Perception and lexical processing of gradient foreign accentedness

by

Vincent Joseph Porretta

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics
University of Alberta

Examining committee:

Benjamin V. Tucker, Supervisor
Juhani Järvikivi, Supervisor
Antti Arppe, Supervisory Committee
Anne-Michelle Tessier, Examiner
Andrea Weber, External Examiner

Abstract

This dissertation examines native speaker perception and processing of the variability inherent to non-native speech, specifically, Mandarin-accented English. In order to accomplish this, subjective ratings were first collected as a measure of perceived foreign accentedness. The influence of linguistic variables (both acoustic and lexical) was investigated with regard to the perception of gradient foreign accentedness. The results of the ratings study indicate that the perception of gradient accentedness is influenced by measures of acoustic distance (i.e., magnitude of difference between a given production and an average native production) as well as properties of the lexical items themselves (e.g., neighborhood density and phonotactic probability). These ratings were then utilized to investigate how the gradient nature of accentedness influences lexical processing across behavioral, (visual word) eye-tracking, and electrophysiological methods. Additionally, it was also possible to examine how self-reported listener experience with Chinese-accented speakers influences the processing of gradient accentedness. These studies provide converging and complementary evidence that processing of these tokens varies non-linearly along the accentedness continuum and by level of listener experience. The electrophysiological results indicate that the pattern of processing, including the allocation of perceptual and attentional resources, changes as a result of increased exposure to Chinese-accented English. The effect of experience is also seen in both reaction time and visual world eye-tracking data. The results of a cross-modal priming study indicate that degree of foreign accent modulates the strength with which lexical representations are primed and that listener experience with the accent in question mitigates this effect. Visual world eye-tracking presents similar results, showing that the time-course of word recognition is slowed as accentedness increases, though the ability to decode the signal is enhanced for listeners with greater experience. Taken together, the results come to bear on our understanding of how gradient foreign-accented speech maps onto linguistic representations and how those representations may change and adapt over time to accommodate the variability inherent to foreign-accented speech.

Preface

The research projects contained within this dissertation received research ethics approval from the University of Alberta Research Ethics Board, Project Name "Perception and Processing of Speech Variability", No. Pro00027392, 13 March 2012. The research conducted for this thesis was done in collaboration with other researchers. The study contained in chapter 2 was carried out with assistance from Dr. P.A. Bolger, Dr. B.V. Tucker, and Dr. A.-J. Kyröläinen. I was responsible for experiment design, data collection, analysis, and manuscript composition. Dr. P.A. Bolger assisted with experiment design, Dr. A.-J. Kyröläinen assisted with concept formation and contributed to manuscript edits, and Dr. B.V. Tucker was the supervisory author, contributing to concept formation and manuscript edits. The study contained in chapter 3 was carried out with assistance from Dr. P.A. Bolger, Dr. R.H. Baayen, and Dr. A. Tremblay. I was responsible for experiment design, data collection, analysis, and manuscript composition. Dr. P.A. Bolger assisted with experiment design, while Dr. R.H. Baayen and Dr. A. Tremblay assisted with data analysis. The study contained in chapter 4 was carried out with assistance from Dr. B.V. Tucker and Dr. J. Järvikivi. I was responsible for experiment design, data collection, analysis, and manuscript composition. Dr. B.V. Tucker and Dr. J. Järvikivi were the supervisory authors, assisting with experiment design and contributing to concept formation and manuscript edits. No part of this thesis has been previously published.

Acknowledgements

I would like to express my appreciation and gratitude to all those who have supported me throughout my studies and during the undertaking of this doctoral dissertation. In particular, I thank my supervisors Dr. Benjamin V. Tucker and Dr. Juhani Järvikivi for their guidance, support, and mentorship. Thank you also to Dr. Antti Arppe, the third member of my supervisory committee, for his guidance and collegiality. I also wish to thank Dr. Patrick Bolger, my first supervisor at the University of Alberta. While it was not possible for him to remain at the university, working with him was a privilege and an important part of my academic formation.

Thank you to the administrative staff of the Department of Linguistics for their tireless support of the students. I also thank my fellow graduate students whose friendship and support have made the road easier.

I would remiss if I did not thank my wonderful family, especially my parents, who have always supported my endeavors, both personal and professional. I have always tried to make them proud. Their unconditional love and support mean the world to me. Lastly, I want to thank my partner. His love, support, guidance and patience are more than I could ever ask for.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Different speakers, even those of the same native language, produce speech sounds that vary, sometimes drastically, within and across individuals. Foreign-accented speech presents a unique case of this variability. We generally have the sense that some non-native speakers have stronger accents than others, and that it typically becomes easier to understand non-native speakers over time, with more interaction. While gradience in perceived foreign accentedness is generally acknowledged, it has not been investigated in regard to lexical processing. Perceived goodness-of-fit speaks to the dimensions along which speech is evaluated and this fit may affect processing in a graded manner. A gradient effect would underscore the role that variability plays in language processing at large. Similarly, little work has been done to investigate how experience with foreign-accented speech might influence processing of gradient foreign accentedness. Experience with the linguistic variability within a specific group of non-native speakers may ultimately change how listeners handle variability in the speech signal. So, foreign-accented speech provides a special environment which can enhance our understanding of how spoken language is processed by native listeners and what this may indicate about how language is represented in the mind.

Here, a multi-methodological approach is taken to address these points in an attempt to provide converging evidence regarding the nature of foreign accent strength and its effect on lexical processing. Specifically, the studies contained in this dissertation seek to address the following questions: 1) Which acoustic and lexical properties influence the perception of gradient foreign accentedness? 2) How does the gradience present in foreign-accented speech affect language processing in real-time? 3) Is there an influence of listener experience on foreign-accented speech processing? 4) What type of information might need to be represented in the mind in order to account for any effects of gradience and experience?

What follows is a review of relevant literature and methodologies. This chapter lays out the concept of gradient foreign accentedness and possible influences on the magnitude of degree of foreign accent. A survey of studies investigating the effect of foreign accented speech on processing is then provided. Additionally, the issue of speech variability and listener experience are presented in relation to lexical processing. The chapter concludes with a brief synopsis of the three papers contained in this dissertation which were designed to address specific aspects of the points mentioned above.

1.1 Foreign-accented speech

A foreign accent is generally thought to arise from the presence of speech patterns (at various levels) that do not align with those of native speakers. Existing literature has begun to explore this phenomenon and indicates a number of speaker-related factors which can influence both the presence as well as the degree of foreign accent. In a review by Piske, MacKay, and Flege (2001), variables such as age of second language (L2) learning, length of residence in an L2-speaking country, gender, formal instruction, motivation, language learning aptitude and amount of first language (L1) use have all been shown to influence a non-native speaker's accent. However, in addition to their review, Piske et al. provide a study showing that both age of L2 learning and amount of continued L1 use appear to have the greatest effect on degree of foreign accentedness.

Because foreign accent is often considered to be a failure to achieve native-like productions of speech sounds (or sound sequences), factors such as age of L2 learning, instruction, and native language usage may capture other underlying articulatory and phonological properties. Sounds are produced by manipulating vocal articulators and speakers with a foreign accent may lack practice in making the necessary articulatory gestures in place and timing (Flege, 1980). The phonology of a speaker's L1 can influence his ability to produce non-native phones by interfering with the speaker's ability to perceive and produce L2-specific phonetic detail. Sounds that are similar between the two languages appear to interfere with each other because the speaker may judge them to be acoustically different realizations of the same category (Flege & Hillenbrand, 1984). Additionally, the phonotactic rules of the speaker's L1 have been shown to affect accent in the L2; specifically, L2 sequences that align with the phonotactics of the speaker's L1 are judged to be less accented than those that do not (Park, 2013). It has been borne out in previous research that non-native productions indeed differ from native productions on a variety of acoustic measures, both temporal and spectral (Flege & Hillenbrand, 1984; Munro, 1993; Wayland, 1997; Baker et al., 2011). Specifically, word duration, vowel duration, voice-onset time and formant values have shown deviation in non-native productions. These deviations may then affect the perception of gradient foreign accentedness.

1.2 Perception of degree of foreign-accentedness

Native listeners have been shown to be sensitive to the acoustic differences between non-native and native productions and can detect the presence of an accent in as little as 30 ms of a burst release (Flege & Hillenbrand, 1984). Listeners appear to use some of this fine-grained acoustic information when determining degree of foreign accentedness. A number of studies have shown that accentedness ratings (for various L1-L2 combinations) can be predicted by many of the acoustic variables cited above (Munro, 1993; Wayland, 1997; Porretta & Tucker, 2012). Witteman, Weber, and McQueen (2013) state that vowels drive differences in perceived accentedness as they can vary significantly from standard forms. Munro (1993) showed this for L1 Arabic productions of English vowels. In that study, native English-speaking listeners rated vowel productions from native Arabic speakers and it was shown that first and second formant frequency measures were most predictive of those ratings. In a similar study, Wayland (1997) found that the ratings (from native Thai speakers) of native English-speaking productions of Thai words were driven by spectral rather than temporal measures; however the set of predictors varied for each of the different Thai tones. In work, preliminary to this dissertation, Porretta and Tucker (2012) found that global accentedness ratings of Chinese-accented English could be predicted by complex interactions between first and second formant values as well as between vowel duration and word duration. This indicates that, at least in the case of Chinese-accented English, the relationship between these measures can affect ratings of foreign accentedness.

All three of the aforementioned studies, employed acoustic distance measures for predicting accentedness ratings, with these distances taken in relation to one or more native speakers. Conceptually, the calculation of distances from a typical native production aligns with the observation that accent strength varies in magnitude across speakers. The concept of perceptual distance is not only related to non-native speech, but also dialectal variation (Floccia, Goslin, Girard, & Konopczynski, 2006; Goslin, Duffy, & Floccia, 2012). As a production becomes more deviant from what a listener considers to be native in his or her dialect, the perceived accent strength should increase in magnitude. However, productions can vary along multiple dimensions, and therefore, by calculating multiple distance measures for a given production, it is possible to begin to understand the type of multi-dimensional comparison that listeners carry out in relation to their stored lexical representations. Despite the findings that there are acoustic correlates to foreign accentedness, the majority of work on foreign-accented speech processing has largely ignored this multidimensional gradience, instead opting to simply dichotomize talkers into accented or non-accented groups.

While it seems clear that accentedness ratings are likely affected by the acoustic properties of the signal, it seems less clear that they may also be affected by properties of the words themselves. In psycholinguistic research, lexical variables have long been shown to affect behavioral decisions and response latencies. One study has looked specifically at the influence of speaker-independent factors on foreign accentedness ratings. Levi, Winters, and Pisoni (2007) examined the effect of word frequency on accentedness ratings assigned to English words produced by native German speakers. They found that words with higher lexical frequency resulted in lower accentedness ratings. A post hoc acoustic analysis of formant values for a subset of their tokens partially ruled out production as a source of the effect. This suggests that not only the properties of the acoustic signal produced by the speaker, but also the nature of items within the lexicon of the listener can affect judgments of accentedness. It is thus possible that lexical variables apart from frequency may also affect the perception of foreign accentedness and perhaps interact with the acoustics of the production. This variability in perceived foreign accentedness may then influence how foreign-accented speech is processed in real time.

1.3 Processing of foreign-accented speech

Judgments of degree of foreign accentedness necessarily entail the processing of the accented speech. However, these judgments do not constitute an online measure of processing. Online processing of accented speech can be examined using a range of psycholinguistic methods, which have been employed in previous research designed to investigate foreign-accented speech. Below is a survey of behavioral, electrophysiological, and eye-tracking studies which have investigated various aspects of foreign-accented speech processing.

1.3.1 Behavioral evidence

Existing behavioral studies suggest that foreign-accented speech likely requires more effortful processing in terms of both intelligibility and comprehensibility. While related, they are different (cf. Derwing & Munro, 1997). Intelligibility refers to a listener's actual comprehension of the intended message; that is, the end result of processing indicating successful (or unsuccessful) understanding. Comprehensibility, on the other hand refers to the relative ease/difficulty in understanding a sample of speech and corresponds to the amount of time or effort required for processing. Work examining intelligibility and comprehensibility of foreign-accented speech has also shown differences between native- and foreign-accented speech. In particular, much work has been focused on how listeners adapt in order to process non-native variability more successfully or efficiently.

Typically measured by transcription accuracy, intelligibility is initially poor, but improves with repeated exposure (cf. Gass & Varonis, 1984; Bradlow & Bent, 2003; Bradlow & Bent, 2008). Furthermore, listeners benefit from exposure to multiple speakers of the same accent type, showing the ability to achieve speaker-independent adaptation (Bradlow & Bent, 2008;

Sidaras, Alexander, & Nygaard, 2009). This points to the role of experience in the processing of non-native speech.

Work on comprehensibility has shown that when listening to foreign-accented speech response latencies are longer in comparison to native speech (Munro & Derwing, 1995; Clarke & Garrett, 2004; Floccia, Butler, Goslin, & Ellis, 2009). Munro and Derwing (1995) found that native listeners took significantly longer to evaluate the truthfulness of sentences spoken in Mandarin-accented English. Additionally, Floccia et al. (2009) found that in a lexical decision task (in sentential context) the introduction of a foreign-accented talker lead to a large perturbation in reaction time, followed by a smaller, but long-lasting increase. They also determined that this difference in reaction time to foreign-accented speech does not improve with repeated exposure. Similarly, Adank and McQueen (2007), failed to find short-term adaptation to an unfamiliar regional accent even after 20 minutes of exposure. These results are in conflict with those of Clarke and Garrett (2004) who found that adaptation, in relation to baseline reaction time, can occur in as little as 2 to 4 sentences. The reason for the difference in results across these studies is unclear, though it may be due to differences in the tasks employed in each. Again, however, these studies tend to dichotomize speech into accented and non-accented groups.

A recent study, though, found that the speed of perceptual adaptation (as measured by reaction times) is dependent on two factors, namely the strength of foreign accent and listener familiarity with the accent (Witteman et al., 2013). A cross-modal priming task was used in which participants heard an accented prime and then made a lexical decision to a written target which either matched or mismatched the prime word. Accent strength was manipulated by including words with three Dutch vowels which native German speakers generally produce with varying degrees of accuracy. The strongly accented words contained a diphthong which does not exist in German and German speakers produce with a perceptually and acoustically different vowel. The medium accented words also contained a diphthong which does not exist in German, but that German speakers produce with a vowel that is phonetically very similar. The weakly accented words contained vowels that are shared between the two languages. Participants were grouped for familiarity based on the university that they attend; participants at a Dutch university near the German border were taken as the high familiarity group and participants at a central Dutch university who reported hearing German-accented Dutch less than once per week from fewer than two speakers were taken as the low familiarity group. While this study used relatively coarse-grained groupings for strength and familiarity, it indicates that both degree of foreign accentedness and listener experience with a particular accent are likely involved in processing. This dissertation attempts to address both of these issues in more detail.

1.3.2 Electrophysiological evidence

Very few studies on foreign-accented speech have employed the event-related potential (ERP) experimental methodology. ERP uses electroencephalography to measure the brain's response to a specific event, in this case a spoken stimulus. Through electrodes placed on the scalp, electrical voltages are measured with high temporal resolution and the resulting voltage fluctuations can be used to understand how processing happens in the mind. Only two studies of this type involving foreign-accented speech are known to the author (Goslin et al., 2012; Hanulíková, van Alphen, van Goch, & Weber, 2012).

