effect, despite both accents aspirating voiceless plosives.

This chapter first describes the subjects who participated in the study (Section 5.1), and then explains how the materials were modified from those used in Experiment Two (Section 5.2). The results are then presented in Section 5.3. Section 5.4 compares the results of Experiments One and Two. Chapter 5 is then concluded by a discussion of the findings (Section 5.5).

5.1 Participants

Similar to Experiment Two, participants were recruited through the University of Hawaiʻi at Mānoa Linguistics Beyond the Classroom program and via word of mouth, and they received partial class credit or a gift card for taking part. 46 participants who did not take part in Experiments One and Two took part in the present experiment. The analysis was conducted on responses from monolingual native speakers of American English who fit the criteria described in Experiment One (Section 3.1). Data from participants who did not meet these criteria were removed.

One additional participant was removed because she knew the non-native talker. After these participants were excluded, responses from 25 participants (15 female, 10 male) were analysed. All participants underwent experimental procedures identical to the participants of Experiments One and Two, including giving written consent prior to the experiment and undergoing the post-experiment interview. A breakdown of the backgrounds of the monolingual native speaker participants is provided in Table 5.1.

Table 5.1: Breakdown of monolingual native speaker participants' demographic information according to gender, age, and their response for the non-native talker's L1.

	Non-native talker L1 = Chinese	Non-native talker L1 = Non-Chinese[9]	Age: min, median, max
Female	1	14	18, 20, 22
Male	2	8	19, 20, 30

5.2 Materials and Procedure

The non-native talker in Experiment Three was a female Mandarin speaker[10] from Shanghai, China, who was in her early 30s. She replaces the Spanish speaker who was the non-native talker in Experiments One and Two. This new non-native talker is a graduate student trained in linguistics who had also been living in the United States of America for approximately 5 years at the time of recording. She recorded the same frame sentence as the Spanish and American English talkers. While the bilabial plosives in the Mandarin talker's frame sentence had slightly shorter VOT for /p/ in *please* (76ms) and much longer VOT in *pick* (84ms) than the respective plosives produced by the American English talker, they were reduced using Praat (Boersma & Weenink, 2019) in order to match the VOT duration across talkers; the VOT durations are the same as those used in the frame sentences in Experiment Two. The method used in shortening the VOT of these tokens is identical to that used to

9 No participant identified the non-native talker as being a Spanish L1 speaker.

10 In addition to being a native speaker of Mandarin, she is also a native speaker of Shanghainese and a heritage speaker of Pudonghua.

shorten the native talker's tokens in Experiment Two (Section 4.2). In addition, the non-native talker used in Experiment Three had a slower speech rate of 3.1 words per second, which is slower than both the other two talkers. The effect this difference in speech rate may have had on responses is discussed in Section 5.4. If listeners use speech rate as a cue, the perceptual boundary should be at a shorter VOT duration for the non-native talker, in fact it should be even shorter than that of the non-native talker used in Experiments One and Two. The non-native talker had a mean F0 of 199Hz (1.96 Bark) for her frame sentence, just over .2 Bark higher than the mean F0 of the native talker's frame sentence (Section 3.2).

The native talker items (frame sentence and syllables) were retained from Experiment Two. Other than the change in non-native talker, the procedure and stimuli construction methods for Experiment Three were also identical to that of Experiments One and Two (Sections 3.2 and 4.2).

5.2.1 Other Acoustic Cues

The /b/ in syllable in the talkers' frame sentences were not explicitly controlled as that was not a word onset plosive. Additionally it occurred in an unstressed syllable. Hence, the VOT of this stop was only measured after the data collection was completed. The native talker had a VOT of 20ms for /b/ while the non-native talker had a VOT of 13ms. If listeners use /b/ as a cue, this would indicate a perceptual boundary for the non-native talker at a similar or slightly shorter VOT duration. Additionally the F0 of the new non-native talker is reported in Table 5 .9 below. While the non-native talker's F0 for *please* is much higher than the native talker's, her F0 for *pick* is similar to that of the native talker's. Therefore, the non-native talker's /p/ should sound more voiceless than the native talker's, at least for *please*, leading to a perceptual boundary at a longer VOT duration.

Table 5.2: F0 is shown in Hertz (Hz) and Bark for the vowels in *please* and *pick* for the native and non-native talkers, and critical syllable (produced by native talker).

	please	*pick*	critical syllable
Native	228Hz, 2.23 Bark	185Hz, 1.82 Bark	153Hz, 1.51 Bark
Non-native	275Hz, 2.66 Bark	188Hz, 1.85 Bark	

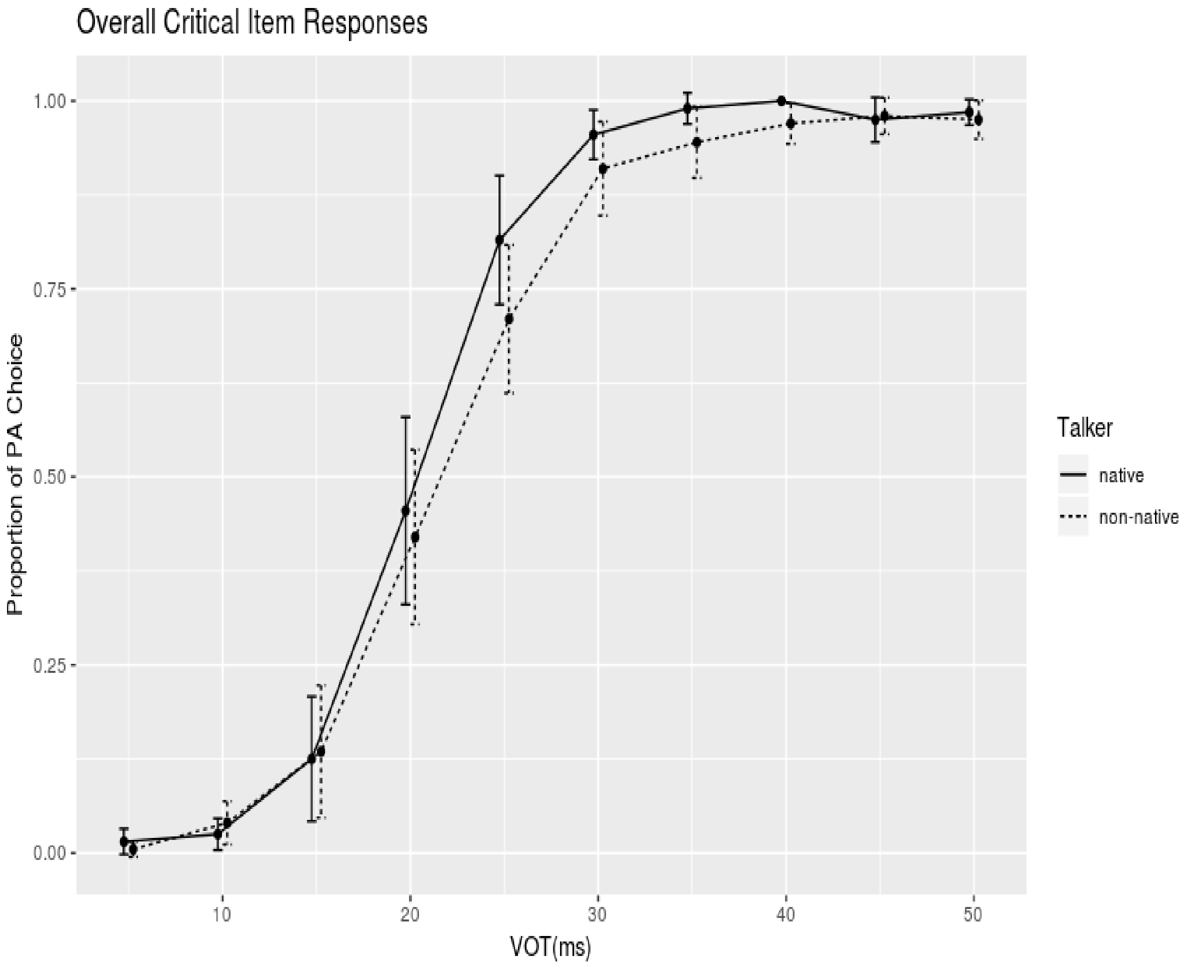

Figure 5.1: Overall proportion of /pa/ responses for critical items in Experiment Three, averaged across participants and blocks and shown separately for the talker who is a native speaker of English (native) and the one who is not (non-native). Voice onset time (VOT) from the relevant step of each continuum is shown in milliseconds (ms) on the x-axis. 95% confidence intervals for each participant's mean responses across blocks are shown.