The two studies have found processing differences between native and non-native speech, although each examines a different aspect of language processing. Hanulíková et al. (2012) found differences for grammatical violation made by native and non-native talkers in the P600, an ERP component known to be sensitive to grammatically anomalous stimuli (cf. Osterhout & Holcomb, 1992). In their study, a native Dutch speaker and a non-native Dutch speaker (native Turkish) produced gender agreement errors in spoken sentences. As expected, a P600 effect was seen for gender violations in native speech, while no P600 was seen for the same gender violations produced with a non-native accented. Additionally the authors examined sentences containing semantic violations, as the N400 component is known to be sensitive to semantically anomalous stimuli (cf. Kutas & Federmeier, 2011). In this study, similar N400 effects for both talkers (i.e., no difference between native and non-native), indicating that there was no general semantic integration problem for the L2 speech. Thus, listeners appear to adjust their expectations about the occurrence of grammatical errors, but not semantic errors when spoken by a non-native talker.

Examining more phonetic questions, Goslin et al. (2012) investigated normalization mechanisms used when processing sentences spoken with either a regional accent or a foreign accent. Speaker normalization refers to process by which talker variability is filtered out of the signal to arrive at underlying phonological categories. To investigate this, they examined the Phonological Mapping Negativity (PMN), an ERP component which corresponds to the mapping of phonetic input onto contextually or experimentally established phonological expectations (cf. Newman & Connolly, 2009). Their results showed that the PMN elicited by the final word of sentences spoken with an unfamiliar regional accent was greater than for those produced in the listener's own accent, while the PMN for foreign accented speech was not present. Foreign accents also resulted in a reduction in N400 amplitude when compared to both unfamiliar regional accents and the listener's own accent. Goslin et al. (2012) interpreted the results as an indication that regional accent-related variations are normalized at the earliest stages of spoken word recognition such that it does not impact lexical access. Conversely, foreign-accented speech does not show the same normalization, as indexed by the PMN, and thus influences later lexical processing.

The authors argue that listeners are able to employ mechanisms at a prelexical processing level to normalize speech coherent with their own native language system (i.e., regional accent), such that lexical processing appears relatively unperturbed. On the other hand, speech variation not consistent with the listener's native language system (i.e., foreign accent) does not appear to be effectively normalized. The unreliable acoustic-phonetic information therefore requires increased reliance upon top-down contextual cues, thus reducing lexical activation. Goslin et al. (2012) frame their results within the context of two hypotheses regarding the processing of non-native speech: The Different Processes Hypothesis and the Perceptual Distance Hypothesis.

The Different Processes Hypothesis states that native and non-native speech will be processed by different normalization mechanisms. Here, native speech relies on a prelexical processing level for normalization, though this is not sufficient to effectively normalize foreign-accented speech which instead increasingly relies on top-down contextual cues. The Perceptual Distance Hypothesis, on the other hand, states that normalization processes for regional-accented speech will simply be attenuated versions of those employed for foreign-accented speech because the regional speech is more perceptually similar to native speech. Goslin et al. (2012) concluded that their results are in line with the Different Processes Hypothesis because the PMN is not greatest for the foreign-accented speech which is the most divergent acoustically. They also indicated that foreign-accented speech presents problems for models of spoken word recognition and the understanding of how accent variation is stored and represented.

While these studies show that processing of foreign-accented speech may indeed differ from native speech, they present somewhat contradictory results, at least with regard to the issue of lexical access and integration. Additionally, they do not adequately address two main issues related to foreign-accented speech, namely, those of degree of accentedness and listener experience with the accent(s) under investigation. Goslin et al. (2012) attempt to address perceptual distance by including regional- and foreign-accented speech, but do not take a more gradient continuum of foreign accentedness into consideration. Furthermore, the influence of listener experience on processing was not examined or controlled.

1.3.3 Eye-tracking evidence

The Visual World (Word) eye-tracking paradigm has been used extensively to investigate the time-course of word recognition (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Dahan, Magnuson, & Tanenhaus, 2001; Huettig & McQueen, 2007; McQueen & Viebahn, 2007). However, it has not been used much to explore online processing of foreign-accented speech. To the knowledge of the author, one study has employed the methodology to investigate the recognition of non-native pronunciation variants (segment-based accent) in L2 listeners (Hanulíková & Weber, 2012). In their study, Hanulíková and Weber (2012) presented three English *TH* (/θ/) pronunciation variants to native English, Dutch, and German participants. Participants' eye gaze was recorded while viewing four written words on screen and listening to sentences containing a final word with one of the variants. For example, participants saw the words *thing*, *ring*, *play*, and *change* on screen and heard a sentence ending in the production *ting*. It was then possible to calculate the amount of looks to each word upon the presentation of the auditory stimulus as a measure of processing.

The three variants were /f/, /t/, and /s/, each corresponding to typical realizations produced by English, Dutch and German speakers respectively. So, English speakers would have more experience with the /f/ variant, Dutch with the /t/ variant, and German with the /s/ variant. It was thus hypothesized that each group would more easily process the variant that is consistent with their own accent, despite the fact that /f/ is the most perceptually similar to /θ/. This was indeed found to be the case; each group produced more looks to the target word for the variant with which they had the most experience, not the variant that is most perceptually similar. Hanulíková and Weber (2012) do not take a firm stance on the representations and mechanisms underlying this processing; however, they do indicate the important role experience in processing variant forms. While this study does not specifically address the L1 lexicon, it supports the possibility that lexical representations are more richly defined based upon experience with a particular type of input (Pierrehumbert, 2001; Pierrehumbert, 2003a).

1.3.4 Multiple methods and converging evidence

Behavioral, electrophysiological, and eye-tracking methods each provide a slightly different vantage point from which to examine foreign-accented speech. Behavioral studies can give indication of both judgments and the result of linguistic processing through measurements of performance. Electrophysiological studies on the other hand provide a continuous measure of stimulus processing over time even when no behavioral response is elicited or when there is no change in behavioral response. Eye-tracking, specifically the visual world/word paradigm, allows for an online and continuous measurement of the information the brain is processing. Converging results across methodologies would begin to provide a fuller view of the effect of gradient foreign accentedness on language processing. The studies contained in this dissertation employ all three of the aforementioned methods in order to understand the phenomenon of gradient perceived foreign accentedness and its effect on lexical processing.

1.4 Lexical processing and representation

The phenomenon of foreign-accented speech comes to bear on our understanding of models of spoken word recognition and lexical representation. The ability of humans to deal with this variability is central to our understanding because foreign-accented speech represents a form of variability that does not fully align with native distributions of the acoustic-phonetic dimensions of speech. How this variability is represented in the mind and handled during lexical processing is still debated. Below is a brief description of current models of spoken word recognition, speech variability, and adaptation through experience.

1.4.1 Spoken word recognition

Models of spoken word recognition attempt to explain the complex process of decoding the speech signal in order to access specific items stored in the mental lexicon. It is generally agreed upon that the recognition of spoken words involves two main aspects, activation and competition. As the speech signal unfolds, lexical items consistent with that input become activated and this activation increases in strength for items that increasingly match the input. It is possible that at any given point, there may be multiple words which could possibly match the input signal. These items then compete, perhaps exerting influence on each other until only the best match to the input remains. While these aspects are not generally disputed, researchers still debate the exact process(es) by which spoken language is understood and the requisite information which must be stored in order for the process to occur.

Modern models differ primarily on the assumptions they make about how processing might take place. One view is that these stages constitute modular units, roughly corresponding to levels of linguistic description, which process specific information and then pass it along to the next stage. Models like this, such as Cohort (W. Marslen-Wilson & Tyler, 1980), Merge (Norris, McQueen, & Cutler, 2000), Shortlist B (Norris & McQueen, 2008), and TRACE (McClelland & Elman, 1986), assume abstract lexical representations. The process thus maps the signal to abstract phonemic categories and passes the rarified information on to the lexicon, stripping away variation. Early models like Cohort rely on hard phonemic decisions, though later models allow for finer-grained detail through phoneme activation units (as in Merge) or phoneme probabilities (as in Shortlist B). TRACE, however, relies on a subphonemic feature level consisting of seven acoustic-phonetic features which unfold over time. Each feature is represented by a continuous vector encoding its degree of presence, which then activates a particular phoneme and then in turn activates a lexical item. This allows for the model to handle variability associated with coarticulation.

Another view takes a decidedly different stance which integrates phonetic, lexical and sentential processing, allowing detailed information from the signal to influence higher levels of processing. Models of this sort do not make the same assumptions or distinctions of how speech is mapped to words (e.g., Klatt, 1979) and how processing might unfold through time (e.g., Gaskell & W. D. Marslen-Wilson, 1997; Gaskell & W. D. Marslen-Wilson, 2002). Because there is a growing body of evidence that talker- and group-level information carried in the speech signal can influence spoken word recognition (Dufour & Nguyen, 2014; Goldinger, 1996; Johnson, 2006; Luce & Lyons, 1998; McLennan & Luce, 2005; Pisoni, 1997; Strand, 2000), it seems unlikely that the output of speech perception is a string of phonemes devoid of non-phonemic variability. Foreign-accented speech presents a special challenge to models of spoken word recognition because of the variability it contains. Obviously, the process by which humans understand words spoken with a foreign accent is capable of handling this non-native variability, even for words produced with a relatively strong accent. Models of this process must then be capable of handling this variability and account for how experience might change how it is handled.

1.4.2 Variability

There are two general accounts for how variability in the speech stream is handled. One, the representational account, asserts that variability in the realization of spoken words is encoded in long-term lexical memories, which in turn affects recognition. The other, the processing account, asserts that lexical representations do not encode variability, which is instead dealt with at the phonemic level before access to the lexicon.

Representational accounts maintain that variation is encoded in lexical representations and that perception can take place without speaker normalization or the conversion of phonetic instantiations to abstract categories before being mapped onto the lexicon (Goldinger, 1996;

Goldinger, 1998; Johnson, 1997; Johnson, 2006; Pierrehumbert, 2001; Pierrehumbert, 2003a; Pisoni, 1997). Within this set of accounts, episodic (exemplar-based) models, such as those proposed by Goldinger (1996) and Johnson (2006) assume that everything about phonetic experiences is stored in detail and becomes part of the representation. This detail subsequently affects the categorization of sounds and the recognition of lexical items. Another representational account (Ranbom & Connine, 2007) claims that multiple abstract representations for variant forms exist in addition to the canonical representation and the link between them is determined by the frequency with which the variant occurs. Alternatively, it may be that phonetic variation is stored in memory with the lexical prototype (Klatt, 1979). This set of accounts easily accommodate the idea that linguistic experience builds up over time and that representations can also include indexical information (e.g., gender, race, and dialect) about specific talkers or groups of talkers. These clusters of representations might then serve to form generalizations about unseen, or in this case unheard, input.

Processing-based accounts, on the other hand, maintain that the lexicon contains a single canonical representation for each word. Under this view, lexical representations are abstract and possibly phonologically underspecified (Lahiri & W. Marslen-Wilson, 1991). In this way, variability in the speech stream must be handled by prelexical processing involving sublexical phonological abstraction at the level of the phoneme (McQueen, Cutler, & Norris, 2006). Research supporting this claim suggests that variability is not part of the lexical representation because perceptual adaptation through experience can influence recognition of previously unheard words containing a phoneme-level deviation. This then points to a processing mechanism by which the signal is first mapped for individual sound segments before accessing the lexicon. Any effect of experience thus affects the process, not the representation. However, these studies have looked exclusively at variation of specific, individual phonemes, rather than variation along multiple dimensions in the entire word.

By examining the processing of gradient foreign accentedness it is possible to have a more nuanced view of the differences, and possible commonalities, between these two broad accounts of variability. Additionally, investigating the effect of experience (with a particular accent) on processing can provide evidence for how representations might evolve over the long term for the purposes of processing accent-related variation.

1.4.3 Experience and generalization

Anecdotally, it seems that foreign-accented speech may become easier to understand with increased exposure. There is also experimental evidence for this effect of experience, most of which comes from behavioral studies. Examining reaction time to visual probe words preceded by sentences produced by non-native speakers of English, Clarke and Garrett (2004) found that processing speed is initially slower for non-native speech compared to native speech but that this difference diminishes within one minute of exposure. This adaptation effect can occur in the span of as little as 2 to 4 sentences. A similar effect for the speed with which listeners adapt perceptually to foreign-accented speech has been shown to be partly dependent upon listener familiarity with the accent (Witteman et al., 2013). Here listener familiarity was defined broadly and based on geographic location.

There have also been two studies which examine the recognition of foreign-accented speech in L2 speakers. In a cross-modal priming experiment, it was shown that accented tokens consistent with the listener's own accent facilitate L2 word processing, while no facilitation was seen for tokens in a different accent (Weber, Broersma, & Aoyagi, 2011). This indicates that linguistic experience with a foreign accent affects the ability to recognize words carrying this accent. A similar result was found using visual word eye-tracking during L2 word recognition in the study described above (Hanulíková & Weber, 2012). Here participants most easily processed pronunciation variants consistent with their own accent, in effect the variant with which they had the most experience, either through proprioception or perception of tokens from talkers with the same language background.

These effects of short and long term experience suggest that listeners are capable of using exposure to variability to facilitate lexical processing, perhaps through representation-based encoding of foreign accent. This effect of exposure has also been demonstrated in transcription tasks measuring the intelligibility of foreign-accented speech. A number of studies have shown that intelligibility is initially poor but improves with repeated exposure to the talker (cf. Gass & Varonis, 1984; Bradlow & Bent, 2003; Bradlow & Bent, 2008). Furthermore, exposure to non-native talkers may lead to generalization for specific accents. Listeners seem to benefit from exposure to multiple talker of the same accent type, and through this exposure have been shown to achieve talker-independent adaptation for a given accent (Bradlow & Bent, 2008; Sidaras et al., 2009). However, this is not the case for an accent which is different from those in the exposure (Weil, 2001). This generalization may be long-term learning through experience which helps listeners avoid repeating the same learning processes over and over again, talker by talker. Despite this research on accent generalization, it remains unclear how listeners get from experience with individual talkers to group-level generalizations.

Despite the evidence for experience-based processing, most current models of spoken word recognition have little architecture for dealing with these apparent changes to the system. Experience is either not present, or relegated to the speech perception system which maps the input to a stipulated representation. There are, however, some models which actively integrate time, experience, and learning into word recognition. One family of these utilizes an Adaptive Resonance framework (Grossberg, Boardman, & Cohen, 1997; Grossberg, Govindarajan, Wyse, & Cohen, 2004; Grossberg & Myers, 2000; Grossberg & Stone, 1986; McLennan & Luce, 2005). Adaptive Resonance models generally abandon the assumptions made by traditional models of spoken word recognition in favor of a system which can learn to make categorical distinctions. In this way, speech units are emergent rather than stipulated (Goldinger & Azuma, 2003). This type of Adaptive Resonance has even been implemented in an exemplar-based model of the lexicon (Johnson, 2006). Another model of spoken word recognition proposed by Plaut and Kello (1999), instead uses distributed connectionism and allows for emergent phonological categories. In this model a discrete set of acoustically plausible dimensions are used as input from which the system derives categories. These types of models which allow for experience to shape recognition may provide a fruitful framework for dealing with the variability inherent to foreign-accented speech.

1.5 Synopsis of dissertation

This dissertation is meant to explore both the perception and processing of gradiently accented non-native speech through a series of three experimental studies. These studies are presented as separate, self-contained journal-style papers. Below is a brief description of the stimuli and statistical methods common to the three papers along with a short description of each study.