5.3 Results

Figure 5.1 presents the overall proportion of 'pa' responses to these steps, averaged across all blocks in the experiment. In contrast with Experiments One

(Figure 3 .1) and Two (Figure 4 .11), a greater proportion of 'pa' responses were made when listening to the native talker compared with the non-native talker, with a larger area under the native talker line. There also appears to be a difference in the VOT range where nativeness affects perception compared to the previous experiments. Here, there is a difference in responses between talkers at the 20 – 35ms steps. Nevertheless, confidence intervals for both talkers overlap at all steps, even for the ones where nativeness affects perception. Overall, there is less variability between talkers compared to Experiments One and Two.

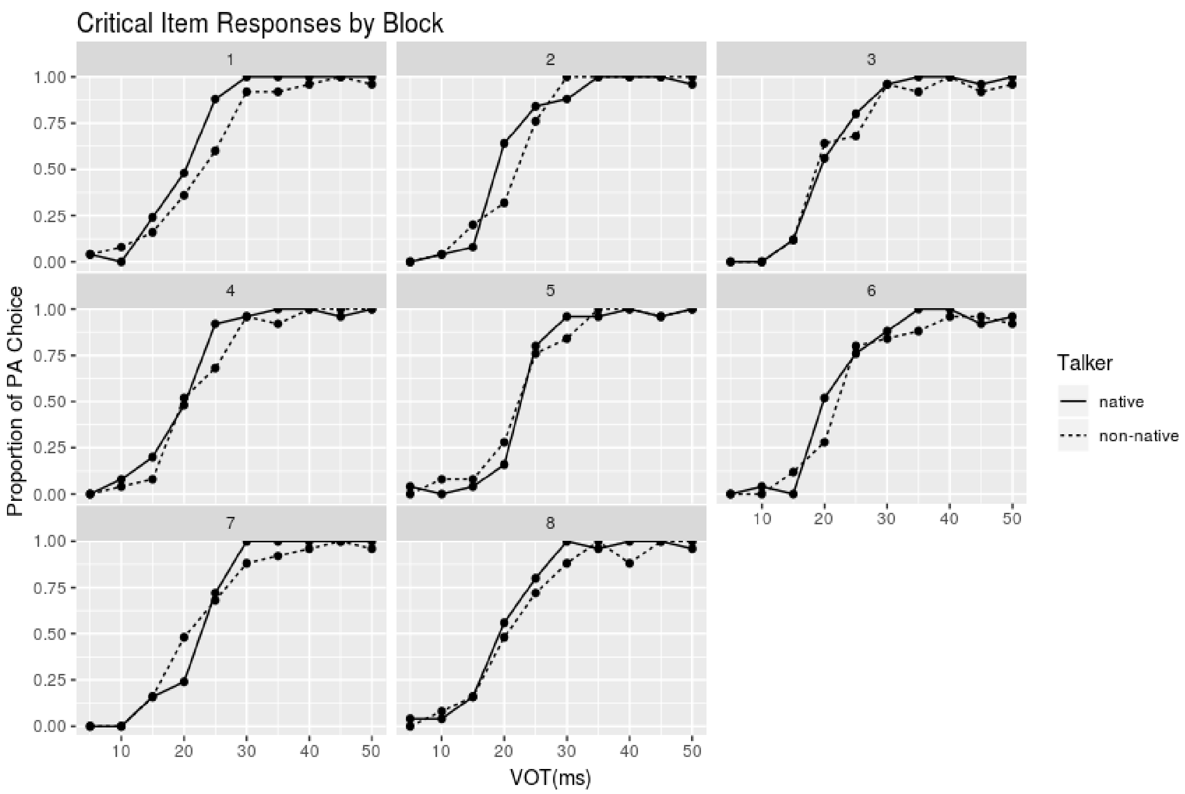

Figure 5.2: Proportion of 'pa' responses for critical items in Experiment Three across the eight experimental blocks, averaged across participants.

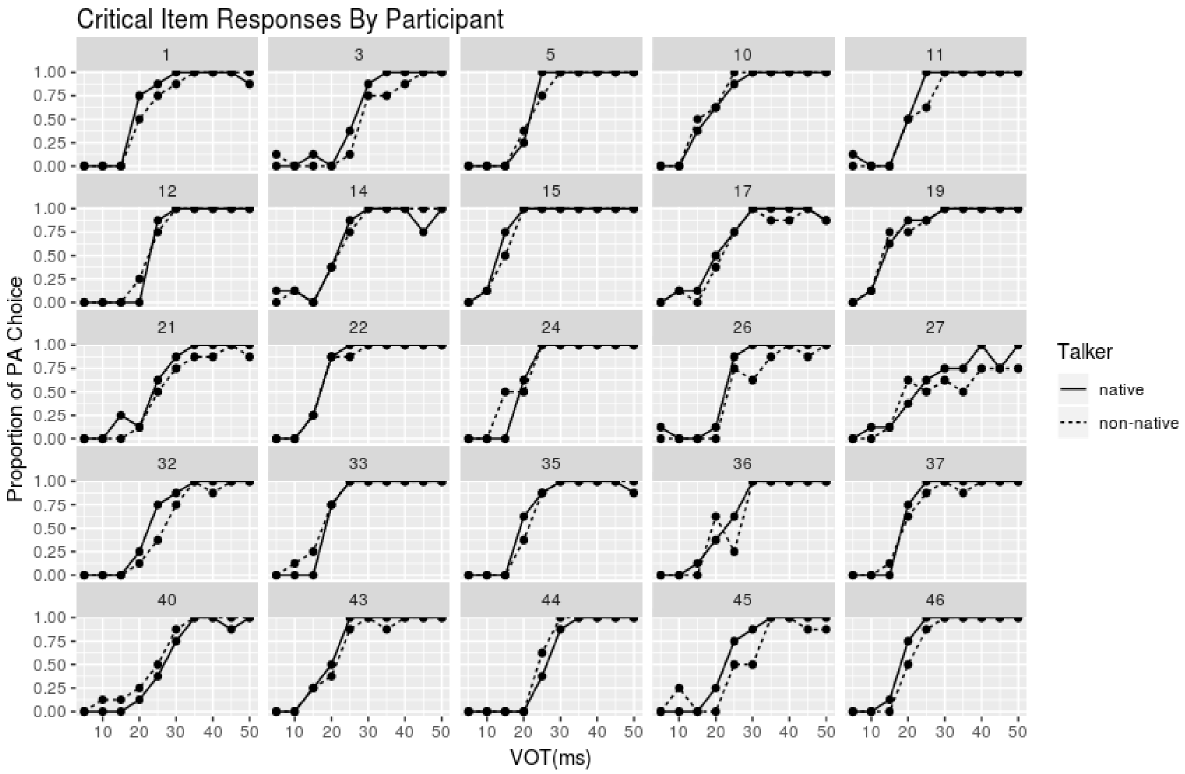

Figure 5.3: Proportion of /pa/ responses for critical items in Experiment Three across participants, averaged across blocks.

As shown in Figure 5.2, there is variability across blocks; there appears to be a small difference in perceptual boundary between talkers in blocks 1, 2, 4, and 5, while there does not appear to be any consistent difference in the mean proportion of /pa/ responses in blocks 3, 6, 7, and 8. Figure 5.3 shows each participant's overall responses averaged across all the blocks. Almost all participants perceived more items as /pa/ for the native talker. Only participants 40 and 44 perceived more items as /pa/ for the non-native talker. Nevertheless, these two participants were retained for analyses as there is no information from their background information and feedback responses that can provide an explanation for their divergent behavior. Compared to Experiments One and Two, the distance between the perceptual boundaries of both talkers appear to be much smaller.

To determine whether the difference in responses when listening to the native and non-native talker was significant, a logistic mixed effects regression model was fit to the responses to the critical syllables. The statistical model[11] included fixed effects of talker (native and non-native), VOT step (10 steps), and block (8 blocks), with a three-way interaction between talker, VOT step and block, two-way interactions of talker and VOT step, talker and block, VOT step and block, and by-participant random slope for VOT step. All predictors are centered. The reference level for talker type is the native talker, while VOT step and block are numerically ordered. The output of this model is reported in Table 5.3. All fixed effects and the interactions between talker type and VOT step, and VOT step and block were revealed to be significant. Despite appearing to lack a consistent shift in Figure 5.2, the significance of block indicates there were fewer /pa/ responses toward the beginning of the experiment, with the number of /pa/ responses increasing over the course of the experiment ($p<.05$). The model also confirmed the graphical observation that there were significantly more /pa/ responses for the non-native talker ($p<.05$). Unlike Experiment One, there are significant interactions of talker type and VOT step ($p<.005$). The negative estimated intercept value for the interaction between talker type and VOT step indicates that there were fewer /pa/ responses for the non-native talker compared to the /pa/ responses for the native talker as the experiment progressed. The VOT step and block interaction also had a negative value, which indicates that a given VOT step is less likely to be perceived as /pa/ as the experiment progressed.

11 The model reported in Table 5.3 was determined to have the best fit for the data through a series of ANOVA model comparisons. Simpler models lacking the non-significant interactions either did not converge or were found to be worse in fit.

Table 5.3: Output of logistic mixed effects model on responses to both talkers' VOT steps in Experiment Three. Model: glmer(Critical Responses ~ Talker*VOT Step*Block + (VOT Step|Participant), data = Native and Non-native Talkers, family = binomial)

	Estimate	Std. Error	z value	Pr(>|z|
(Intercept)	2.3697	0.3551	6.673	< .0001
Talker	-0.6811	0.1560	-4.365	< .0001
VOT Step	0.3509	0.0261	13.428	< .0001
Block	-0.0749	0.0338	-2.216	0.0267
Talker:VOT Step	-0.0539	0.0168	-3.212	0.0013
Talker:Block	0.0705	0.0676	1.043	0.2969
VOT Step:Block	-0.0038	0.0036	-1.035	0.3009
Talker:VOT Step:Block	0.0013	0.0073	0.172	0.8633

5.3.1 Participants who Reported Noticing Splicing in Stimuli

Of the 25 participants who were analysed, 13 reported after being prompted that they noticed that syllables from both sets of talkers were produced by the same talker and spliced into the frame sentences. Figure 5.4 compares the responses of these participants with the other participants who reported being unaware of the splice. The responses appear to group together according to whether the participants perceived the splice in the stimuli, especially in the 20ms VOT step. However, the confidence intervals suggest that this difference might be due to noise and small sample size. Nevertheless, the model presented in Table 5.3 was modified with the addition of Splice as a fixed effect. This model did not converge. Another modified model with a simplified random effect of by-participant slope did converge but Splice was not significant. Regardless of whether participants noticed the splice, both groups still appeared to have different perceptual boundaries for both talkers.

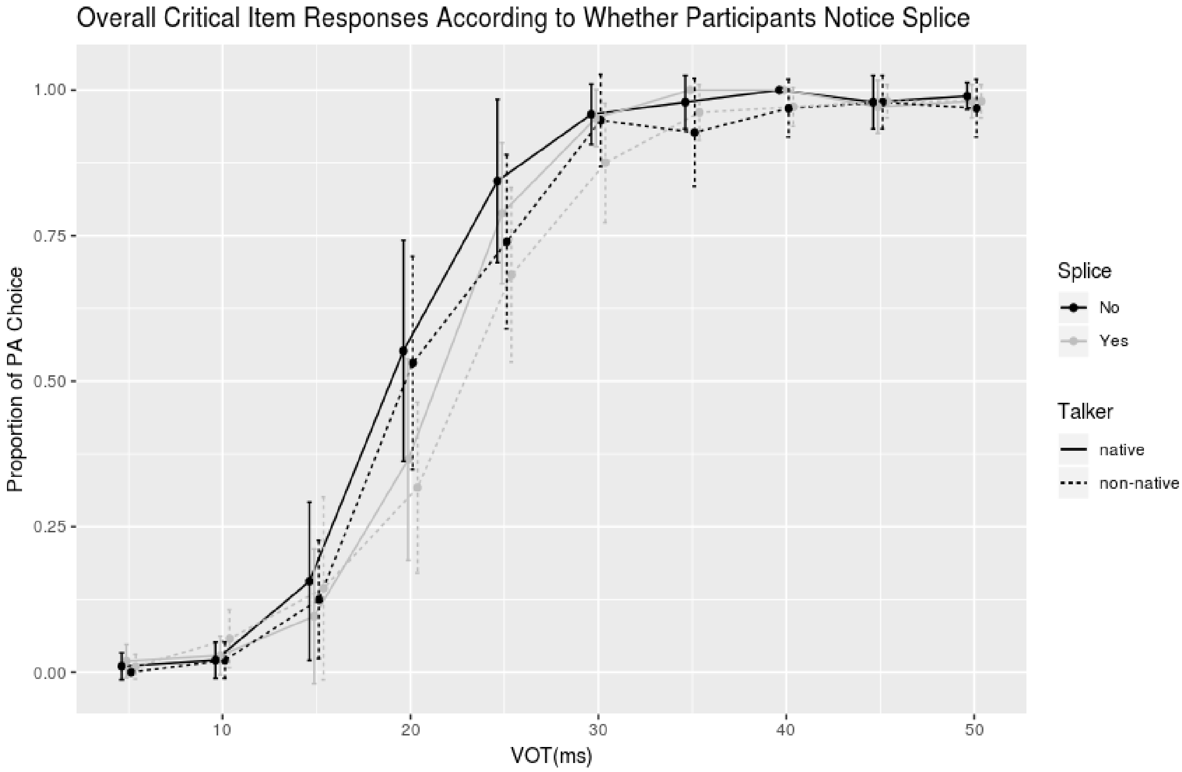

Figure 5.4: Proportion of /pa/ responses for critical items in Experiment Three, averaged across participants and blocks. Proportions are shown separately for participants who reported noticing the splice after being asked explicitly (grey, N=13) and those who reported that they did not notice the splice (black, N=12), and for whether the talker was a native speaker of English (solid) or not (dotted).

It is worth noting that participants who noticed the splice tend to have a lower perceptual boundary across all three experiments than those who do not. Nevertheless it does not seem to affect the degree to which talker accent has an effect.

5.3.2 Participants who Perceived the Non-native Talker as an L1 Mandarin Speaker

Only 3 of the participants, Participants 26, 45 and 46, perceived the non-native talker as being a native speaker of Mandarin. Consequently, the overall difference between participants who perceived the non-native talker as being a native speaker of Mandarin compared to other languages was not graphed. Based on Figure 5.3, the responses of these participants did not seem very different compared to the other participants, with all three having more /pa/ responses for the native talker and only having small differences in perceptual boundaries between both talkers.

5.4 Comparison of Experiments Two and Three

Experiment Three introduced a frame sentence produced by a talker with a different non-native accent from the non-native accent in Experiment Two. Despite having similar aspiration patterns to the native talker, listeners still perceived the syllables spliced in the non-native talker frame sentence differently from those spliced in the native talker frame sentence. However, the patterns in perception were reversed: in contrast with results from Experiments One and Two, listeners have a perceptual boundary for the native talker at a shorter VOT duration and are more likely to perceive syllables in the native talker condition as being /pa/. These findings, in conjunction with those of Experiments One and Two suggest that the perceptual boundaries of all talkers are such that listeners will be most likely to identify a syllable as /pa/ when the frame sentence is produced by the Spanish-accented, non-native talker and least likely when it is produced by a Chinese-accented, non-native talker.

Figure 5.5 which compares the responses from Experiments Two and Three show that the situation described above is not straightforward. There is a large difference in the perceptual boundary for the native talker between both experiments, having ‘shifted’ to a shorter VOT step when paired with the Chinese talker, even shorter than that of the Spanish talker in Experiment Two. This difference in perceptual boundary is apparent for the 15 – 30ms VOT steps where confidence intervals do not overlap. The perceptual boundaries for both non-native talkers were adjacent to each other with overlapping confidence intervals.