1.5.1 Stimuli

In this dissertation, I have limited myself to one particular foreign accent, namely, Chinese-accented English. The purpose for this is two-fold. First, there are many speakers in Canada with Chinese as their mother tongue; In the 2011 Canadian Census, approximately 1 million people reported Chinese (inclusive of Mandarin, Cantonese, and not specified) as their first language. This represents around 3% of the population and is the second largest group of speakers with a first language other than English behind those who speak aboriginal languages. Second, by restricting the L2, it allows for a more tightly controlled examination of the variability inherent to the accent of a specific group. Additionally, because we are dealing with one accent group, it is possible to examine the influence of experience interacting with Chinese-accented speakers on processing.

The stimuli used throughout this dissertation were retrieved from the Northwestern University Wildcat Corpus of Native- and Foreign-Accented English (Van Engen et al., 2010). The Wildcat Corpus is a freely available resource and speech samples from nine male native Chinese speakers and seven male native English speakers were retrieved. The recordings consisted of a list of English words. A subset of 40 monosyllabic words was selected and extracted from each speaker's recording. The use of these stimuli across the three studies contained in this dissertation allows for comparability of results, affording the possibility to provide converging evidence based on the same stimuli across multiple methodologies.

1.5.2 Statistical methods

Given the assumption that foreign accentedness is gradient, it is important to employ statistical methods that treat it as such. As stated above, the majority of studies on foreign-accented speech dichotomize speech into factors, typically accented vs. unaccented, or high accent vs. low accent. While this may be expedient for the purpose of analysis, it completely disregards the fact that listeners understand that accentedness is a degree of magnitude. Additionally, some productions, while coming from a non-native talker, may be more similar to a native production than to talkers of the same first language. Dichotimization for the purposes of achieving a factorial design not only imposes a priori categorization of productions, but also renders it impossible to understand the effect of accent gradience.

For this reason, the statistical analyses contained in this dissertation make use of generalized additive mixed modeling (GAMM) (Hastie & Tibshirani, 1990; Wood, 2006). An important feature of GAMM is that it does not assume a linear relationship between predictors as does ANOVA or linear models. Therefore it is capable of handling non-linearities in the data and allows for the possibility that gradient accent affects processing differently at different points along the continuum. Additionally, GAMM allows for complex interactions between two or more continuous predictors such as degree of accentedness and time. Lastly, GAMM allows for the inclusion of random effects, similar to linear mixed-effects models (cf. Wood, 2006; Wood, Scheipl, & Faraway, 2013).

With regard to its application, GAMM has been successfully used to model linguistic data. Specifically, it has been applied to dialectal variation (Wieling, Nerbonne, & Baayen, 2011; Wieling, Montemagni, Nerbonne, & Baayen, 2014), event-related potentials (Kryuchkova, Tucker, Wurm, & Baayen, 2012; A. Tremblay & Baayen, 2010), prosody (Arnold, Wagner, & Baayen, 2013), and reaction times (Baayen, 2010a; Baayen, 2010b). GAMM has been used extensively outside of linguistics, particularly in the field of ecology (cf. Zuur, Ieno, Walker, Saveliev, & Smith, 2009).

1.5.3 Specific studies

The first study, which is set out in Chapter 2, aims to address the acoustic and lexical variables that drive the perception of gradient foreign accentedness. This study employs a ratings task in which participants judge on a 1 to 9 scale how native-like specific word productions are. The productions presented to raters are from one native English speaker and nine native Chinese speakers. The influence of both speaker-dependent (i.e., acoustic) and speaker-independent (i.e., lexical) variables on these ratings are examined. The speaker-dependent variables are included as acoustic distances measured in relation to a native acoustic reference. This reference point is a typical native speaker value calculated by averaging measures from the remaining native English speakers noted above. The speaker-independent variables are related to properties of individual words, not influenced by the speech signal. Here, lexical frequency (Hasher & Zacks, 1984), phonotactic probability (Vitevitch & Luce, 1999), and phonological neighborhood density (Luce & Pisoni, 1998) are considered. In this way it is possible to examine in a single model which acoustic and lexical variables affect the perception of foreign-accentedness and how. The model indicates that the perceived accentedness is indeed affected by both acoustic distance and probabilistic information contained in the lexicon. Specifically, the shorter the acoustic distances and the better the fit to probable lexical properties, the less accented a token is judged to be. This suggests that the perception of variation (at least at the word level) is affected by more than an L2 talker's ability to approximate typical native speaker values on acoustic measures. The results are discussed in terms of matching variability in the input to multidimensional representations.

Chapter 3 presents the second study which investigates how foreign accentedness affects early auditory processing of native and Chinese-accented English. Using an electrophysiological measure, this study examines the brain's early response to foreign-accented speech as well as the influence of listener experience with Chinese-accented speakers. Here, event-related potentials (ERPs) were recorded for spoken English words varying in degree of foreign accentedness, the same tokens utilized in the previous study. In this study three auditory-evoked potentials are considered; the N100-P200 complex and the Phonological Mapping Negativity (PMN). The model indicates amplitude changes for each of the components across the continuum of foreign accentedness. Specifically, gradient patterns along the continuum emerge for the N100, P200, and PMN. This suggests that variability in the input affects involuntary brain reponse in a graded fashion. Additionally, amplitude modulations differ between experience groups, in some cases showing a reversal of pattern along the accentedness continuum. It seems that experience with Chinese-accented speakers leads to underlying changes in auditory processing, perhaps through long-term adaptation to accent-specific variability. The results are discussed in relation to online processing of speech variability and long-term exposure to the variability inherent to foreign-accented speech.

The third and final study, found in Chapter 4, contains two separate experiments which examine the effect of gradient foreign accentedness on lexical activation strength and the time-course of word recognition. In both experiments, a subset of the previous stimuli (still representing the full continuum of accentedness) were presented to participants. The first of these is a lexical decision task involving cross-modal priming. This experiment allows for the investigation of how strongly gradiently accented words prime their intended target lexical representations. The reaction time model indicates that as degree of accentedness increases, the effect of the prime decreases. The second experiment uses visual word eye-tracking to examine the time-course of spoken work recognition when processing words along the accentedness continuum. The eye-tracking model indicates that the time-course of recognition differs across the accentedness continuum, slowing access as accentedness increases. In addition, listener experience with Chinese-accented English significantly influences both activation strength and word recognition in the two experiments, respectively, mitigating the effect of accentedness. Taken together, this suggests that degree of foreign accentedness influences the strength with which a token activates the corresponding representation as well as the time-course of online word recognition. This suggests that multiple dimensions map as a set of properties, the fit of which influences activation and certainty during recognition. Additionally, the effect of listener experience gives an indication of at least some generalized adaptation to Chinese-accented talkers. The results are discussed in terms of complex, multidimensional lexical representations and long-term, generalized adaptation to foreign-accented speech.

References

Adank, P. & McQueen, J. M. (2007). The effect of an unfamiliar regional accent on spoken-word comprehension. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)* (pp. 1925–1928). Dudweiler: Pirrot.

Arnold, D., Wagner, P., & Baayen, R. H. (2013). Using generalized additive models and random forests to model German prosodic prominence. In *Proceedings of Interspeech 2013* (pp. 272–276). Lyon, France.

Baayen, R. H. (2010a). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon*, *5*(3), 436–561.

Baayen, R. H. (2010b). The directed compound graph of English. An exploration of lexical connectivity and its processing consequences. In S. Olson (Ed.), *New impulses in word-formation (Linguistische Berichte Sonderheft 17)* (pp. 383–402). Hamburg: Buske.

Baker, R. E., Baese-Berk, M., Bonnasse-Gahot, L., Kim, M., Van Engen, K. J., & Bradlow, A. R. (2011). Word durations in non-native English. *Journal of Phonetics*, *39*(1), 1–17.

Bradlow, A. R. & Bent, T. (2003). Listener adaptation to foreign accented English. In M. Sole, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003)* (pp. 2881–2884). Barcelona, Spain.

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

Clarke, C. M. & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, *116*(6), 3647–3658.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*(4), 317–367.

Derwing, T. M. & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition*, *20*(1), 1–16.

Dufour, S. & Nguyen, N. (2014). Access to talker-specific representations is dependent on word frequency. *Journal of Cognitive Psychology*, *26*(3), 256–262.

Flege, J. E. (1980). Phonetic approximation in second language acquisition. *Language Learning*, *30*(1), 117–134.

Flege, J. E. & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America*, *76*(3), 708–721.

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

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

Gaskell, M. G. & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, *12*(5–6), 613–656.

Gaskell, M. G. & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology*, *45*(2), 220–266.

Gass, S. & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning*, *34*(4), 65–87.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1166–1183.

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

Goldinger, S. D. & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, *31*(3–4), 305–320.

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*, *122*(2), 92–102.

Grossberg, S., Boardman, I., & Cohen, M. (1997). Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(2), 481–503.

Grossberg, S., Govindarajan, K. K., Wyse, L. L., & Cohen, M. A. (2004). ARTSTREAM: A neural network model of auditory scene analysis and source segregation. *Neural Networks*, *17*(4), 511–536.

Grossberg, S. & Myers, C. W. (2000). The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychological Review*, *107*(4), 735–767.

Grossberg, S. & Stone, G. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review*, *93*(1), 46–74.

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887.

Hanulíková, A. & Weber, A. (2012). Sink positive: Linguistic experience with *th* substitutions influences nonnative word recognition. *Attention, Perception, & Psychophysics*, *74*(3), 613–629.

Hasher, L. & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, *39*(12), 1372–1388.

Hastie, T. & Tibshirani, R. (1990). *Generalized additive models*. Monographs on Statistics & Applied Probability 43. Chapman and Hall/CRC.

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

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mulinnex (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego: Academic Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, *7*(3), 279–312.

Kryuchkova, T., Tucker, B. V., Wurm, L. H., & Baayen, R. H. (2012). Danger and usefulness are detected early in auditory lexical processing: Evidence from electroencephalography. *Brain and Language*, *122*(2), 81–91.

Kutas, M. & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647.

Lahiri, A. & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, *38*(3), 245–294.

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2007). Speaker-independent factors affecting the perception of foreign accent in a second language. *Journal of the Acoustical Society of America*, *121*(4), 2327–2338.

Luce, P. A. & Lyons, E. A. (1998). Specificity of memory representations for spoken words. *Memory & Cognition*, *26*(4), 708–715.

Luce, P. A. & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36.

Marslen-Wilson, W. & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1–71.

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

McLennan, C. T. & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 306–321.

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

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

Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, *36*(1), 39–66.

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

Newman, R. L. & Connolly, J. F. (2009). Electrophysiological markers of pre-lexical speech processing: Evidence for bottom–up and top–down effects on spoken word processing. *Biological Psychology*, *80*(1), 114–121.

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

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*(3), 299–370.

Osterhout, L. & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806.

Park, H. (2013). Detecting foreign accent in monosyllables: The role of L1 phonotactics. *Journal of Phonetics*, *41*(2), 78–87.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. L. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam: John Benjamins.

Pierrehumbert, J. B. (2003a). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2–3), 115–154.

Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*(2), 191–215.

Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mulinnex (Eds.), *Talker variability in speech processing* (pp. 9–32). San Diego: Academic Press.

Plaut, D. C. & Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWhinney (Ed.), *The emergence of language* (pp. 381–415). Carnegie Mellon Symposia on Cognition Series. Taylor & Francis.

Porretta, V. & Tucker, B. V. (2012). Predicting accentedness: Acoustic measurements of Chinese-accented English. *Canadian Acoustics*, *40*(3), 34–35.

Ranbom, L. J. & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*(2), 273–298.

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in spanish-accented speech. *Journal of the Acoustical Society of America*, *125*(5), 3306–3316.

Strand, E. A. (2000). *Gender stereotype effects in speech processing* (Unpublished doctoral dissertation, Ohio State University, Columbus, Ohio).

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

Tremblay, A. & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151–173). London: The Continuum International Publishing Group.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, *53*(4), 510–540.

Vitevitch, M. S. & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*(3), 374–408.

Wayland, R. (1997). Non-native production of Thai: Acoustic measurements and accentedness ratings. *Applied Linguistics*, *18*(3), 345–373.

Weber, A., Broersma, M., & Aoyagi, M. (2011). Spoken-word recognition in foreign-accented speech by L2 listeners. *Journal of Phonetics*, *39*(4), 479–491.

Weil, S. (2001). Foreign accented speech: Encoding and generalization. *Journal of the Acoustical Society of America*, *109*, 2473–2473.

Wieling, M., Montemagni, S., Nerbonne, J., & Baayen, R. H. (2014). Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and sociodemographic variation using generalized additive mixed modeling. *Language*, *90*(3), 669–692.

Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLoS ONE*, *6*(9), 1–14.

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, *75*(3), 537–556.

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Boca Raton: Chapman & Hall/CRC Press.

Wood, S. N., Scheipl, F., & Faraway, J. J. (2013). Straight forward intermediate rank tensor product smoothing in mixed models. *Statistics and Computing*, *23*(3), 341–360.

Zuur, A., Ieno, E., Walker, N., Saveliev, A., & Smith, G. (2009). *Mixed effects models and extensions in ecology with R*. New York: Springer.

Chapter 2

Perception of gradient foreign accentedness

2.1 Introduction

Traditionally, non-native speakers have been thought to have a foreign accent because they fail to produce native-like speech; that is, they do not produce second language (L2) speech sounds (or sequences of L2 speech sounds) as a native speaker would. The existing literature on foreign-accented speech indicates a wide variety of variables which may influence both the presence and the degree of foreign accent (Piske et al., 2001). Despite previous research, it is unclear which factors may affect the perception of gradient foreign accentedness.

Speaker-dependent variables such as age of L2 learning, length of residence in the target language country, gender, formal instruction, motivation, language learning aptitude and amount of native language (L1) use have been shown to influence a non-native speakers accent. As indicated by Piske et al. (2001), both age of L2 learning and amount of continued L1 use

A version of this chapter has been submitted for publication. Porretta, V., Kyröläinen, A.-J., & Tucker, B. V. (submitted). Perceived foreign accentedness: Acoustic distances and lexical properties. *Attention, Perception, & Psychophysics* appear to have the greatest effect on degree of foreign accentedness. If foreign accent is indeed a failure to achieve native-like productions, factors such as age of L2 learning, instruction, and native language usage may capture other underlying articulatory and phonological properties. Speech sounds are produced by manipulating vocal articulators and non-native speakers may lack practice in making the necessary articulatory gestures in terms of both place and timing, thus resulting in a perceived accent (Flege, 1980). The phonology of a speaker's L1 may also influence his/her ability to produce non-native phones by interfering with the speaker's ability to produce L2-specific phonetic detail. Speech sounds that are similar between the two languages appear to interfere with each other because the speaker may judge them to be acoustically different realizations of the same category (Flege & Hillenbrand, 1984). Additionally, the phonotactic rules of the speaker's L1 have been shown to affect accent in the L2; specifically, L2 sequences that align with the phonotactics of the speaker's L1 are judged to be less accented than those that do not (Park, 2013). Given this, non-native productions indeed differ from native productions on a variety of temporal and spectral acoustic measures (Flege & Hillenbrand, 1984; Munro, 1993; Wayland, 1997; Baker et al., 2011). Specifically, this research has shown that word duration, vowel duration, voice-onset time and formant values in non-native productions deviate from typical native speaker values.