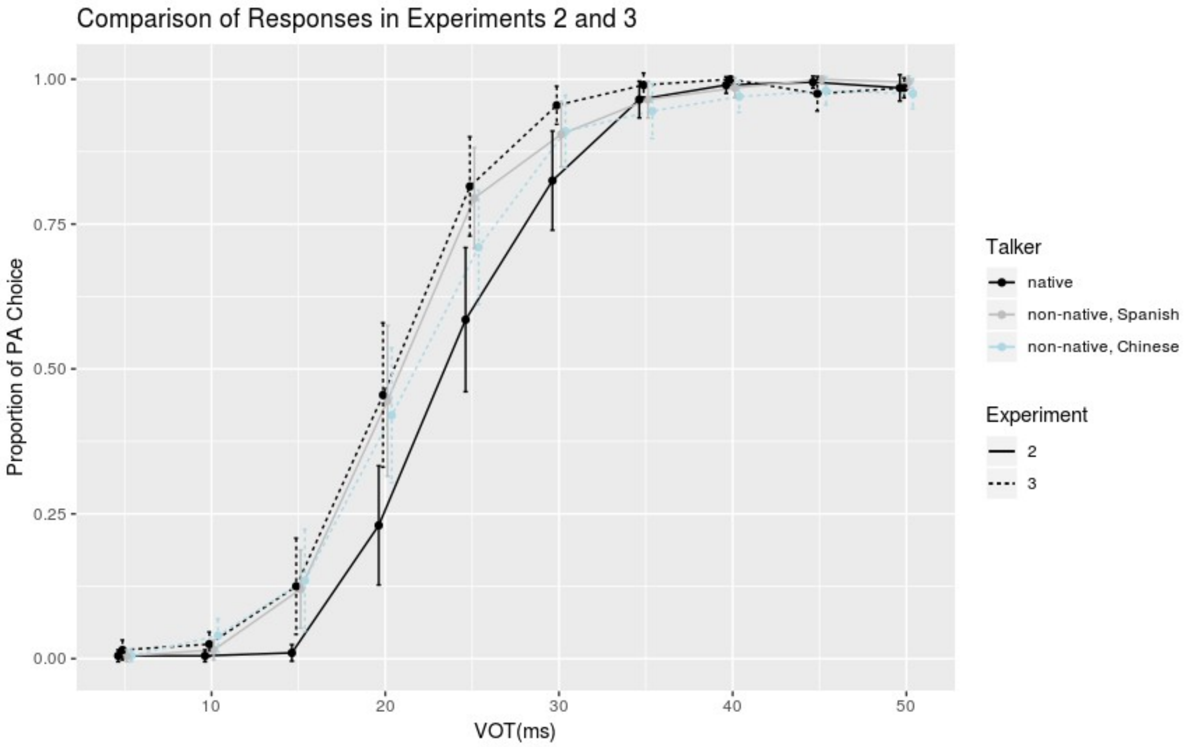

Figure 5.5: Overall proportion of /pa/ responses for critical items in Experiments Two and Three, averaged across participants and blocks and shown separately according to Experiment Two (solid) and Experiment Three (dotted), and for the talker who is a native speaker of English (black), who is Spanish non-native (grey) and who is Chinese non-native(light blue). Voice onset time (VOT) from the relevant step of each continuum is shown in milliseconds (ms) on the x-axis. 95% confidence intervals for each participant's mean responses across blocks are shown.

5.5 Discussion

Although listeners in Experiment Three perceive the syllables spliced in native talker and non-native talker frame sentences differently, they have different patterns of perception for the talkers compared to the previous experiments. The perceptual boundary for the native talker is at a lower VOT step compared to that for the non-native talker. Likewise, participants have more /pa/ responses when listening to the native talker. Both of these findings contrast with results from Experiments One and Two, in which more /pa/ responses were observed for tokens when listeners were exposed to the non-native talker. This reversal in the pattern of perceptual boundaries provides evidence which supports the hypothesis that listeners' perception of /pa/

and /ba/ was not influenced by a stereotype about accented speech in general, where non-native speakers are a general group. This opposing pattern of the perceptual boundaries for the Spanish and Chinese talkers in relation to the perceptual boundary for the native talker concurs with the results of McCullough & Clopper (2016) where Mandarin L1 speakers of English were rarely grouped with Spanish L1 speakers of English in a free association task. The present study shows that these different groups of non-native speakers are not perceived as being similar through an indirect task.

While the VOT in *please* and *pick* in the frame sentence are similar across talkers, as are the VOT of word medial /b/ and the pitch, the speech rate and F0 of the vowel in the talkers' *please* differs across talkers. This offers a few possible explanations for the difference in perceptual boundaries for the different talker conditions. The frame sentence produced by the non-native talker is 0.7 words per second slower than the native talker. According to Miller et al. (1997), a slower speech rate is associated with longer VOT. Listeners compensate for this difference, so that they tend to to have a perceptual boundary at a longer VOT duration for talkers who use a slower speech rate. If the listeners in Experiment Three use speech rate as a cue, they should have a perceptual boundary at a longer VOT duration for the non-native talker compared to the native talker, which is observed in the findings. Further, the F0 of the vowel in the non-native talker's *please* is higher than that of the native talker's, possibly indicating to the listeners that the non-native talker produces /p/ with longer VOT than the native talker, leading them to form a representation where the perceptual boundary for the non-native talker is at a longer VOT than that of the native talker. While the results of Experiment Three are in line with both these interpretations, the present study is not set up to probe which acoustic cues besides

VOT is responsible for cueing a difference in perceptual boundaries between native and non-native talkers.

The Experiment One and Two comparison presented in Section 4.4 appears to suggest holding the VOT cues of the frame sentence constant will induce a ‘shift’ in the perceptual boundary of the non-native talker. However, the comparison of Experiments Two and Three appear to suggest that the perceptual boundary for a native talker can also be adjusted. The perceptual boundary for the native talker appears to ‘shift’ when comparing the results of Experiments Two and Three, while the perceptual boundaries for the non-native talkers were relatively similar in those experiments. Another possible factor which might have an effect on perceptual boundary and was not controlled for in all the experiments is speech rate. As briefly discussed in Section 3.4, a talker with a lower speech rate would be perceived as having a perceptual boundary at an increased VOT value (Miller et al., 1986). Such an explanation could explain why the perceptual boundary for native talker is at a shorter VOT in Experiment Three since the Chinese talker had a slower speech rate. However, this speech rate explanation is not supported by the findings of Experiments One and Two where the Spanish talker also has a lower speech rate than the native talker yet the perceptual boundary for the Spanish talker is at a shorter VOT. Nevertheless, this direct comparison of different experiments should be taken with a grain of salt as the experiments were conducted with different sets of participants. Future research comparing explicitly the effect of multiple non-native talkers with different accents would be worthwhile.

Altogether, the results from Experiment Three provide evidence that there is an effect of whether a talker is a native or non-native speaker on a listener’s perceptual boundary between /ba/ and /pa/ even when the non-native talker realise plosives similarly to the native talker. This difference in perceptual boundaries patterns differently to that seen for the Spanish L1 talker and the native talker, providing evidence that listeners do not treat non-native talkers as a general group.

Chapter 6 General Discussion

This chapter aims to summarise the results of Experiments One, Two and Three and discuss their theoretical implications. I begin by providing a summary of the individual experiments (Section 6.1). The implications of these findings are discussed in Section 6.2 This is followed by a section on the limitations of the study (Section 6.3). The chapter is then concluded by an overall summary (Section 6.4).

6.1 Summary of Experiment Results

Experiment One compared responses towards a syllable continuum spliced into unmodified frame sentences produced by a native speaker of American English and a native speaker of Spanish from Mexico. Experiment Two retained the same set of talkers but held the VOT of the word initial /p/ in both frame sentences constant. Experiment Three introduced a new non-native talker, a Mandarin native speaker from China, in place of the Spanish talker. Both the VOT of the word initial /p/ in native talker and non-native talker frame sentences were also held constant. Table 6.1 summarises the results of all three experiments.

Table 6.1: Summary of results across experiments.

	1	2	3
Different perceptual boundaries for native and non-native talkers	Yes	Yes	Yes
Perceptual boundary for non-native talker is at a shorter VOT duration	Yes	Yes	No
Perceptual bounday for native talker is at a shorter VOT duration	No	No	Yes
Significant effect of Block	Yes	Yes	Yes

Experiment One established that listeners have different perceptual boundaries for an identical /pa/-/ba/ continuum when those syllables were inserted into frame sentences produced by two different talkers: one who is native speaker of American

English and the other who speaks Mexican Spanish-accented English. It was hypothesised that listeners will have a perceptual boundary between /p/ and /b/ at a shorter VOT duration for the Mexican Spanish talker who naturally produces relatively unaspirated voiceless plosives (shorter VOT) compared to the American English talker who naturally produces aspirated voiceless plosives (longer VOT). As hypothesised, listeners have a crossover point at a step with a shorter VOT duration for the non-native talker. Overall, the listeners also perceived more syllables as /pa/ when inserted into the non-native frame sentence.

As both talkers' frame sentences in Experiment One contained VOT cues signaling the talkers' aspiration patterns, Experiment Two was conducted to determine if this difference in perceptual boundary is influenced by the listener incorporating talker accentedness or by the VOT cues in the frame sentences. The native talker's frame sentence used in Experiment One was modified for Experiment Two by shortening the VOT duration of word initial /p/ to the same duration as those present in the non-native talker's frame sentence. Experiment Two was then run with a new set of participants. Its results confirmed that listeners still have different perceptual boundaries for both talkers even when the VOT cues in their frame sentences are of identical duration, although the difference in perceptual boundaries appears to have reduced. This reduction appears to be a consequence of the perceptual boundary for the non-native talker being shifted to a longer VOT duration. Overall, the findings of Experiment Two suggest that talker accentedness is sufficient to influence a listener's perceptual boundary.

Next, Experiment Three tested whether this talker accentedness effect is limited to the particular accent used in Experiments One and Two by replacing the non-native frame sentence spoken by a talker with a different accent who naturally produces aspirated voiceless plosives (longer VOT) like the native talker. Experiment Three was then run with a different set of participants from Experiments One and Two. Despite both talkers' frame sentences having VOT cues of equal duration, the results of Experiment Three confirmed that listeners have different perceptual boundaries for a non-native talker of a different accent. However, in contrast to Experiments One and Two, the listeners' crossover point for the non-native talker was at a longer VOT duration than that of the native talker and listeners also perceived more syllables as /pa/ when inserted into the native frame sentence. This difference between Experiments One and Two versus Experiment Three is evidence that non-native talkers who speak different native languages are not perceived as a single general non-native accented group.

The difference in perceptual boundaries between native and non-native talkers throughout all the experiments can seemingly be explained by the listener forming a representation of the talkers' perceptual boundaries through other acoustic cues besides word initial VOT, such as speech rate, pitch and VOT of word medial /b/, especially for Experiments Two and Three. However, the stimuli that the listener is responding to, the steps in the continuum, only differ in terms of VOT. A representation constructed solely from acoustic cues in the frame sentence would not be able to account for differences in the listeners' responses across talkers in Experiments Two and Threee, as such a representation would not have the necessary acoustic information to evaluate the steps in the continuum. Instead, it is possible that those other acoustic cues serve to activate a representation or prototype of the talker's bilabial plosives in the listener's mind, which contain some information about what kind of VOT the talker would produce for /b/ or /p/. This VOT information is then evaluated against the steps in the continuum. In effect, these other acoustic cues are characteristic of the talker's accent group and help to activate talker information.

6.1.1 Effect of Block

Outside of the main results arising from the factors manipulated, there is another common trend across all the experiments. All mixed effects models (Table 3 . 3, Table 4 .6, and Table 5 .10) fitted to the responses in the three experiments have shown a significant effect of Block. Specifically there are fewer /pa/ responses as the experiments progressed through the blocks. This is unlikely to be an adaptation effect as adaptation is commonly found to occur within very few items in the relatively simple task used in this study (Clark & Garrett, 2004). Further, figures (Figure 3.7, Figure 4.3, Figure 5.2) which separated the responses according to block also did not reveal a discernable systematic pattern. A possible explanation for this is that listeners start out tending to select more items as /pa/ and subsequently compensate by selecting more /ba/ as the experiment progresses. As the keys corresponding to each choice is not switched during the experiment, listeners might feel the need to select the other key to balance out their selection. This compensatory selection might be less systematic compared to genuine adaptation and less likely to be observable visually in graphs.

6.2 Implications

The experiments reported in this dissertation exploring the effects of talker accentedness as a social prime is linked with research done on the effects of social primes on category perception. The results are in line with recent work demonstrating that talker-based information affects the perceptions of sounds (Sections 2.1.2 and

2.1.3). It establishes that similar to other types of talker-related information such as gender, age, and dialect, a talker's accentedness affects listeners' perceptual boundaries of speech sounds. However the contrasting patterns of differences in perceptual boundaries seen in Experiments One and Two versus Experiment Three indicates that non-native talker accent is not treated as a singular type. This is in line with the findings of recent research conducted on non-native accent classification (Atagi & Bent, 2013; McCullough & Clopper, 2016). In free classification tasks where listeners were not limited by provided accent labels, Mandarin native speakers and Spanish native speakers were rarely grouped together, often at a rate of less than 20% (McCullough & Clopper, 2016). Thus, listeners are able to differentiate non-native accents well enough to activate representations of different talkers in memory, even from only one frame sentence.

The effect of using cues to activate information about an individual talker's speech sounds is generally explained using experience-based models such as the exemplar model (Lacerda, 1997; Pierrehumbert, 2001). In an exemplar model, listeners store the individual experiences they encounter. These tokens are stored together with a cloud of other similar tokens which represent a category. The category which is activated forms a representation or recalls a prototype of how a member of that category produces a speech sound. A hypothetical process of perception of the syllables used in the present study is as follows: listeners hear the frame sentence which contains cues which activates categories for native and non-native speech. The tokens indexed to these categories are used to form representations of each talker's speech sounds, which is reflected in the listeners' responses to the experimental items. As the listeners recruited for the study are monolingual native speakers, a listener's cloud of tokens representing native speakers will likely be much larger and more consistent than the cloud representing non-native speakers. The unusual VOT duration of /p/ in the native talker in Experiment Two might not be salient enough for listeners to adjust their representation of /b/-/p/, leading to the lack of difference between the native speaker perceptual boundaries in Experiments One and Two. In contrast, the cloud representing non-native speakers might have more variation between listeners, as the monolingual listeners would vary more in their experience of non-native accented speech. This would be a possible explanation for the lack of a consistent pattern across the participants within each experiment (Figure 3.8,Figure 4.4 and Figure 5.3). The difference in responses most listeners have towards the talkers within each experiment also indicate that the listeners do not just default towards using one representation for all talkers, but actually use the acoustic cues available to activate a specific representation for each talker.

The difference in the perception of non-native and native speech at the sound level offers a few potential explanations for the findings of previous research on talker accentedness. Where non-native accents were found to negatively affect tasks like lexical access, sentence processing and overall comprehension (Section 2.1.4), it is possible that the discrepancy in perception between native and non-native speech at the speech sound level is a contributing factor. Firstly, most listeners appear to form different perceptual boundaries for native and non-native talkers, even when the non-native talker is a native speaker of a language with similar plosive realisations to American English. This might result in the formation of inaccurate category representations for the non-native talker, which might hinder or delay lexical access and comprehension. Secondly, even when the correct VOT cues are available in the frame sentence (eg., Experiment One), listeners might still not form accurate category boundaries between voiced and voiceless plosives. The crossover point of /b/-/p/ in Experiment One was at a VOT duration of about 20ms (Figure 3.6) but the non-native talker realises some of her word initial /p/ with shorter VOT duration (eg., /p/ of *pick* in the frame sentence has a VOT of 14ms). Hence, it is possible that listeners might misperceive some plosives as their voicing counterparts (eg., intended /p/ word as /b/ word), leading to listeners mishearing words. While high-frequency words are less likely to be misheard, a mismatch between phonological representation and actual speech arising from a non-native accent has been shown to negatively impact word recognition of lower-frequency words ((Imai, Walley, & Flege, 2005)).

The findings of this dissertation also add nuance to recent research on how native speakers process non-native speech. These studies suggest that listeners process the linguistic input of non-native speakers in less detail, instead deferring to reliable contextual information such as the situation where the interaction is taking place (Lev-Ari, 2015a, 2015b; Lev-Ari et al., 2018; Lev-Ari & Keysar, 2012). This was because non-native speakers are more likely to make incorrect lexical choices and listeners have to reconstruct the non-native speakers' intended meanings from context. This dissertation does not specifically investigate listeners' perception of non-native speech at the lexical level. However, the different native-non-native perceptual boundaries seen in Experiments One and Two versus Experiment Three do indicate that listeners do pay sufficient attention to the acoustics of non-native speech to form different perceptual boundaries for different non-native talkers. This is by no means a counter to the claims of that series of research, merely an indication that the perception of non-native speech is a field which requires more research.