Given the acoustic differences between non-native and native productions, native listeners appear sensitive to them and can detect the presence of an accent in as little as 30 ms of a burst release (Flege & Hillenbrand, 1984). Listeners seem to use this fine-grained acoustic information when rating degree of foreign accentedness. A number of studies have shown that accentedness ratings (for various L1-L2 combinations) can be predicted by many of the acoustic variables discussed above (Munro, 1993; Wayland, 1997; Porretta & Tucker, 2012). Witteman et al. (2013) state that vowels drive differences in perceived accentedness as they can vary significantly from standard forms and Munro (1993) showed this for L1 Arabic productions of English vowels. In that study, five native English-speaking listeners (all linguists) rated accentedness of vowel productions from native Arabic speakers. It was shown that first formant (F1) and second formant (F2) frequency measures were most predictive of those ratings, though predictors varied by vowel. In a similar study, Wayland (1997) found that ratings (from native Thai speakers) of native English-speaking productions of Thai words were driven by spectral rather than temporal measures; however the set of predictors varied for each of the different Thai tones. Prior to the study presented here, Porretta and Tucker (2012) found that global accentedness ratings of Chinese-accented talkers could be predicted by interactions between F1 and F2 values as well as between vowel duration and word duration taken from a separate set of words from the same speaker. Thus, at least in the case of Chinese-accented English, the relationship between measures can affect perceived foreign accentedness.

Because it is often noted that accentedness appears to be gradient, the concept of perceptual distance has been advanced for both regional- and foreign-accented speech (Clarke & Garrett, 2004; Floccia et al., 2006; Goslin et al., 2012). Perceptual distance acknowledges non-native deviations and places accentedness on a sliding scale according to a particular accent's acoustic distance from a given native variety. This scale has foreign accents occupying the far end and regional accents somewhere in the middle, closer to the native accent. The underlying assumption of perceptual distance is that a single mechanism handles the processing of this variation (cf. Floccia et al., 2006). This continuum can be examined through the application of distance measures, calculated in relation to a typical native production. This aligns conceptually with the observation that accent strength varies in magnitude across speakers. Furthermore some acoustic studies (Munro, 1993; Wayland, 1997; Porretta & Tucker, 2012) have attempted to explore the foreign accentedness continuum by employing acoustic distance measures. These studies have found that acoustic distances can be used to predict accentedness ratings, suggesting that native listeners match non-native tokens to the distributional properties of their learned native language. This matching process can be considered as an evaluation of the goodness-of-fit of a token relative to the representation of a native-like production. As noted in Pierrehumbert (2003b), any given language occupies a particular region of phonetic space. It may be that the region occupied by a foreign-accent does not match directly with that of the native accent. Thus, this conceptualization of accentedness, can be quantified as distances between the regions along any psycho-acoustic dimension. By calculating multiple distance measures, it is possible to begin to understand the type of multi-dimensional comparison that listeners carry out. Despite the findings that acoustic distance correlates with foreign accentedness, the majority of work on foreign-accented speech processing has largely ignored this multidimensional gradience, instead opting to simply dichotomize talkers into accented or non-accented groups.

While it may be that perceived accentedness is likely affected by the acoustic properties of the signal, it seems less intuitive that it may also be affected by speaker-independent properties of words themselves. These properties include lexical frequency – an estimation of how often a word occurs in a language (cf. Hasher & Zacks, 1984); phonological neighborhood density – an estimation of how many phonologically similar words exist in a language (cf. Luce & Pisoni, 1998); and phonotactic probability – an estimation of how probable a given phonological sequence is in a language (cf. Vitevitch & Luce, 1999). From a psycholinguistic perspective, properties of words and their relationship to each other contribute to the organization of the mental lexicon (cf. Pierrehumbert, 2003b). These lexical variables have long been shown to affect behavioral measures such as reaction time, especially in native spoken word recognition. However, only a few studies have examined lexical variables as they relate to the recognition of foreign-accented speech (Imai, Walley, & Flege, 2005) or the perception of foreign accentedness (Levi et al., 2007).

Imai et al. (2005) investigated the role of phonological neighborhood density in the recognition of native- and Spanish-accented English words. While the focus of their study was to investigate the lexicon of second language learners, the data from their native English controls is particularly relevant here. Using an offline measure of word recognition (transcription accuracy), they showed that lexical frequency and neighborhood density influenced word recognition of Spanish-accented English. Words with higher lexical frequency were recognized better than low frequency words, while recognition was better for words from sparse neighborhoods than for words from dense neighborhoods. While their task was inherently different than the one presented here, it shows that variables like lexical frequency and phonological neighborhood density are still likely at play when dealing with foreign-accented speech. In a task more similar to the one presented here, Levi et al. (2007) examined the influence of lexical frequency on foreign accentedness ratings assigned to English words produced by native German speakers. They found that words with higher lexical frequency resulted in lower accentedness ratings. These two studies, taken together, suggest that not only the properties of the acoustic signal produced by the speaker, but also aspects of lexical representations in the mental lexicon can affect recognition of accented speech and judgments about it. These lexical properties may mediate the process by which listeners match new input to native-like representations and evaluate goodness-of-fit.

The existing research on the perception of foreign accentedness indicates that speaker-dependent factors, particularly those contained within the speech signal, influence the perception of gradient foreign accent. Additionally, this perception may also be affected by speaker-independent factors. However, these speaker-dependent and -independent variables have not been examined together, nor with the same foreign accent. Thus, the present study investigates the perception of gradient foreign accentedness and the potential factors influencing the matching process. An investigation of this sort, examining the perception of gradient foreign accentedness, provides an opportunity to develop an understanding of what contributes to the perception of variability in spoken language. Specifically it indicates the type of information that must be represented in the lexicon and the relationship among representations.

In this study we ask: 1) How do speaker-dependent variables (temporal and spectral acoustic distance measures) influence perceived foreign accentedness of Chinese-accented English? 2) Do speaker-independent variables (lexical measures in addition to frequency) affect the perception of foreign accentedness? 3) If yes, how do these lexical variables influence perception? That both lexical and acoustic properties might affect the perception of gradient foreign accent bears on the process by which non-native variation is matched against a typical native-like production. With regard to speaker-dependent variables, it is predicted that greater acoustic distance from a native reference point will result in stronger perceived foreign accentedness for all acoustic measures considered due to the magnitude of the mismatch. As for speaker-independent variables, it is predicted that easing the matching process (by means of increased activation or probability of the lexical item) will result in weaker perceived foreign accentedness. If speaker-dependent and speaker-independent variables are shown to be factors contributing to degree of perceived foreign accentedness, this approach brings forth the dimensions involved in assessing the goodness-of-fit of non-native productions. The dimensions affecting the evaluation then speak to the contents of representations and the relationship among them.

2.2 Methodology

2.2.1 Participants

Thirty participants (24 female, 6 male) were recruited from the University of Alberta campus area and ranged in age from 18 to 33 years old (M = 22.1, SD = 3.97). All reported having normal hearing and being native speakers of North American English. Of the 30, 5 reported speaking a heritage language other than North American English (French = 3, Italian = 1, West Indies English = 1). Participants received $10 for completing the experimental task.

2.2.2 Materials

Recordings of nine male native Chinese speakers and seven male native English speakers were retrieved from the NU Wildcat Corpus of native- and foreign-accented English (Van Engen et al., 2010). The Chinese speakers were all listed as native speakers of Mandarin Chinese. Each recording contained a word list read three times by a single talker and a subset of 40 monosyllabic words was selected for use in this study (see Table A.1 in the Appendix). Additionally, global accentedness ratings were obtained from the corpus; a detailed description of the ratings task can be found in Van Engen et al. (2010). This global rating was based on each talker's reading of the Stella Passage (Weinberger, 2013) and was judged on a scale of 1 to 9 (1 corresponding to no foreign accent and 9 corresponding to a very strong foreign accent). The selected Chinese speakers represent a broad range of accentedness based on these mean global ratings (M = 5.96, SD = 1.35, min. = 3.1, and max. = 7.41).

For the present study, the first repetition of the word list was chosen and extracted to serve as the individual stimuli for this study. Five measurements were taken from each token for each talker: 1) word duration; 2) vowel duration (of both monophthongs and diphthongs); 3) F1 frequency; 4) F2 frequency; and 5) F3 frequency. While we do not believe that these are the only acoustic properties that a listener may make use of, previous research has indicated that these are likely contributing factors to foreign accentedness. The word and vowel boundaries of the tokens from each speaker were segmented by hand in Praat

(Boersma & Weenink, 2011) inspecting both the waveform and spectrogram. The beginning of the word was marked at consonant burst/onset while the end of the word was marked after consonant aspiration or frication. For word-initial stops with negative VOT, the beginning was marked at the onset of voicing. The beginning of each vowel was marked at the beginning of regular glottal pulses and onset of the voicing bar while the end the vowel was marked at the decrease in F2 energy. Vowels adjacent to sonorant consonants presented difficulty and changes in amplitude were used to identify the boundary, which was then verified auditorily.

Using an automatic script in Praat, temporal measurements of word duration and vowel duration were extracted along with three spectral measurements, F1–F3 measured at the midpoint. Formant values which appeared to be mispredicted values (n = 17) were hand inspected in the spectrogram and new measurements were taken.

Stimuli

The recordings of the nine native Chinese speakers along with one native English speaker were taken as the stimuli for this experiment. The other 6 native English speakers, which we will refer to as the Native Acoustic Reference, were used to calculate mean values of the acoustic variables from which distance measures could be calculated. As such, they were not included as talkers in the ratings task. Individual sound files were created for each token which were then normalized for amplitude. These stimuli were arranged into 10 blocks, each containing all 40 words from the word list (4 words from each talker). Within each block word order was pseudo-randomized. This was done to ensure that raters would hear all talkers in all blocks thus representing the full range of productions.

2.2.3 Procedures

Participants completed the task seated at a computer in a quite room. Stimuli were presented in DMDX (K. I. Forster & J. C. Forster, 2003) with over-the-ear headphones adjusted to a comfortable volume. Participants were instructed that they would see a word on-screen and hear an auditory token of that word. Presenting the written word in conjunction with the auditory stimulus has been shown to make perceived differences between native and non-native speakers more salient (cf. Levi et al., 2007). Participants were asked to rate how much of a foreign accent each stimulus had on a scale from 1 (no foreign accent) to 9 (very strong foreign accent). This response was made via computer keyboard. If a response was not made within six seconds, the program automatically proceeded to the next trial. Short, self-paced breaks were provided between each block.

2.3 Analysis and results

2.3.1 Ratings

The goal of this study is to examine the factors which affect the perception of gradient foreign accentedness. Because we are interested in the underlying concept of accentedness as a continuum, we ultimately approximate it by using mean rating values for each item similar to other studies (cf. Balota, Pilotti, & Cortese, 2001; Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012). Thus, before calculating and modeling these mean values, it was necessary to verify three aspects of the ratings obtained in this study: 1) that raters were using the scale in the appropriate manner (i.e., that they did not reverse the scale); 2) the use of the ratings scale across all participants; 3) the spread of ratings assigned to each talker; and 4) agreement of raters across items.

Inspection of ratings

In total, 12000 ratings were collected, 14 of which contained missing values (0.12% of the data). These missing values were approximately equally spread across the 10 talkers and were removed before proceeding with the analysis. The ratings were first visually inspected and it was verified that no participant reversed the rating scale. Second, the use of the scale across participants was inspected. The distribution of the assigned ratings indicates that no single value was over- or under-used (see Table 2.1). Third, the spread of ratings for each talker was visually inspected and all talkers appeared to be rated across a reasonable spread of the scale. Based on these inspections, no further data were removed.

Table 2.1: Distribution of the assigned ratings.

Rating	Frequency
1	1356
2	1234
3	1497
4	1611
5	1451
6	1531
7	1497
8	901
9	908

Rater agreement

To assess the agreement among raters we calculated two types of Intraclass Correlation Coefficient (ICC): ICC2 and ICC2k, implemented in the package psych, version 1.4.5, of the statistical environment R, version 3.1.0 (R Development Core Team, 2014). The ICC is the ratio of the variance in the data explained by the variance between raters (Shrout & Fleiss, 1979). When the ICC is at one, this indicates a perfect agreement between raters. The ICC2 measure of agreement takes items and raters as random effects when each rater provides a rating for each token. That is, items and raters are treated as random samples from larger populations. Thus, the results of the ICC2 are generalizable to the larger population. For the ICC2, reliability is calculated for a single rater's ratings. ICC2k is similar, except reliability is calculated for mean of rater's ratings. Here, ICC2k is of particular interest because the goal is to rate each item by a set of raters and take the final rating of a token to be the mean of the set. We are not interested in training listeners to rate accented speech; rather, we are interested in the perception of foreign accentedness on average along this scale. The ICC2k indicates a high level of agreement on the mean accentedness value when the raters are viewed as a random sample (see Table 2.2).

Table 2.2: Intraclass Correlation statistics.

Type	ICC	F-value	df1	df2	p-value	lower bound	upper bound
ICC2	0.46	32.14	386	11194	< 0.01	0.42	0.5
ICC2k	0.96	32.14	386	11194	< 0.01	0.96	0.97

Mean ratings

Mean item ratings were then calculated for each token by averaging over raters, henceforth referred to as Mean Item Rating. Item means range from 1.03 to 8.73 with standard deviations from 0.18 to 2.5. When these item means are aggregated by talker, the talker means range from 2.4 to 6.6 with standard deviations from 1.8 to 2.33. The distribution of the Mean Item Rating by talker is illustrated in Figure 2.1.

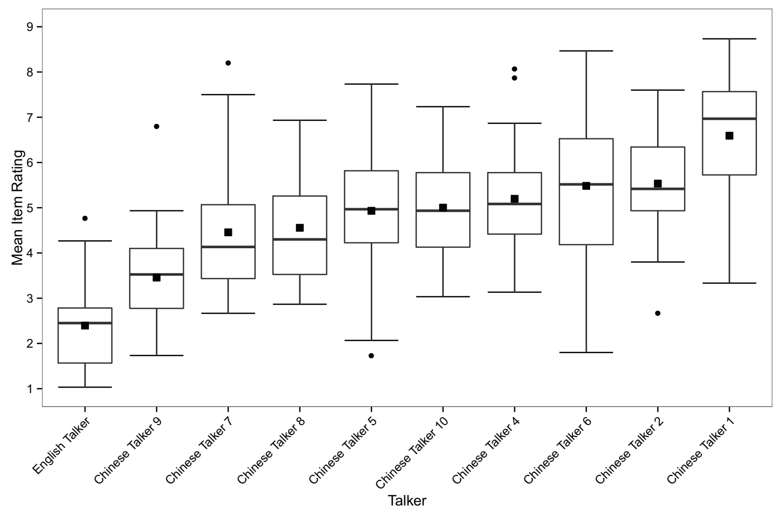

Figure 2.1: Boxplots for Mean Item Rating by talker. The black squares indicate the grand average over items for each talker.

Interestingly, these are highly correlated ($r(8) = 0.9$, $p < 0.001$) with the global accentedness rating (Stella rating) available for each talker in the NU Wildcat Corpus (Van Engen et al., 2010). This further verifies the results of the Intraclass Correlation statistics and indicates that the raters in the present study arrived at similar judgments on average as the listeners in the Stella ratings from the Wildcat Corpus. This suggests that their judgments of accentedness may have been based on similar information. Thus, we assume that the underlying concept of degree of accentedness is likely related across the studies.

2.3.2 Model

The Mean Item Ratings were modeled using generalized additive mixed modeling (GAMM) (Wood, 2006), implemented in the package mgcv, version 1.7-29. GAMM does not assume a linear relationship between predictors as does ANOVA or linear models and allows for complex interactions between two or more numeric predictors. Because, accentedness is thought to be gradient, GAMM allows us to model the possibly wiggly functional form of the predictors (and interactions). Thus, GAMM is capable of handling possible non-linearities in the data. This point is particularly important for the present study due to the working conception of foreign accentedness as a continuum. It may be that predictors affect this continuum differently at different points along it. Additionally, GAMM allows for the inclusion of random effects, similar to linear mixed-effects models (cf. Baayen, Davidson, & Bates, 2008). A detailed description of the model fitting process is presented below, following a description of the input variables.