6.3 Limitations

While the results here have provided some insights into how listeners process native and non-native speech, there are a few limitations to the Experiments conducted. Firstly, the syllables tested only consisted of the /b/-/p/ continuum. It is possible that the perception of other plosives might be affected differently by non-native speech. Secondly, the present study only focuses on monolingual native speakers as a listener group. Using native speakers who have more experience with other languages or non-native accents will allow the investigation of listener experience as a factor in the perception of non-native speech. While some data was collected from non-monolingual native speakers over the course of this study, that data was not analysed for this dissertation as these participants differed significantly in terms of their language background. Thus, I did not obtain enough data from participants with similar enough language backgrounds to make a meaningful analysis. Thirdly, only one frame sentence from one non-native talker was used in each experiment. More talkers from the same non-native accent will provide stronger evidence that listeners are accessing a non-native speech category for that group of talkers, instead of using a representation of a particular talker. For example, the talker's accent might remind a listener of their friend, instead of activating a representation for a non-native speaker group.

6.4 Overall Summary

In sum, Experiment One provides evidence that listeners have different perceptual boundaries for /pa/-/ba/ syllables when inserted into frame sentences produced by talkers of different accentedness. Experiment Two then controlled the VOT cues in each talker's frame sentence to determine that talker accentedness can cause a difference in perceptual boundaries between those talkers, albeit at a smaller magnitude than in Experiment One. A cross-experiment analysis of Experiments One and Two suggests that the reduced difference in perceptual boundary is due to listeners using the shortened VOT cues in the native talker's frame sentence to construct a perceptual boundary for the non-native talker, instead of directly revising their perceptual boundary of the native talker. Experiment Three found that listeners still have a difference in perceptual boundary for native and non-native talkers even when the frame sentence is produced by a non-native talker who has similar VOT production patterns to the native talker. Nevertheless, the pattern of difference was reversed from the previous experiments, indicating that listeners do not perceive non-native talkers of different accents as a single general group. I argue that this reversal is due to the acoustic cues present in the non-native talker's frame sentence in Experiment Three caused listeners to activate a different representation of a talker.

Chapter 7 Conclusion

The goal of this dissertation was to investigate the effects of a non-native accent on listeners' perception of speech sounds. The work reported in this dissertation has provided some insight into this aspect of speech perception, while also revealing more areas to uncover. In Chapter 1, I introduced the goals and rationale of the study, and outlined the structure of the dissertation. Chapter 2 provided a review of the literature covering an overview of important research conducted on the role of talker-related information in the process of speech perception. Chapter 3 introduced the experimental methodology used in the study and reported the results of Experiment One which tested whether there are differences in how listeners perceive identical syllables which are spliced into frame sentences spoken by talkers with different accents. Chapter 4 reported the findings of Experiment Two which investigated whether listeners incorporate specific voice onset time (VOT) cues embedded in the frame sentence. Chapter 5 then reported the findings of Experiment Three which determined if effects found in Experiments One and Two can be observed with a different non-native accent. Chapter 6 consisted of an overall discussion which discussed the implications of all 3 experiments as a whole, provided a theoretical explanation for the findings obtained, and situated this study within the field of non-native speech perception.

The conclusions of this dissertation are as follows: Listeners perceive speech sounds differently according to the nativeness of the talker. Even when the critical syllables were produced by the same native talker, a frame sentence produced by different talkers is a sufficient cue for talker nativeness. This difference in perception is not solely due to VOT cues provided in the frame sentences. However, there is a possibility that other acoustic cues were used by listeners to form a perceptual representation of non-native speech sounds. Nevertheless, listeners do not treat all non-native talkers as a single general category; relative to the native talker, listeners were more likely to perceive the Spanish talker as producing more /pa/, while listeners were more likely to perceive the Mandarin talker as producing more /ba/.

Appendix A: Language Background Questionnaire

Language Background Questionnaire

In this questionnaire, we ask for information about the languages you know. We also ask for some basic information about yourself, including your contact information (email, phone). Please be assured that this information will not be shared with anybody outside the Language Analysis and Experimentation Laboratories (LAE Labs) at the University of Hawai'i at Manoa. If you have any questions while filling in the questionnaire, please feel free to ask the experimenter.

* Required

Personal Information

1. **First Name** *

2. **Last Name** *

3. **Sex** *
Mark only one oval.
- ◯ Female
- ◯ Male

4. **Month of Birth** *
Mark only one oval.
- ◯ 1
- ◯ 2
- ◯ 3
- ◯ 4
- ◯ 5
- ◯ 6
- ◯ 7
- ◯ 8
- ◯ 9
- ◯ 10

5. **Year of Birth** *
Mark only one oval.

- 1998
- 1997
- 1996
- 1995
- 1994
- 1993
- 1992
- 1991
- 1990
- 1989
- 1988
- 1987
- 1986
- 1985
- 1984
- 1983
- 1982
- 1981
- 1980
- 1979
- 1978
- 1977
- 1976
- 1975
- 1974
- 1973
- 1972
- 1971
- 1970
- 1969
- 1968
- 1967
- 1966
- 1965
- 1964

- 1963
- 1962
- 1961
- 1960
- 1959
- 1958
- 1957
- 1956
- 1955
- 1954
- 1953
- 1952
- 1951
- 1950
- Before 1950

6. **Place of Birth** *
   (city, state, country)

7. **Where did you grow up? Which neighbourhood? Where did you spend the most amount of time?** *

8. **Country where you currently reside? Did you live in another country? For how long?** *

9. **Current Occupation**

10. **(For local residents) Do you speak Pidgin? If you don't, do you**

11. **Highest level of schooling completed** *
Mark only one oval.

- Less than High School
- Graduated High School
- Professional training/2-year college
- 4-year college
- Graduate School

12. **E-mail address** *

13. **Telephone number**
(xxx-xxx-xxxx)

Personal Information (continued)

14. **Is your hearing normal?** *
Mark only one oval.

- Yes
- No

15. **May we contact you in the future to ask for your participation in other language research studies?** *
Mark only one oval.

- Yes
- No

Your Language History

(If any of your answers are English, please indicate the variety or varieties of English; e.g. 'American English', 'Australian English', 'Hawaii Creole English/Pidgin', etc.). If you are giving multiple answers, please separate them with commas (e.g. 'American English, Pidgin, Chinese')

16. **What was the first language you learned as a child?** *

17. **What language do you feel most comfortable speaking in casual conversation now?** *

18. **What language(s) was/were spoken in your home when you were growing up? ***

19. **What is your mother's native language? ***

20. **What language(s) did your mother use with you when you were a child? ***

21. **What is your father's native language? ***

22. **What language(s) did your father use with you when you were a child? ***

Your Language History (continued)

Please answer the following questions about your abilities with English and other

23. **A) Speaking English** *

Mark only one oval.

- ( ) 0
- ( ) 1
- ( ) 2
- ( ) 3
- ( ) 4
- ( ) 5
- ( ) 6
- ( ) 7
- ( ) 8
- ( ) 9
- ( ) 10

24. **B) Understanding spoken English** *

Mark only one oval.

- ( ) 0
- ( ) 1
- ( ) 2
- ( ) 3
- ( ) 4
- ( ) 5

25. **C) Reading English** *

Mark only one oval.

- ( ) 0
- ( ) 1
- ( ) 2
- ( ) 3
- ( ) 4
- ( ) 5
- ( ) 6
- ( ) 7
- ( ) 8
- ( ) 9
- ( ) 10

26. **D) Writing English** *

Mark only one oval.

- ( ) 0
- ( ) 1
- ( ) 2
- ( ) 3
- ( ) 4
- ( ) 5

27. **E) Overall English Proficiency** *
Mark only one oval.

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10

Your Language History (continued)

Please answer the following questions about your language abilities

Language Proficiency (other than English)

Please list all the languages you know, other than English, below. For each language, please indicate your proficiency on a scale from zero to ten (zero being worst, ten being best)

28. **Language 1**

29. **Proficiency in Language 1**
Mark only one oval.

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8

30. **Language 2**

31. **Proficiency in Language 2**
Mark only one oval.

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10

32. **Language 3**

33. **Proficiency in Language 3**
Mark only one oval.

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8

34. **Proficiency in Language 4**
Mark only one oval.