With regard to the application of GAMM, it has been previously applied successfully to model a variety of linguistic data. In particular, it has been used for investigating dialectal variation (Wieling et al., 2011; Wieling et al., 2014), event-related potentials (Kryuchkova et al., 2012; A. Tremblay & Baayen, 2010), prosodic prominence (Arnold et al., 2013), and reaction times (Baayen, 2010a; Baayen, 2010b). Outside of language, it has been used extensively in the field of ecology (cf. Zuur et al., 2009).

Input variables

Acoustic variables Formant values (F1–F3) were log-normalized for comparison across speaker-specific vowel spaces. Vowel-to-word ratio was calculated by dividing the vowel duration of a word by the total duration of that word. If accentedness is a result of non-native productions approaching native-like acoustic targets (to varying degrees), quantifying the distance from native speaker norms allows examination of how variation along different variables affects perceived accentedness. Native speaker norms are presumed to be the acoustic values of a typical speaker. Here, as in other studies (Munro, 1993; Wayland, 1997), this is operationalized as a Native Acoustic Reference obtained by averaging across multiple native speakers for a particular variable. The six native English speakers subjected to acoustic analysis for calculation of the Native Acoustic Reference were not included as talkers in the ratings task. For each numeric variable, the token value of each talker was subtracted from the Native Acoustic Reference. The absolute value of that difference yielded a positive number representing the magnitude of the distance between the production in question and the Native Acoustic Reference. The acoustic variables are summarized in Table 2.3. An example of Log F1 and Log F2 for the vowel /ɛ/ in *bet* is presented in Figure 2.2.

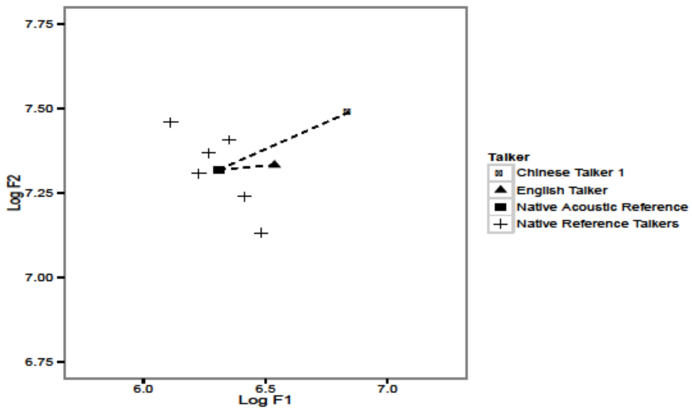

Figure 2.2: Log F1 and Log F2 distances for the vowel /ɛ/ in 'bet'. are shown as a dotted line in relation to the Native Acoustic Reference (square symbol) for the most accented Chinese talker (dot symbol) and the English talker (triangle symbol).

Lexical variables Lexical frequencies for each word were retrieved from COCA (Davies, 2008) and the number of phonological neighbors for each word was extracted from the English Lexicon Project (Balota et al., 2007). As in other studies, neighbors are defined as words with a one-phoneme difference (e.g., /bɛt/ is neighbors with /bæt/, /bɛg/, and /sɛt/). Furthermore, the phonotactic probability of each word was calculated using the Phonotactic Probability Calculator provided online by Vitevitch and Luce (2004). Both of these have been argued to capture the process of matching words to phonological templates when searching the mental lexicon (Pierrehumbert, 2003b). The lexical variables are summarized in Table 2.3.

Standardization and Residualization Pairwise correlations of the input variables were computed to check for possible collinearity among them. In case of collinearity, it is possible to assess the effect of a collinear input variable, though it is difficult to reliably assess the direction of the effect (Dormann et al., 2013). All continuous input variables were standardized to reduce spurious correlation between them by subtracting the mean and dividing by one standard deviation. However, even after standardization, Phonotactic Probability remained

Table 2.3: Summary of input variables.

Variable	Range	*M*	*SD*
Log F1 Distance	0–1.13	0.16	0.16
Log F2 Distance	0–1.04	0.15	0.14
Log F3 Distance	0–0.5	0.09	0.07
Vowel-to-Word Ratio Distance	0–0.32	0.06	0.05
Log COCA Frequency	5.93–14.55	9.39	1.75
Neighborhood Density	1–58	27.5	15.34
Phonotactic Probability	0.05–0.33	0.17	0.06

correlated with Neighborhood Density ($r(398) = -0.44$, $p < 0.001$). Similiar to Jaeger (2010), residualization was carried out to remove collinearity between these variables by fitting a linear model to the standardized Phonotactic Probability as function of the standardized Neighborhood Density. We took Phonotactic Probability to be the response variable because it requires an accumulation of phonemic sequences over words in the lexicon, whereas Neighborhood Density only requires phonological words and the relationship between them. Thus, it seems that, at least conceptually, Neighborhood Density is a simpler variable compared to Phonotactic Probability. The residuals from this model were extracted and, as expected, after residualization, the correlation between Phonotactic Probability and Neighborhood Density was removed ($r(398) = 0$, $p = 1$). Importantly, the residualized Phonotactic Probability is highly correlated with the original measure ($r(398) = 0.9$, $p < 0.001$); thus, the residualized variable can be interpreted in the same manner as the original.

Model fitting and evaluation

The input variables described above were fit to the response variable (Mean Item Rating) with by-Word and by-Talker random intercepts. By-Word random intercepts allow for the possibility that some words may be more likely than others to sound accented. By-Talker random intercepts allow for the possibility that different talkers may be more likely than others to sound accented. For each predictor, a nonlinear functional relation with the response variable was allowed for by using a smooth function. For interactions, tensor product smooths were used, allowing for a wiggly surface (Baayen, 2010b; Wood, 2006).

The model was fit using the backwards step-wise elimination procedure described in Zuur et al. (2009). All input predictors were first included as main effects and the number of smoothing parameters was optimized using Maximum Likelihood. Second, their contribution to the model was evaluated using two criteria. The first criterion was the estimated p-value of the smoothing parameter. The estimated p-value of the smoothing parameter indicates whether or not the functional form of the predictor is different from zero. If greater than the conventional alpha level of 0.05, the predictor was considered for removal. The second criterion was the Akaike information criterion (AIC) (Akaike, 1998). The use of AIC is an information-theoretic approach which supplies information on the strength of evidence for a particular model given the data. Importantly, AIC indicates how much information is lost when a particular predictor is removed from the model; therefore, lower AIC values indicate increased evidence for a given model. This approach is particularly advantageous as it is not affected by the order of variables entered into the model.

Upon removal of the predictor, the AIC value of that model was compared to that of the model containing the predictor in question. If the difference in AIC was less than 2, the predictor was removed and we continued with the model fitting procedure (Burnham & Anderson, 2002). Through this process, Log F3 Distance and Log COCA Frequency were eliminated. Finally, motivated interactions among acoustic and lexical predictors were included and examined using the same criteria. Doing so, two interactions emerged; one between Log F1 Distance and Log F2 Distance, and another between Residualized Phonotactic Probability and Neighborhood Density. Thus the model consists of the following predictors: Random intercepts for Word and Talker, main effect for Vowel-to-Word Ratio Distance, interaction between Log F1 Distance and Log F2 Distance, and interaction between Residualized Phonotactic Probability and Neighborhood Density. Visual inspection showed that the residuals of this model were approximately normally distributed, thus meeting the assumptions of the model and indicating a good fit to the data.

To evaluate the model we calculated ΔAIC values for main effects and interactions in the model (see Table 2.4). This allows us to rank the predictors in terms of strength of evidence. This is done by subtracting the AIC of the model including the predictor (i.e., the full model) from the AIC of the model without the predictor (i.e., the simpler model). As a rule of thumb, a $\Delta < 2$ suggests substantial evidence for the simpler model. Δ values between 3 and 7 indicate considerably less support for the simpler model, while $\Delta > 10$ indicates that the simpler model is very unlikely (Burnham & Anderson, 2002). Additionally, we calculated AIC weight values for the model variants (see Table 2.4). AIC weights are computed on the set of models and give a proportion indicating how often a particular model would be chosen as the most likely given the data. These two measures give an indication of the importance of the predictor within the model and the likelihood of a particular model.

Looking at the ΔAIC values, it can be seen that the smooth function for Vowel-to-Word Ratio Distance produces the largest change and thus indicates that it plays a large role in explaining the data. The interaction of Log F1 Distance and Log F2 Distance also plays an important role, similar in impact to Vowel-to-Word Ratio Distance. While the interaction of Phonotactic Probability and Neighborhood Density plays a smaller role, it still contributes to the model likelihood. As seen in Table 2.4, based on the AIC weights, the full model would be chosen as most likely 91% of the time.

Table 2.4: Model AIC, ΔAIC, and AIC weight for Ratings model.

Model	df	AIC	ΔAIC	AIC Weight
Full model	34.66	1214.96	NA	0.91
w/o tensor: Phonotactic Prob., Neighborhood Dens.	41.67	1219.61	4.66	0.09
w/o tensor: Log F1 Distance, Log F2 Distance	33.62	1241.75	26.79	0
w/o smooth: Vowel-to-Word Ratio Distance	34.7	1243.46	28.5	0

After evaluating the full model, it was refitted using Restricted Maximum Likelihood to optimized the number of smoothing parameters as recommended in Zuur et al. (2009). This final model, reported below, explains 62.9% of the deviance showing that the model is able to capture important facets of gradient foreign accentedness.

Results

The results of the final model are reported along with visualizations of the effects. Table 2.5 presents the statistics for the parametric and smooth terms in the model. The column labeled 'edf' indicates the estimated degrees of freedom of the smooth functions. When equal to 1, as in the case of the Vowel-to-Word Ratio Distance measure, the effect is approximately linear.

The effects are visualized below in Figure 2.3. Importantly, zero on all axes represents the mean.

Table 2.5: Generalized additive mixed model reporting parametric coefficient (Part A) and estimated degrees of freedom (edf), reference degrees of freedom (Ref.df), F and p values for the tensor products and random effects (Part B) for the Ratings model.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
intercept	4.7604	0.3305	14.4022	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
tensor: Res. Phonotactic Prob., Neighborhood Dens.	3.5260	3.7253	7.4733	< 0.0001
smooth: Vowel-to-Word Ratio Distance	1.0001	1.0001	31.9035	< 0.0001
tensor: Log F1 Distance, Log F2 Distance	3.2523	3.4632	10.9498	< 0.0001
random smooth: Word	17.9162	36.0000	1.0107	0.0006
random smooth: Talker	8.7521	9.0000	34.8724	< 0.0001

The interaction of the Log F1 Distance and Log F2 Distances was included as F1 and F2 are established variables involved in vowel categorization. The result of this interaction is presented as a regression surface in Figure 2.3 (top left panel). The contour lines represent the estimated Mean Item Rating values fitted by the model. As both Log F1 and Log F2 Distances increase, mean item ratings also increase. This indicates that increased spectral deviations from the Native Acoustic Reference result in higher mean item ratings. Additionally, this increase in accent strength appears more precipitous along Log F1 Distance, compared to Log F2 Distance.

A linear main effect of Vowel-to-Word Ratio Distance can be seen in Figure 2.3 (top right panel). Vowel-to-Word Ratio represents the amount of the word duration that is subsumed by the duration of the vowel. Vowel-to-Word Ratio Distance represents how far that ratio is from the Native Acoustic Reference for a given word. Increased deviation from the Native Acoustic Reference is correlated with higher mean item ratings.

The interaction of Neighborhood Density and Residualized Phonotactic Probability was included as both variables are involved in word identification and phonological matching (Imai et al., 2005; Luce & Pisoni, 1998; Pierrehumbert, 2003b). The result of this interaction is presented as a regression surface in (Figure 2.3, bottom left). Again, the contour lines represent the estimated Mean Item Rating values fitted by the model. As Neighborhood Density and Residualized Phonotactic Probability both increase, mean item ratings decrease. In particular, mean item ratings are high across neighborhood densities when residualized phonotactic probability is low (to the left in this panel). Likewise, mean item ratings are high across Residualized Phonotactic Probability when Neighborhood Density is low (to the bottom in this panel). However, when both Neighborhood Density and Residualized Phonotactic Probability are high, accentedness ratings steadily decrease. This indicates that when the phonemic sequence is probable and the word has many phonological neighbors, mean item ratings are lower.

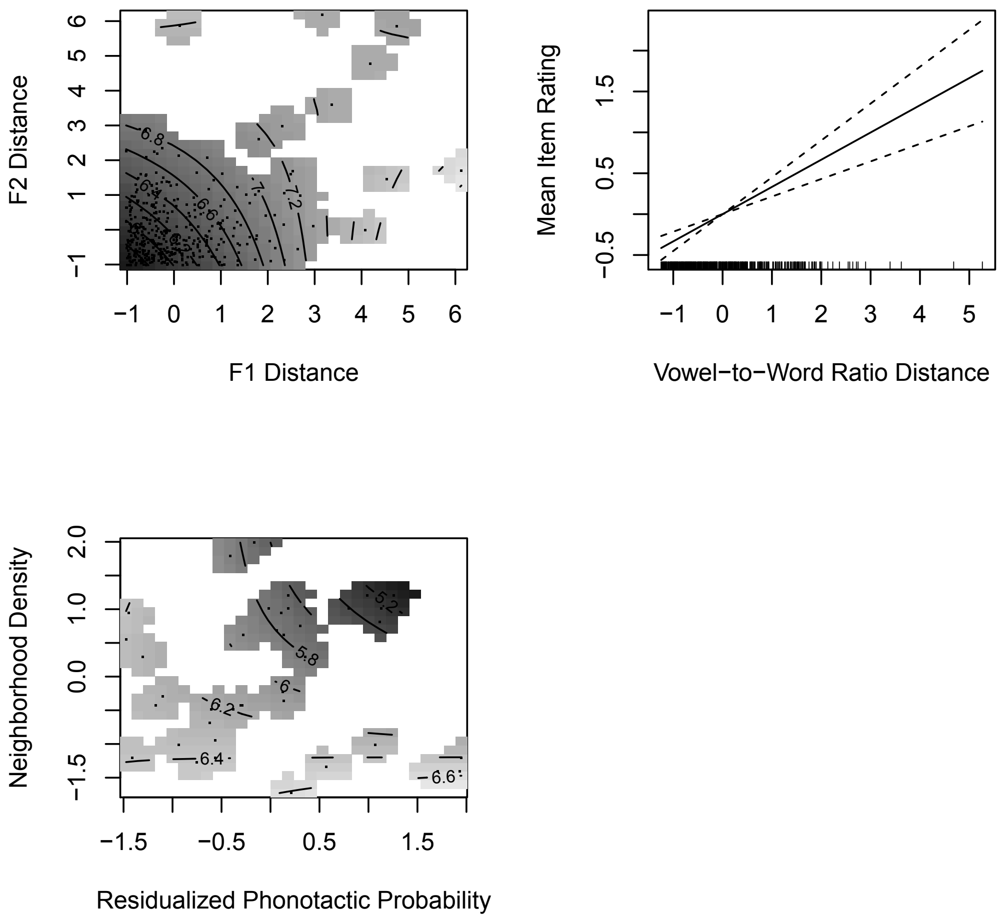

Figure 2.3: Effects of final model. Top left panel: Contour plot of the interaction of F1 distance and F2 distance. Top right panel: Partial effect of Vowel-to-Word Ratio Distance with 95% confidence bands. Bottom left panel: Contour plot of the interaction of residualized Phonotactic Probability and Neighborhood Density. For both contour plots, light gray represents higher mean item ratings (stronger accent) and black represents lower mean item ratings (weaker accent).