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10

35. **Language 4**

____

36. **Language 5**

____

37. **Proficiency in Language 5**
Mark only one oval.

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7

Appendix B: Background of Monolingual Native Listeners

Experiment One:

	Gender	Age	Childhood hometown (city or region)	Perceived L1 of non-native talker
2	Female	25	Stony Brook, New York	Korean
4	Male	20	Western Washington state	Not sure (Asian)
5	Female	19	Kailua (Oahu) Hawaiʻi	Not sure (Filipino or English)
6	Male	24	Nampa, Idaho	English
7	Female	23	Kaneohe, Hawaiʻi	Filipino
14	Female	20	Nuʻuanu, Hawaiʻi	Not sure (Asian language)
18	Male	35	Berkeley, California; Big Island, Hawaiʻi	Filipino or Pacific Islander
20	Male	21	Salt Lake, Hawaiʻi	Chinese
24	Male	21	Hudson River Valley, New York	Tagalog
29	Male	20	Makiki, Hawaiʻi	Spanish
31	Female	18	Mililani, Hawaiʻi	Not sure (Southeast Asian)
33	Male	21	Waipio Gentry, Hawaiʻi	A Filipino language
36	Female	21	Orange County, California	Spanish
37	Female	23	Sierra Madre, California	Spanish
38	Female	22	Oceanside, California; Durango, Colorado	Spanish
43	Male	23	Pearl City, Hawaiʻi	Indian
45	Female	20	Santa Barbara; Ventura, California	French or Indian
46	Male	19	Pacifica, California	Not sure (Asian)
47	Female	25	Redding, California; Missouri	Not sure (probably Vietnamese or Southeast Asian)
48	Female	28	Pleasanton, California; Portland, Oregon	Spanish
49	Female	24	Kaneohe, Hawaiʻi	Filipino or Ilocano
50	Female	30	Hawaiʻi Kai, Hawaiʻi	Spanish
51	Female	22	Bend, Oregon	Spanish
52	Male	20	Silicon Valley, California	Arabic
53	Male	19	Makiki, Hawaiʻi	Spanish

Experiment Two:

	Gender	Age	Childhood hometown (city or region)	Perceived L1 of non-native talker
5	Female	19	Salt Lake, Hawaiʻi	Russian
8	Male	21	Salt Lake, Hawaiʻi	Chinese
9	Female	21	Mililani, Hawaiʻi	Micronesian
13	Female	19	Waipio, Hawaiʻi	Filipino
16	Female	22	Wahiawa, Hawaiʻi	Spanish
17	Male	19	Nuʻuanu, Hawaiʻi	Filipino

19	Male	20	Hilo, Hawaiʻi	Spanish
20	Female	19	Nuʻuanu, Hawaiʻi	Not sure (European)
21	Female	21	Pukalani, Hawaiʻi	Vietnamese
22	Female	21	Kaimuki, Hawaiʻi	Korean
23	Female	23	Kahala, Hawaiʻi	Filipino
24	Male	18	Mililani, Hawaiʻi	Hispanic language
26	Female	20	Kaimuki, Hawaiʻi	Not English (Micronesian)
29	Male	19	Vancouver, Washington	Indian language
30	Male	30	Olympia, Washington	Spanish
32	Male	22	Hawaiʻi Kai, Hawaiʻi	Spanish
33	Male	21	Wailuku; Hilo, Hawaiʻi	Spanish
35	Male	18	Kahala, Hawaiʻi	Filipino
36	Male	24	ʻEleʻele, Hawaiʻi	Indian language
37	Female	19	Aina Haina, Hawaiʻi	Not sure (maybe English)
40	Female	20	Nuʻuanu, Hawaiʻi	Spanish
41	Female	19	Moanalua	Indian language
44	Male	19	Bay Area, California; Lanai, Hawaiʻi	Tagalog
49	Female	19	Kaimuki, Hawaiʻi	Filipino or Spanish
51	Female	20	Kailua, Oʻahu, Hawaiʻi	Chinese

Experiment Three:

	Gender	Age	Childhood hometown (city or region)	Perceived L1 of non-native talker
1	Male	20	Pearl City; Hawaiʻi Kai, Hawaiʻi	Melanesian (or Southeast Asian)
3	Female	19	Hanalei, Hawaiʻi	French
5	Female	21	El Segundo, California	Tagalog
10	Male	20	Kapolei, Hawaiʻi	Indian language
11	Female	20	Nuʻuanu, Hawaiʻi	Japanese
12	Female	22	Aiea, Hawaiʻi	Not sure (Asian language)
14	Male	20	Anchorage, Alaska; Cape Coral, Florida	Portuguese
15	Female	18	Baltimore, Maryland	Pacific Islander or Asian language
17	Female	18	Atlanta, Georgia; St. Petersburg, Florida	Indonesian
19	Female	20	Puna, Hawaiʻi	Japanese
21	Female	18	San Jose, California	Japanese
22	Male	23	Miami, Florida	Tamil
24	Female	19	Escondido, California	Indian language
26	Male	30	Honolulu, Hawaiʻi; Seattle, Washington	Chinese
27	Male	20	Kaimuki, Hawaiʻi	Filipino

32	Female	21	Tamuning, Guam	Filipino
33	Male	20	Wailuku, Hawaiʻi	Filipino
35	Male	21	Hawaiʻi Kai; Kahala, Hawaiʻi	Vietnamese
36	Male	25	Kaneohe, Hawaiʻi	Tagalog
37	Female	20	Columbia, South Carolina	English (stronger accent)
40	Female	21	San Francisco, California	Indian language
43	Female	18	Grand Junction, Colorado	Asian language
44	Female	20	Kaneohe, Hawaiʻi	Indian language
45	Female	22	Hawaiʻi Kai, Hawaiʻi	Chinese
46	Male	19	Hilo, Hawaiʻi	Chinese

References

Arslan, L. M., & Hansen, J. H. L. (1997). A study of temporal features and frequency characteristics in American English foreign accent. *The Journal of the Acoustical Society of America*, *102*(1), 28–40. https://doi.org/10.1121/1.419608

Atagi, E., & Bent, T. (2013). Auditory free classification of nonnative speech. *Journal of Phonetics*, *41*(6), 509–519. https://doi.org/10.1016/j.wocn.2013.09.003

Babel, M., & McGuire, G. (2013). Listener Expectations and Gender Bias in Nonsibilant Fricative Perception. *Phonetica*, *70*(1–2), 117–151. https://doi.org/10.1159/000354644

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, *114*(3), 1600–1610. https://doi.org/10.1121/1.1603234

Bradlow, A. R., & Bent, T. (2008). Perceptual Adaptation to Non-Native Speech. *Cognition*, *106*(2), 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

Campbell-Kibler, K. (2009). The nature of sociolinguistic perception. *Language Variation and Change*, *21*(01), 135. https://doi.org/10.1017/S0954394509000052

Casasanto, L. S. (2008). Does Social Information Influence Sentence Processing? *Proceedings of the Annual Meeting of the Cognitive Science Society*, *30*(30), 7.

Chen, L., Chao, K.-Y., & Peng, J.-F. (2007). VOT productions of word-initial stops in Mandarin and English: A cross-language study. *ROCLING 2007 Poster Papers*, 303–317.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647–3658. https://doi.org/10.1121/1.1815131

Clopper, C. G., & Pisoni, D. B. (2004). Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change*, *16*(01). https://doi.org/10.1017/S0954394504161036

Clopper, C. G., & Pisoni, D. B. (2006). Effects of region of origin and geographic mobility on perceptual dialect categorization. *Language Variation and Change*, *18*(02). https://doi.org/10.1017/S0954394506060091

Derwing, T. M., & Munro, M. J. (1997). ACCENT, INTELLIGIBILITY, AND COMPREHENSIBILITY. *Studies in Second Language Acquisition*, *19*(01). https://doi.org/10.1017/S0272263197001010

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, *42*(04), 476. https://doi.org/10.1017/S026144480800551X

Drager, K. (2011). Speaker Age and Vowel Perception. *Language and Speech*, *54*(1), 99–121. https://doi.org/10.1177/0023830910388017

Eckert, P. (2008). Variation and the indexical field1: VARIATION AND THE INDEXICAL FIELD. *Journal of Sociolinguistics*, *12*(4), 453–476. https://doi.org/10.1111/j.1467-9841.2008.00374.x

Flege, J. E., & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *The Journal of the Acoustical Society of America*, *83*(2), 729–740. https://doi.org/10.1121/1.396115

Floccia, C., Butler, J., Goslin, J., & Ellis, L. (2009). Regional and Foreign Accent Processing in English: Can Listeners Adapt? *Journal of Psycholinguistic Research*, *38*(4), 379–412. https://doi.org/10.1007/s10936-008-9097-8

Foulkes, P. (2010). Exploring social-indexical knowledge: A long past but a short history. *Laboratory Phonology*, *1*(1). https://doi.org/10.1515/labphon.2010.003

Hanulíková, A. (2018). The effect of perceived ethnicity on spoken text comprehension under clear and adverse listening conditions. *Linguistics Vanguard*, *4*(1). https://doi.org/10.1515/lingvan-2017-0029

Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, *48*(4). https://doi.org/10.1515/ling.2010.027

Hay, J., Drager, K., & Gibson, A. (2018). *HEARING R-SANDHI: THE ROLE OF PAST EXPERIENCE*. *94*(2), 45.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*(4), 458–484. https://doi.org/10.1016/j.wocn.2005.10.001

Hayes-Harb, R., Smith, B. L., Bent, T., & Bradlow, A. R. (2008). The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. *Journal of Phonetics*, *36*(4), 664–679. https://doi.org/10.1016/j.wocn.2008.04.002

Imai, S., Walley, A. C., & Flege, J. E. (2005). Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners. *The Journal of the Acoustical Society of America*, *117*(2), 896–907. https://doi.org/10.1121/1.1823291

Johnson, K. (2005). Speaker Normalization in Speech Perception. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 363–389). https://doi.org/10.1002/9780470757024.ch15

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics*, *27*(4), 359–384. https://doi.org/10.1006/jpho.1999.0100

Kang, O., & Rubin, D. L. (2009). Reverse Linguistic Stereotyping: Measuring the Effect of Listener Expectations on Speech Evaluation. *Journal of Language and Social Psychology*, *28*(4), 441–456. https://doi.org/10.1177/0261927X09341950

Kim, J. (2016). Perceptual Associations Between Words and Speaker Age. *Laboratory Phonology*, *7*(1), 18. https://doi.org/10.5334/labphon.33

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, *107*(1), 54–81. https://doi.org/10.1016/j.cognition.2007.07.013

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review, 13*(2), 262–268. https://doi.org/10.3758/BF03193841

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2007). *First Impressions and Last Resorts: How Listeners Adjust to Speaker Variability: (527342012-194)* [Data set]. https://doi.org/10.1037/e527342012-194

Lev-Ari, S. (2015a). Adjusting the Manner of Language Processing to the Social Context: Attention Allocation During Interactions with Non-native Speakers. In R. K. Mishra, N. Srinivasan, & F. Huettig (Eds.), *Attention and Vision in Language Processing* (pp. 185–195). https://doi.org/10.1007/978-81-322-2443-3_11

Lev-Ari, S. (2015b). Comprehending non-native speakers: theory and evidence for adjustment in manner of processing. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01546

Lev-Ari, S., Ho, E., & Keysar, B. (2018). The Unforeseen Consequences of Interacting With Non-Native Speakers. *Topics in Cognitive Science*. https://doi.org/10.1111/tops.12325

Lev-Ari, S., & Keysar, B. (2012). Less-Detailed Representation of Non-Native Language: Why Non-Native Speakers' Stories Seem More Vague. *Discourse Processes, 49*(7), 523–538. https://doi.org/10.1080/0163853X.2012.698493

Lisker, L., & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. WORD, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Magen, H. S. (1998). The perception of foreign-accented speech. *Journal of Phonetics*, *26*(4), 381–400.

McCullough, E. A., & Clopper, C. G. (2016). Perceptual subcategories within non-native English. *Journal of Phonetics*, *55*, 19–37. https://doi.org/10.1016/j.wocn.2015.11.002

McGowan, K. B. (2011). *The Role of Socioindexical Expectation in Speech Perception*. University of Michigan.

McGowan, K. B. (2015). Social expectation improves speech perception in noise. *Language and Speech*, *58*(4), 502–521.

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1609–1631. https://doi.org/10.1037/a0011747

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review*, *15*(6), 1064–1071. https://doi.org/10.3758/PBR.15.6.1064

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33–B42.

Miller, J. L., O'Rourke, T. B., & Volaitis, L. E. (1997). Internal structure of phonetic categories: Effects of speaking rate. *Phonetica 54*, *54*(3–4), 121–137.

Munson, B., Jefferson, S. V., & McDonald, E. C. (2006). The influence of perceived sexual orientation on fricative identification. *The Journal of the Acoustical Society of America*, *119*(4), 2427–2437. https://doi.org/10.1121/1.2173521

Niedzielski, N. (1999). The Effect of Social Information on the Perception of Sociolinguistic Variables. *Journal of Language and Social Psychology*, *18*(1), 62–85. https://doi.org/10.1177/0261927X99018001005

Norris, D., James M., M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238. https://doi.org/10.1016/S0010-0285(03)00006-9

Oh, J. S., Au, T. K.-F., & Jun, S.-A. (2010). Early childhood language memory in the speech perception of international adoptees. *Journal of Child Language*, *37*(05), 1123–1132. https://doi.org/10.1017/S0305000909990286

Oh, J. S., Jun, S.-A., Knightly, L. M., & Au, T. K. (2003). Holding on to childhood language memory. *Cognition*, *86*(3), 12.

Podesva, R. J., Reynolds, J., Callier, P., & Baptiste, J. (2015). Constraints on the social meaning of released /t/: A production and perception study of U.S. politicians. *Language Variation and Change*, *27*(01), 59–87. https://doi.org/10.1017/S0954394514000192

Porretta, V., Tremblay, A., & Bolger, P. (2017). Got experience? PMN amplitudes to foreign-accented speech modulated by listener experience. *Journal of Neurolinguistics*, *44*, 54–67. https://doi.org/10.1016/j.jneuroling.2017.03.002

Porretta, V., Tucker, B. V., & Järvikivi, J. (2016). The influence of gradient foreign accentedness and listener experience on word recognition. *Journal of Phonetics*, *58*, 1–21. https://doi.org/10.1016/j.wocn.2016.05.006

Romero-Rivas, C. (2016). *The effects of foreign-accented speech on language comprehension and retrieval processes* (Universitat Pompeu Fabra). Retrieved from https://www.tdx.cat/bitstream/handle/10803/399504/tcrr.pdf?sequence=1

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, *33*(4), 511–531. https://doi.org/10.1007/BF00973770

Rubin, D. L., & Smith, K. A. (1990). Effects of accent, ethnicity, and lecture topic on undergraduates' perceptions of nonnative English-speaking teaching assistants. *International Journal of Intercultural Relations*, *14*(3), 337–353. https://doi.org/10.1016/0147-1767(90)90019-S

Strand, E. A. (1999). Uncovering the Role of Gender Stereotypes in Speech Perception. *Journal of Language and Social Psychology*, *18*(1), 86–100. https://doi.org/10.1177/0261927X99018001006

Strand, E. A., & Johnson, K. (1996). 2. Gradient and Visual Speaker Normalization in the Perception of Fricatives. In D. Gibbon (Ed.), *Natural Language Processing and Speech Technology* (pp. 14–26). https://doi.org/10.1515/9783110821895-003

Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition*, *119*(1), 131–136. https://doi.org/10.1016/j.cognition.2010.10.018

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *38*(4), 579–590. https://doi.org/10.1037/h0080868

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00577

Vieru, B., & Boula, P. (2011). Characterisation and identification of non-native French accents. *Speech Communication, 53*(3), 292–310.

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1993). *F* 0 gives voicing information even with unambiguous voice onset times. *The Journal of the Acoustical Society of America*, *93*(4), 2152–2159. https://doi.org/10.1121/1.406678

Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2013). Roles of Voice Onset Time and F0 in Stop Consonant Voicing Perception: Effects of Masking Noise and Low-Pass Filtering. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1097–1107. https://doi.org/10.1044/1092-4388(2012/12-0086)

Yamaguchi, T., & Pḗtursson, M. (2012). Voiceless Stop Consonants in Malaysian English: Measuring the VOT Values. *Asian Englishes*, *15*(2), 60–79. https://doi.org/10.1080/13488678.2012.10801330