2.4 Discussion and conclusions

The present study aimed to examine, in a single model, the acoustic and lexical variables that influence accentedness ratings assigned to 40 English words. These words spoken by ten different talkers: one native English-speaking talker and nine native Chinese-speaking talkers. Based on previous studies, it was predicted that temporal and spectral acoustic distance measures (i.e., speaker-dependent factors) would influence the perception of gradient foreign accentedness. Specifically, increased distance from the Native Acoustic Reference (i.e., mean native speaker value) would fit less well to a native-like representation and would thus result in stronger perceived foreign accentedness. In addition, it was predicted that lexical variables (speaker-independent factors) would also influence perception along the continuum of accentedness. Previous research using ratings suggested that higher lexical frequency would result in lower perceived foreign accentedness. It was predicted that increased probability and activation of a particular item within the mental lexicon might ease the matching process and lead to lower perceived foreign accentedness.

The analysis of the individual ratings indicates that when raters are viewed as a randomly sampled group, the mean item ratings offer a highly reliable measure for the accentedness continuum. This was further corroborated by the high correlation between the mean talker ratings (based on word-specific ratings collected in this study) and the global accentedness ratings for each talker available in the NU Wildcat Corpus.

The mean item ratings were modeled using GAMM. The results indicate that, indeed, both acoustic and lexical variables affect the perception of gradient foreign accentedness. This suggests that both are involved in the matching of a particular token to the representation of what constitutes a native-like production. The content of the acoustic signal (temporal and spectral properties) is what must be decoded and matched. This signal is dependent on the abilities of the talker to approximate a native-like production. However, lexical properties such as frequency, phonotactic probability, and number of phonological neighbors are properties of the lexicon and are not dependent on a talker's production. These may also affect the matching process, and subsequently the perception of foreign accentedness. By including both acoustic and lexical predictors in the model, we control for the effect of one on the other.

The results of this study show that multiple acoustic measures predict variability of foreign accentedness ratings when considered as distances from a Native Acoustic Reference. The present study replicates the results of Munro (1993), who showed that distance variables predict accentedness ratings assigned to English vowels spoken by native Arabic speakers. Here, we demonstrate that with Chinese-accented English words showing that the distance of both F1 and F2 values from typical native productions positively correlates with the strength of perceived accentedness. Also, as suggested by Witteman et al. (2013), deviations in vowel quality appear to be a driving force in the perception of foreign accentedness. Based on the ΔAIC results, it appears that the interaction between the Log F1 and Log F2 distances is quite important in the assessment of foreign-accentedness. This stands to reason, particularly for monosyllabic words, as F1 is a cue to vowel height and F2 a cue to vowel frontness. Jointly, these measures relate to vowel categorization and variability along these frequencies may lead to possible miscategorization which may particularly influence the perception of monosyllabic words.

Additionally, the present results add to those of previous studies by showing the role of temporal properties such as word and vowel durations, more specifically, their relationship to each other. Here, the temporal relationship of Vowel-to-Word Ratio indicates the amount of the word subsumed by the vowel and the distance of this from a typical native value positively correlates with the strength of perceived accentedness. It appears that speakers who are more able to approximate this durational pattern are perceived as less accented. This, taken together with the spectral properties, supports the idea that listeners assess acoustic features including word duration, vowel duration, and formant values produced by the speaker and compare them against native speaker values. It should be noted, however, that we do not believe these to be the only acoustic properties that listeners use to evaluate degree of foreign accent. It is likely that other properties (e.g., VOT, formant transitions/trajectories, etc.) also influence judgments of accentedness.

These acoustic results support the concept of perceptual distance (cf. Floccia et al., 2006) as a framework for understanding the phenomenon of foreign accent. As seen here, perceptual distance can be quantified along any number of acoustic parameters and the magnitude of these deviations generally lead to stronger perceived accentedness. Listeners may maintain fine phonetic and distributional information about what constitutes a native-like production. This could be represented either as a set of exemplars (Johnson, 1997; Johnson, 2006; Pierrehumbert, 2001; Pierrehumbert, 2003a; Walsh, Möbius, Wade, & Schütze, 2010) or a single prototype (Iverson & Kuhl, 1995; Samuel, 1982) which encodes multiple dimensions. In either case, this information is then available for matching new input produced by different speakers. Listeners, instead, seem to be able to judge the goodness-of-fit of a given token to the representation of a native production along multiple acoustic dimensions. Non-native speakers who are more successful at producing speech sounds closer to typical native values are indeed perceived as more native-like. It is also interesting to note that not all of the words spoken by the native talker were rated by listeners as perfectly native. This indicates that even among native tokens, perceptual distance may be taken into consideration by listeners for handling variation in native speech, suggesting the employment of similar mechanisms for both native and non-native speech. However, given the results of the model, perceptual distance is not the only influence on degree of perceived foreign accentedness.

The lexical properties of particular words also appear to exert influence on the perception of gradient foreign accentedness. This is particularly important as it shows that factors, independent of the speaker's ability to phonetically approximate native productions, influence this perception. The present model tests these lexical predictors when all other variables are held constant. That said, we fail to replicate the results of Levi et al. (2007) showing that lexical frequency influences the ratings of foreign accentedness. In their study, Levi et al. (2007) showed that as lexical frequency increases, perceived foreign accentedness decreases. However, lexical frequency was not a significant predictor in our model. One reason for this may be that there were not enough words to provide sufficient power for detecting the effect. Our 40 monosyllabic words were extracted from the NU Wildcat Corpus and we therefore could not control the distribution across lexical frequency. Levi et al. (2007) specifically controlled for frequency in three levels (i.e., low, mid, and high) and had a greater number of words. They did perform a separate, post hoc acoustic analysis of formant values for a subset of their tokens to partially rule out production as a source of the differences seen in levels of lexical frequency. However, they did not specifically control for this in the ratings experiment.

The model does indicate the influence of both neighborhood density and phonotactic probability, and in particular, their interaction. Specifically, as both neighborhood density and phonotactic probability increase, perceived accentedness decreases. This is similar to work on well-formedness (Hay, Pierrehumbert, & Beckman, 2004; Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997). Vitevitch et al. found that ratings of well-formedness of English nonse words reflect the probability of phonotactic sequences. Subsequently, Hay et al. found similar results indicating that well-formedness is gradient with this probability. According to Pierrehumbert (2003b), lexical neighborhoods and phonotactics provide general information about the lexicon and involve matching an input to some phonological template. The interaction in the present model suggests that denser neighborhoods may provide greater activation for the matching process especially when the phonemic sequence is probable. Thus, when the matching process is made easier through increased probability of the phonemic sequence and activation within dense neighborhoods, perceived accentedness is reduced. Based on research on spoken word recognition, it might be expected that sparse neighborhoods would lead to reduced perceived foreign accentedness (Imai et al., 2005; Luce & Pisoni, 1998). However, it is important to note that the task of rating tokens is rather different than word recognition, particularly with the present design. Here the task was offline. Because raters saw the written form of the word on screen while the auditory stimulus was played, the set of possible candidates is restricted automatically, therefore bypassing auditory word recognition. Participants were then judging how native the token was compared to their knowledge of what is an acceptable native production of the word rather than first having to identify the auditory word form. This disentangles the recognition and matching processes.

Another interesting result is the relative strength of evidence for a given predictor in the final statistical model. In this study, ΔAIC values used to rank the predictors. This ranking indicates that the speaker-dependent variables have the strongest influence on perceived foreign accentedness. The speaker-independent variables were also found to play a role, but to a lesser extent. Importantly, the contribution of these variables is unique and beyond that of the speaker-dependent variables. So, it seems that properties of the lexicon can influence perceived foreign accentedness beyond the effect of perceptual distance.

The goodness-of-fit of a given token is not solely dependent on the relationship between the acoustic signal and a native-like representation; rather, this fit is affected by probabilistic information about a particular lexical item as well as its relation to other items. This acoustic and lexical matching process appears to be multi-faceted in nature, such that the shorter the acoustic distances and the better the fit to probable lexical properties, the less accented a token is judged to be. Whether or not the native-like representation is based on a single prototype or a set of exemplars which give rise to a prototype, the results point to a lexical representation that contains multidimensional probabilistic and distributional information about what constitutes a native-like accent as well as various properties of lexical items. Listeners (in this case, raters) likely use all of this learned information when evaluating a token's goodness-of-fit within their native language. Matching processes occur across multiple dimensions and perceived foreign accentedness may reflect the ease with which a match occurs across both acoustic and lexical properties, rather than strict perceptual distance. It thus seems that the perception of variation (at least at the word level) is affected by acoustic distance from native-like representations as well as properties specific to the lexical items themselves. Therefore, an L2 talker's ability to approximate typical native speaker values on acoustic measures is only part of what affects the strength of perceived foreign accentedness, with probabilistic properties of the lexicon modulating the perception of accentedness. So, perceived foreign accentedness may be better described as an index of the ease of the matching process.

Moving forward from these results, it would be interesting to examine the role of listener experience with foreign-accented speech. It may be that the distributional information against which a non-native token is compared, changes over time through experience. If this is the case, one could expect that raters with differing levels of experience may find non-native variability more or less acceptable, due to changes in the distribution which facilitate the matching process. This would indicate that language experience influences the development of the representation of what constitutes a native production for individual listeners.

References

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of Hirotugu Akaike* (pp. 199–213). New York: Springer. (Reprinted from Proceedings of the Second International Symposium on Information Theory, B.N. Petrov and F. Caski, eds., Akademiai Kiado, Budapest, 1973, 267–281)

Arnold, D., Wagner, P., & Baayen, R. H. (2013). Using generalized additive models and random forests to model German prosodic prominence. In *Proceedings of Interspeech 2013* (pp. 272–276). Lyon, France.

Baayen, R. H. (2010a). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon*, *5*(3), 436–561.

Baayen, R. H. (2010b). The directed compound graph of English. An exploration of lexical connectivity and its processing consequences. In S. Olson (Ed.), *New impulses in word-formation (Linguistische Berichte Sonderheft 17)* (pp. 383–402). Hamburg: Buske.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Baker, R. E., Baese-Berk, M., Bonnasse-Gahot, L., Kim, M., Van Engen, K. J., & Bradlow, A. R. (2011). Word durations in non-native English. *Journal of Phonetics*, *39*(1), 1–17.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, *29*(4), 639–647.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459.

Boersma, P. & Weenink, D. (2011). Praat: Doing phonetics by computer [Version 5.3.61].

Burnham, K. P. & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: Springer.

Clarke, C. M. & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, *116*(6), 3647–3658.

Davies, M. (2008). The Corpus of Contemporary American English (COCA): 400+ million words, 1990–present.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46.

Flege, J. E. (1980). Phonetic approximation in second language acquisition. *Language Learning*, *30*(1), 117–134.

Flege, J. E. & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America*, *76*(3), 708–721.

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

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

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*, *122*(2), 92–102.

Hasher, L. & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, *39*(12), 1372–1388.

Hay, J., Pierrehumbert, J., & Beckman, M. E. (2004). Speech perception, well-formedness and the statistics of the lexicon. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic interpretation: Papers in laboratory phonology 6* (pp. 58–74). Cambridge, UK: Cambridge University Press.

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. *Journal of the Acoustical Society of America*, *117*(2), 896–907.

Iverson, P. & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, *97*(1), 553–562.

Jaeger, F. T. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mulinnex (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego: Academic Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499.

Kryuchkova, T., Tucker, B. V., Wurm, L. H., & Baayen, R. H. (2012). Danger and usefulness are detected early in auditory lexical processing: Evidence from electroencephalography. *Brain and Language*, *122*(2), 81–91.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990.

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2007). Speaker-independent factors affecting the perception of foreign accent in a second language. *Journal of the Acoustical Society of America*, *121*(4), 2327–2338.

Luce, P. A. & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36.

Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, *36*(1), 39–66.

Park, H. (2013). Detecting foreign accent in monosyllables: The role of L1 phonotactics. *Journal of Phonetics*, *41*(2), 78–87.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. L. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam: John Benjamins.

Pierrehumbert, J. B. (2003a). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2–3), 115–154.

Pierrehumbert, J. B. (2003b). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probability theory in linguistics* (pp. 177–228). Cambridge, MA: The MIT Press.

Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*(2), 191–215.

Porretta, V., Kyröläinen, A.-J., & Tucker, B. V. (submitted). Perceived foreign accentedness: Acoustic distances and lexical properties. *Attention, Perception, & Psychophysics*.

Porretta, V. & Tucker, B. V. (2012). Predicting accentedness: Acoustic measurements of Chinese-accented English. *Canadian Acoustics*, *40*(3), 34–35.

R Development Core Team. (2014). *R: A language and environment for statistical computing*. Version 3.1.0. R Foundation for Statistical Computing. Vienna, Austria.

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

Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428.

Tremblay, A. & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151–173). London: The Continuum International Publishing Group.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, *53*(4), 510–540.

Vitevitch, M. S. & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*(3), 374–408.

Vitevitch, M. S. & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers*, *36*, 481–487.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, *40*(1), 47–62.

Walsh, M., Möbius, B., Wade, T., & Schütze, H. (2010). Multilevel exemplar theory. *Cognitive Science*, *34*(4), 537–582.

Wayland, R. (1997). Non-native production of Thai: Acoustic measurements and accentedness ratings. *Applied Linguistics*, *18*(3), 345–373.

Weinberger, S. H. (2013). Speech accent archive, George Mason University [http://accent.gmu.edu].

Wieling, M., Montemagni, S., Nerbonne, J., & Baayen, R. H. (2014). Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and sociodemographic variation using generalized additive mixed modeling. *Language*, *90*(3), 669–692.

Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLoS ONE*, *6*(9), 1–14.

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, *75*(3), 537–556.

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Boca Raton: Chapman & Hall/CRC Press.

Zuur, A., Ieno, E., Walker, N., Saveliev, A., & Smith, G. (2009). *Mixed effects models and extensions in ecology with R*. New York: Springer.

Chapter 3

Early auditory processing of gradient foreign-accented speech

3.1 Introduction

When we interact with non-native speakers of English we intuitively know that some speakers are more difficult to understand that others. This difficulty may be related to the strength of their accent. However, despite even the strongest of accents, we generally understand with relative ease, and most people would say that the more time you spend with non-native speakers, the easier they become to understand. There is a considerable body of literature which has examined our ability to comprehend foreign-accented speech as well as our ability to adapt to it behaviorally over short- and long-term exposures (Gass & Varonis, 1984; Bradlow & Bent, 2003; Clarke & Garrett, 2004; Bradlow & Bent, 2008; Witteman et al., 2013). Despite this behavioral evidence, little electrophysiological research has been done to examine how foreign accent affects speech processing. This study aims to explore early auditory processing of foreign-accented speech, taking into consideration both degree of accentedness and listener experience with accented speakers.

3.1.1 Foreign-accented speech

It is believed that foreign-accented speech is perceived as different from native speech because it deviates from the typical distributions of native speaker temporal and spectral acoustic targets. While non-native speakers of a language may fail to reach native-like acoustic targets for a number of reasons (cf. Flege, 1980; Flege & Hillenbrand, 1984), of interest here is how the magnitude of those deviations affects processing. Native listeners appear to be quite sensitive to deviation from native speaker norms. Listeners are capable of detecting accent in as little as a single 30 ms burst release (Flege, 1984) and these deviations have been shown to correlate with how foreign accentedness is perceived (Baker et al., 2011; Munro, 1993; Wayland, 1997; Porretta & Tucker, 2012). While deviations can occur for any segment, Witteman et al. (2013) state that vowels likely drive differences in perceived accentedness as they can vary significantly from standard forms. This observation is borne out in both Porretta and Tucker (2012) and Porretta, Kyröläinen, and Tucker (2013) who show that the interaction between F1 distance and F2 distance (i.e., how deviant the vowel is from a native target in the vowel space) is the strongest acoustic predictor of perceived accentedness. Additionally, it has been shown that lexical variables such as word frequency may also play a role in the perception of foreign accentedness (Levi et al., 2007; Porretta et al., 2013). Despite this work on the acoustic correlates of foreign accentedness, much of the work related to processing of foreign-accented speech has largely ignored this multidimensional gradience, instead opting to simply dichotomize talkers into accented and non-accented groups.

Based on the results of behavioral studies, it is likely that foreign-accented speech requires more effortful processing as response latencies are longer when listening to foreign-accented speech in comparison to native speech (Munro & Derwing, 1995; Floccia et al., 2009). Munro and Derwing (1995) found that listeners took significantly longer to evaluate the truthfulness of sentences spoken in Mandarin-accented English. Floccia et al. (2009) found that in a word identification task introduction of a foreign-accented talker lead to a large perturbation in reaction time, followed by a smaller, but long-lasting increase. Additionally, they determined that reaction time to foreign-accented speech does not improve with repeated exposure to the talker as does intelligibility (cf. Gass & Varonis, 1984; Bradlow & Bent, 2003; Clarke & Garrett, 2004; Bradlow & Bent, 2008). However, in a cross-modal priming study, Witteman et al. (2013) found that the speed of perceptual adaptation is dependent on two factors, namely the strength of foreign accent and listener familiarity with the accent. This raises interesting questions regarding the online processing of gradient accentedness and the role of listener experience in that processing.

3.1.2 ERPs and early auditory processing

One avenue by which to study the processing of foreign-accented speech is the examination of event-related potentials (ERPs). Auditory-evoked electrophysiological responses can provide evidence of how listeners deal with accent-related variation during the earliest stages of speech processing. The N100 and P200 components (N100-P200 complex) have long been interpreted as indices of auditory processing because the properties of sounds and their contexts create a chain of components representing the adult auditory-evoked potential. While these two components co-occur, they can be dissociated and are linked to different generators and topographies, thus underlining that they represent different processes (Crowley & Colrain, 2004; Hillyard, Hink, Schwent, & Picton, 1973). Within the existing literature, this complex of waves is generally thought to represent preattentive, bottom-up processes.

The N100 component, which generally peaks between 50 and 150 ms, has been shown to modulate according to acoustic properties of stimuli (Näätänen & Picton, 1987). It is a negativity that is generally maximal when measured by midline central or temporal electrodes (Vaughan Jr. & Ritter, 1970; Näätänen & Picton, 1987) and has been shown to have multiple generators in the primary and secondary auditory cortices (Näätänen & Picton, 1987). Specifically it is sensitive to properties such as intensity, frequency, and other acoustic changes, indexing the arrival of information to the auditory cortex (Martin, Tremblay, & Stapells, 2007). The component may reflect a number of processes including attention to sound arrival, the reading out of sensory information from the auditory cortices, the formation of a stimulus sensory memory, or the facilitation of efficient sound processing (Näätänen & Picton, 1987). Additionally, it has been shown to modulate with oriented attention, state of arousal, and habituation. When attention is focused on the stimulus and during heightened states of arousal, the N100 has been shown to increase in amplitude (Hillyard et al., 1973;

Näätänen & Picton, 1987). Conversely, N100 amplitude has been shown to decrease over time as a result of habituation to stimuli (Näätänen & Picton, 1987; Ross & K. L. Tremblay, 2009).

With regard to language, the N100 is sensitive to stress patterns in word segmentation, voice onset time, and segmental context of assimilated speech sounds. Sanders and Neville (2003) showed that a word segmentation effect found in native English listeners is not present in Japanese speakers listening to English speech. They argue that this indicates that non-native listeners do not use the acoustic differences in the stimuli in the same way that native listeners do for the purposes of segmenting the speech signal. The N100 has been shown to vary with voice onset time of consonants (Steinschneider, Volkov, Noh, Garell, & Howard, 1999) and modulation of the component has been linked to perceptual learning (K. L. Tremblay, Kraus, McGee, Ponton, & Otis, 2001; K. L. Tremblay & Kraus, 2002; K. L. Tremblay, Shahin, Picton, & Ross, 2009). Additionally, there is evidence that the N100 reflects early categorization of assimilated speech sounds (Bien & Zwitserlood, 2013). While generally thought to reflect strictly bottom-up processing, a recent study suggests that expectation-driven phonological mismatch can be seen as early as the N100 (Brunellière & Soto-Faraco, 2014).

The P200 component of the N100-P200 auditory-evoked complex is associated with a number of processes similar to those of the N100; however the two can be dissociated and the P200 appears to reflect additional processes. Generated in Brodmann area 22 of the brain, it generally peaks between 150 and 250 ms, reaching its maximum over the vertex (Crowley & Colrain, 2004). While it is part of the auditory evoked response its functional significance is not totally understood. Having been linked to a number of processes in a variety of modalities, the P200 seems to index: attention (Bernal et al., 2000; Carretié, Mercado, Tapia, & Hinojosa, 2001; Wolach & Pratt, 2001); item encoding and memory retrieval (Stelmack, Saxe, Noldy-cullum, Campbell, & Armitage, 1988; Curran & Dien, 2003); word and speech sound repetition (Evans & Federmeier, 2007; Misra & Holcomb, 2003; Ross & K. L. Tremblay, 2009); auditory stimulus familiarity (Shahin, Bosnyak, Trainor, & Roberts, 2003; Shahin, Roberts, Pantev, Trainor, & Ross, 2005); and auditory perceptual training (Atienza, Cantero, & Dominguez-Marin, 2002; Sheehan, McArthur, & Bishop, 2005; Tong, Melara, & Rao, 2009; K. L. Tremblay et al., 2009).

With regard to auditory stimuli, the P200 displays a relative increase in amplitude to the repetition of speech sounds (Ross & K. L. Tremblay, 2009; Shahin et al., 2005). Additionally, training on specific speech sounds leads to a similar increase in amplitude (cf. Sheehan et al., 2005; Tong et al., 2009; K. L. Tremblay et al., 2009). This effect can be seen even when no explicit response is required of participants (Ross & K. L. Tremblay, 2009; Shahin et al., 2003; Shahin et al., 2005). Importantly, this effective increase in amplitude is sometimes seen after a day or more interval between recording sessions indicating a longer term change in processing (Atienza et al., 2002; Sheehan et al., 2005; K. L. Tremblay et al., 2009).

It is hypothesized that the P200 indexes the comparison of perceptual aspects of an auditory stimulus to stored memory traces, reflecting enhanced perceptual discrimination and matching processes. Given that the response of the component can change over time and with training, it can be described as plastic and thus may represent neural reorganization. This experience seems to facilitate rapid and preattentive access to perceptual representations which build up over time and may require sleep in order to consolidate.

The Phonological Mapping Negativity, or PMN, is another auditory-evoked response. This component is a negative-going wave which reaches maximum amplitude between 250 and 350 ms post stimulus onset and generally has a midline fronto-central (Desroches, Newman, & Joanisse, 2009; Lee, Harkrider, & Hedrick, 2012) or centro-parietal (Hagoort & Brown, 2000) distribution. Seen only in the auditory modality, it is believed to correspond to an early, phonological stage of word processing that maps phonetic input onto phonological representations and early acoustic-phonetic mechanisms in speech comprehension (Connolly & Phillips, 1994; Newman, Connolly, Service, & McIvor, 2003; Newman & Connolly, 2009; Steinhauer & Connolly, 2008). Moreover, the PMN has been described as a “goodness of fit” measure between the expected phonology and the acoustic–phonetic input contained in the signal (Newman & Connolly, 2009).

This component has predominantly been studied in relation to word-initial, single phoneme manipulations in sentential context, priming, and phoneme deletion tasks. The

Referred to in previous literature as the Phonological Mismatch Negativity, the N200 and the N250.

negativity becomes pronounced (i.e. increased amplitude) for stimuli which do not match an established phonological expectation. With regard to native speech, it appears that the component is an all-or-nothing response, as multiple mismatches produce equally large negativities as single segment mismatches (Newman et al., 2003).

Once thought to be an early N400 or a component of similar function in the lexical selection process (cf. Hagoort & Brown, 2000), the PMN has been shown to be functionally distinct from the N400 (Connolly & Phillips, 1994). It does not appear to be influenced by the activation of lexical competitors (Newman et al., 2003) nor by semantically related items (Connolly & Phillips, 1994; Diaz & Swaab, 2007). Additionally, this component is not influenced by the lexical status of the target being elicited by nonwords (Connolly, Service, D'Arcy, Kujala, & Alho, 2001, 2; Newman & Connolly, 2009) and nonsense syllables (Lee et al., 2012).

ERP components such as these can therefore be useful in the examination of early processing of non-native speech signals. Linguistic processing has been shown to occur early on in the unfolding of the acoustic signal (cf. Kryuchkova et al., 2012). Foreign-accented speech may thus affect this early processing in listeners, which has not yet been thoroughly explored, particularly in regard to gradient foreign accentedness.

3.1.3 ERPs and foreign-accented speech

While ERPs have not been used extensively to investigate foreign-accented speech, two recent ERP studies have found differences in processing between native and non-native speech (Hanulíková et al., 2012; Goslin et al., 2012). While the studies examine different aspects of language, they both illustrate the need for further investigation of foreign-accented speech processing. Hanulíková et al. (2012) found differences in the P600 component for grammatical violation made by native and non-native talkers. In their study a native Dutch speaker and a non-native Dutch speaker (native Turkish) produced gender agreement errors in spoken sentences. As expected, a P600 effect was seen for gender violations in native speech, while no P600 was seen for the same gender violations produced by the non-native (accented) speaker. Additionally they examined sentences containing semantic violations and found similar N400 effects for both talkers, indicating there was no general integration problem for world knowledge in the L2 speech. The authors concluded that the lack of a P600 effect indicates that listeners use foreign accent as a cue for non-nativeness and may subsequently modify their expectations of errors as a function of speaker characteristics.

Examining earlier effects, Goslin et al. (2012) investigated normalization mechanisms used when processing sentences spoken with either a regional accent or a foreign accent. Their results showed that the PMN elicited by the final word of sentences spoken with an unfamiliar regional accent was greater than for those produced in the listener's own accent, while the PMN for foreign accented speech was not present. Foreign accents also resulted in a reduction in N400 amplitude when compared to both unfamiliar regional accents and the listener's own accent. The authors claim that regional accent-related variations are normalized at the earliest stages of spoken word recognition such that it does not impact lexical access. Conversely, foreign-accented speech does not show the same normalization and thus influences later top-down lexical processing.

Goslin et al. (2012) argue that listeners are able to employ mechanisms at a prelexical processing level to normalize speech coherent with their own native language system (i.e., regional accent), such that lexical processing appears relatively unperturbed. On the other hand, speech variation not consistent with the listener's native language system (i.e., foreign accent) does not appear to be effectively normalized. The unreliable phonetic/acoustic information therefore requires increased reliance upon top-down contextual cues, thus reducing lexical activation (indexed by the N400). The authors frame their results in relation to two hypotheses of foreign-accented speech processing: the Different Processes Hypothesis and the Perceptual Distance Hypothesis. The Different Processes Hypothesis states that there are functional differences in the processing and representation of native and non-native accents. The Perceptual Distance Hypothesis on the other hand states that the underlying processing mechanism is the same and that differences between native and non-native accent variation will manifest as a function of perceptual distance from the home accent. The authors conclude that their results are in line with the Different Processes Hypothesis because PMN amplitude is not greatest for the foreign-accented speech which is the most divergent acoustically. However, they do not address the topic of expectation, which is likely involved in the modulation of the PMN (cf. Diaz & Swaab, 2007).

While these two studies show that processing of foreign-accented speech may indeed differ from native speech, they do present contradictory results. Additionally, these studies do not fully address the issue of gradient accentedness. Goslin et al. (2012) attempt to address perceptual distance by including home, regional-, and foreign-accented speech, but do not take into consideration the role of listener experience in the processing of that speech.

3.1.4 The present study

This study aims to examine the roles of both degree of foreign accentedness and listener experience with the accent in question. In particular, are there early auditory processing differences along the accentedness continuum when listening to native English and Chinese-accented speech? Additionally, does experience with Chinese-accented speakers influence early auditory processing of gradient foreign accentedness?

Given the data from reaction time studies, along with the limited extant ERP studies, we expect that increased degree of foreign accentedness will modulate ERP components associated with early auditory processing. Also, given the data regarding perceptual adaptation to foreign-accented speech, we expect that increased listener experience with Chinese-accented English will result in processing changes across variation in accentedness. Therefore, we expect that waveform differences will emerge in the regions of 100-200ms (N100), 200-300ms (P200), and 250-350ms (PMN) along the continuum of foreign accentedness and for different levels of listener experience with Chinese-accented speakers.

3.2 Methodology

3.2.1 Participants

Twenty-four participants (female n = 16, male n = 8) were recruited from the University of Alberta campus area and ranged in age from 18 to 29 years old (M = 20.7, SD = 3.03). They received $25 for their participation in the study. All reported having normal hearing and being native speakers of North American English. Of the 24, three reported being left-hand dominant.

3.2.2 Materials

Recordings of nine male native Chinese speakers and one male native English speaker were retrieved from the NU Wildcat Corpus of native- and foreign-accented English (Van Engen et al., 2010). Each recording contained a word list read three times by a given talker and a subset of that list (40 monosyllabic words) was used for this study. Here, the first repetition was chosen and extracted to serve as the individual stimuli for this study. For more information regarding the segmentation and extraction of the stimuli please see Porretta and Tucker (2012) or Porretta et al. (2013). The sound files were normalized for amplitude across words and speakers. The stimuli were blocked by talker, beginning with the native English talker. The presentation of the non-native talkers was arranged so that accentedness would generally increase based on the holistic ratings available in the Wildcat Corpus (Van Engen et al., 2010). This was done so that the words would generally become more difficult to understand as the experiment progressed and counteract any effect of priming from previous tokens. However, it was also necessary to make an attempt to counterbalance talker order. Therefore, the nine non-native talkers were arranged in order of increasing accentedness and then split into three groups of three. The arrangement of talkers within these groups was then counterbalanced to produce six presentation orders. This allowed for talker order to be different across participants, yet still generally increasing in accentedness throughout the experiment. Within each block, words were pseudo-randomized. Ten filler words from the native English talker were placed between blocks and flanked with basic arithmetic problems. The arithmetic problems served as a distractor task and the filler items were included as an accent "reset" to minimize the effect of the previous accented talker on the subsequent.

3.2.3 Procedures

Participants were seated in front of a computer screen and keyboard in an electromagnetically shielded and sound-attenuated booth. Stimuli were presented binaurally using ER-1 insert earphones (Etymotic Research, Inc.), as these do not interfere with EEG recording. E-Prime 2.0 (Psychology Software Tools, Inc.) was used for stimulus presentation and synchronization with the ERP recording system. Participants were instructed to stare at a white fixation cross in the middle of a black screen which appeared at the beginning of each trial. Their task was to pay careful attention to the stimuli in order to understand the words as they were spoken. An inter-stimulus interval ranging from 500–1000 ms was used to allow the ERP signal to return to baseline. Short, self-paced breaks were allowed between each block.

Participants were fitted with a nylon BioSemi cap with 32 Ag/AgCl active electrodes (10/20 layout). Additionally, two electrodes were placed on the left and right mastoids for off-line re-referencing. Two electrodes were also placed at the outer canthi of the right and left eyes for recording horizontal electro-oculograms, while two others were placed above and below the left eye for recording vertical electro-oculograms. The EEG signal was recorded using a BioSemi Active II digital amplification system with an input range of -262 μV to +262 μV per bit and the signal was band-pass filtered on-line from 0.01 to 100 Hz. For each stimulus item, the EEG signal was recorded starting 200 ms prior to the onset of the stimulus and continued for 1300 ms post onset.

After the EEG portion of the experiment was completed, participants responded to a series of demographic questions (see Table A.2 in the Appendix) including details of language experience. Language experience information included native and second language, along with amount of experience interacting with non-native speakers and particularly speakers with a Chinese accent.

The digitized EEG signal was pre-processed with Brain Vision Analyzer software (version 2.0; Brain Products, Germany). The 32 channels were first re-referenced to the left and right mastoid electrodes and the signal was then downsampled from 8192 Hz to 128 Hz A band-pass filter was applied from 0.05 to 30 Hz using a Butterworth filter (time constant = 0.3183). Additionally, the signal was baseline corrected over the 200 ms preceding stimulus onset. Finally, the EEG data were exported from Brain Vision Analyzer for further processing in R, version 3.1.0 (R Development Core Team, 2014). Ocular movements were corrected using the icaOcularCorrection package (A. Tremblay, 2011).

3.3 Analysis and results

3.3.1 Analysis

The EEG data were analyzed using generalized additive mixed modeling (henceforth, GAMM) (Hastie & Tibshirani, 1990; Wood, 2006). GAMM has previously been applied successfully to model a variety of linguistic data. In particular, it has been used for investigating dialectal variation (Wieling et al., 2011; Wieling et al., 2014), prosodic prominence (Arnold et al., 2013), reaction times (Baayen, 2010a; Baayen, 2010b), and importantly event-related potentials (Kryuchkova et al., 2012; A. Tremblay & Baayen, 2010). GAMM has a number of advantages in this regard. First, GAMM does not assume a linear relationship between predictors as does ANOVA or linear models and thus it more readily handles non-linearities, something which is inherent to EEG data. Second, GAMM allows one to model the surface of an interaction between two numeric predictors. This point is particularly important for the present study due to the working conception of gradient foreign accentedness. Because accentedness is thought to be gradient, GAMM allows us to model the, possibly wiggly, surface representing the change in voltage as one moves through time and across accentedness values. Third, given the time-series nature of EEG data, GAMM allows for the control of autocorrelation in the data. Autocorrelation deals with the correlation between datapoints in a time-series; a measurement at timepoint t is correlated to differing degrees with a measurement at timepoint t-i, depending on the lag. The analysis was carried out in R (R Development Core Team, 2014) utilizing mgcv, version 1.7-29 (Wood, 2014) and eRp, version 0.9.29 (A. Tremblay, Baayen, & Hendrix, 2013) packages.

Grand average

Prior to modeling, we inspected the grand average waveforms (over participants and items) at all electrode sites from 0 to 500 ms. These are presented together in Figure 3.1. As can be seen, a stereotypical auditory-evoked response is present, primarily over fronto-central and centro-parietal sites. The N100 is prominent around electrode CP1, peaking at approximately 156 ms. The P200 is prominent around Fz, with its peak at approximately 242 ms. This

Figure 3.1: Grand Average at each of the 32 electrode sites.

is consistent with previous descriptions of the complex. The N100-P200 complex is then followed by the PMN peaks at approximately 344 ms around electrode Cz. Again, this is consistent with its description.

Predictor variables

Accent Rating: Prior to the present experiment, accentedness ratings were collected from 30 native English-speaking participants on the 40 words spoken by each of the ten talkers (total of 400 items). This ratings task was performed separately from the EEG study by different listeners. The participants in the ratings task listened to each item and judged it on a scale from 1 to 9 in which 1 represents no foreign accent (i.e., completely native production) and 9 represents a strong foreign accent (i.e., highly non-native production). From these ratings, item and talker means were calculated to serve as measures of accentedness associated not only with each talker but with each experimental item. This was done to provide a more fine-grained measure of accentedness because it is possible that a talker produces different lexical items with varying degrees of accentedness. These item ratings have been shown to be highly correlated with measures of acoustic distance (Porretta et al., 2013) and when averaged over talker correlate significantly ($r(9) = 0.9$, $p < .001$) with the holistic accentedness rating for each talker available in the Wildcat Corpus (Van Engen et al., 2010). This indicates that the raters were likely using similar acoustic information for making their judgments as those in the corpus.

Experience: This variable was calculated from two questions in the questionnaire: 1) "On a weekly basis, how often do you interact with non-native speakers of English?" and 2) "What percentage of those interactions include speakers with a Chinese accent?" The participants were asked to respond to the first question on a scale of 0 (Never) to 10 (Daily). To the second question, participants responded with a percentage. These two values were used to estimate the amount of experience each participant currently had interacting with Chinese-accented speakers. This was done by first multiplying the response to question one by 10 and dividing the response to question two by 100. These two values were then multiplied together. The final value was then multiplied by 100 to obtain a value ranging between zero and 100 (see

Figure A.1 in the Appendix). This experience variable was split into quantiles and then grouped (quantile 1 and quantiles 2–4). This produced a factor (Experience) with two levels which we refer to as Low and High. This allowed for an empirical definition of Experience level based on the sample rather than an predefined criterion for determining High versus Low. Therefore, the Low Experience group represents those participants who have little-to-no experience interacting with Chinese-accented speakers and the High Experience group represents those with moderate-to-considerable experience. In the subsequent analysis, this factor variable is referred to as Experience.

Time: Time in milliseconds served as an independent variable. This allows us to model the entire waveform over the period of interest, in this case 100–400 ms, capturing the three components of interest.

Cartesian coordinates of the electrodes: The Cartesian coordinates of each electrode site were included as predictors X and Y. This allows us to model the entire topography of the scalp through Time in a single model. In doing so, it is possible to avoid multiple comparisons among electrode sites.

Response variable

Mean Amplitude: Mean Amplitude served as the dependent variable. Averaging was carried out within the two Experience groups. Within each group, amplitude was averaged over participants by Item, Electrode, and Time. This was done to help reduce noise in the signal as well as to make the model more computationally tractable. Averaging was not done over items as this would have required discretizing accentedness, which is not in line with the research questions of this study.

Model fitting and evaluation

The input variables described above were fit to the response variable (Mean Amplitude) with by-Item random intercepts. By-Item random intercepts allow for the possibility that different items may have different baseline values at the beginning of the time-series. An interaction between Rating and Time was included, using a thin plate regression spline with restricted maximum likelihood to optimize the number of smoothing parameters. This tensor product allows the interaction to have a nonlinear functional relationship with the response variable and is suitable for modeling interactions between variables on different scales (Wood, 2006). For the interaction of Rating and Time, the Cartesian coordinates of the electrodes were included which allows the response variable to vary as a function of the spatial location of the electrodes. Finally, to this tensor product the factor variable of Experience was included. Thus the model consists of the following predictors: Random intercepts for Item and an interaction between Time, Rating, X and Y, by Experience. Additionally, an AR-1 correlation parameter, $\rho = 0.917$, was included to control for autocorrelation in the timeseries data.

The model was fit using a backwards step-wise elimination procedure (cf. Zuur et al., 2009). The inclusion of predictors in the model was evaluated using two criteria. The first criterion was the estimated *p*-value of the smoothing parameter or parametric component. The estimated *p*-value of the smoothing parameter or parametric component indicates whether or not the functional form of the predictor is different from zero. If greater than the conventional alpha level of 0.05, the predictor was considered for removal. The second criterion was Maximum Likelihood (ML) score comparison (Zuur et al., 2009). Here ML comparison is preferred over AIC as AIC becomes less reliable (i.e., anti-conservative) when autocorrelation is taken into account.

Upon removal of the predictor, the ML score of that model was compared to that of the model containing the predictor in question (i.e., the full model), indicating whether or not the inclusion of the predictor significantly improves model likelihood. Through this process, no predictors were eliminated from the model.

The model was refitted for data within -/+2.5 standard deviations of the residuals of the model (1.25% removed, 12508 datapoints) (cf. Baayen, 2008). Re-inspection indicated that the residuals were approximately normally distributed.

Additionally, ML score comparison was done in order to evaluate the terms of interest included in the full model (see Table 3.1). Here, the Chi-square statistic indicates a significant difference in likelihood between the models, given the degrees of freedom, and *p*-values below the conventional alpha level of 0.05 indicate evidence for the more complex model (i.e., the full model). In all cases, the full model including Item, Time, Rating, X, Y, and Experience is the most likely.

Table 3.1: Maximum Likelihood comparison between full model and model variants (simpler model) for EEG model.

Model	Score	Edf	Chisq	Difference	p-value
Full model	1.4488×10^6	31	NA	NA	NA
w/o Experience	1.4491×10^6	17	351.7	14	< 0.0001
w/o Rating	1.4488×10^6	17	61.9	14	< 0.0001

Again, after evaluation the full model was refit with method set to REML (Restricted Maximum Likelihood) and trimmed as above (cf. Zuur et al., 2009). This final model, reported below, accounts for 22.6% of deviance explained showing that the model is able to capture important facets of variation in the signal over time.

3.3.2 Results

Model

The results of the final model are reported below along with visualizations of the effects (both topography and time). The coefficients of the model are found in Table 3.2.

Table 3.2: Generalized additive mixed model reporting parametric coefficient (Part A) and estimated degrees of freedom (edf), reference degrees of freedom (Ref.df), F and p values for the tensor products and random effects (Part B) for EEG model.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
intercept	0.1490	0.0608	2.4498	0.0143
Experience	-0.3345	0.0211	-15.8233	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
by-item random intercepts	386.8200	398.0000	22.7060	< 0.0001
tensor: Time, Accent, X, Y, by Experience: High	207.9129	281.1615	60.6997	< 0.0001
tensor: Time, Accent, X, Y, by Experience: Low	309.8508	401.6387	37.0355	< 0.0001

Topography

Figure 3.2 presents scalp topographies at three points in time. In the left-hand panel we see the topography at 156 ms, which corresponds to the N100 component. Here, the negativity (indicated by dark gray) has a slightly left-lateralized distribution and peaks around electrode CP1. In the middle panel is the topography at 242 ms, corresponding to the P200 component.

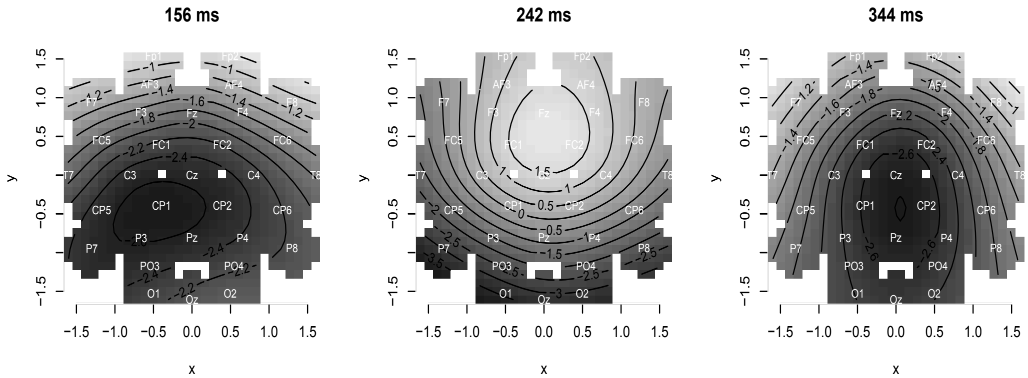

Figure 3.2: Scalp topography at 156 ms, 242 ms, and 344 ms. Darker shading indicates decreased amplitude while, lighter shading indicates increased amplitude.

The positivity (indicated by light gray) is distributed over the fronto-central region, peaking at electrode Fz. The right-hand panel shows the topography at 344 ms and corresponds to the PMN component. The negativity (indicated by dark gray) has a slightly posterior central distribution and peaks near electrode Cz.

The regions indicated here coincide with the peak amplitudes present in the grand average waveforms seen above in Figure 3.1. These are more or less in line with the locations at which the N100, P200, and PMN components have been reported to reach their maximum amplitude (Näätänen & Picton, 1987; Crowley & Colrain, 2004). The topographic results here indicate the electrode sites at which the amplitude peaks occurred. Therefore, based on this, we further examine the temporal waveforms at three electrode locations: CP1 for effects on the N100 component, Fz for effects on the P200 component, and Cz for effects on the PMN component.

Time

Below we present visualization of the effects of Rating and Experience through Time. For each component of interest, we examine Amplitude for the specific time period related to the component at the electrode site corresponding to its peak.

N100: Figure 3.3 presents Amplitude from 100 to 200 ms across Accent Rating for both high and low experience groups at electrode CP1. Looking at the left panel (High Experience), we see that Amplitude gradually decreases as accent becomes stronger. From the right panel (Low Experience), we see a very different pattern in which Amplitude increases slightly and remains relatively constant across the accentedness continuum. However, for items having a high accent rating, the peak appears to shift slightly forward relative to word onset.

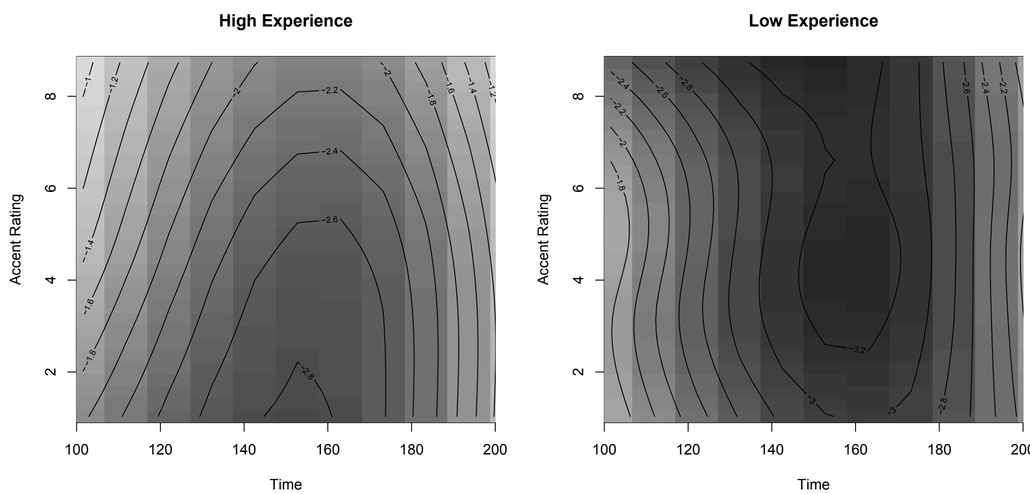

Figure 3.3: Contour plot of effect of Accent Rating over Time by Experience at electrode CP1. Darker shading indicates decreased amplitude while, lighter shading indicates increased amplitude.

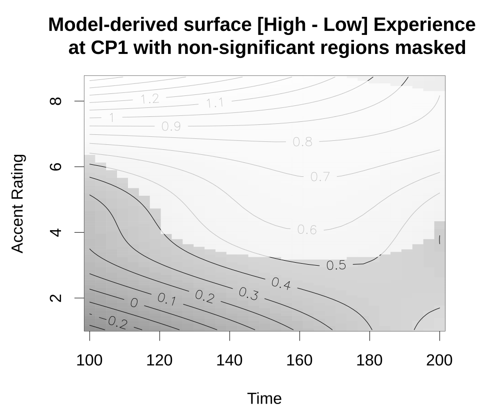

Figure 3.4: Model derived difference surface at CP1.

Figure 3.4 presents a difference surface (High experience minus Low experience) for the predicted values at CP1 through time along the accentedness continuum. As can be seen, there is a significant difference in performance between the two groups along the continuum of accentedness and time, indicating a gradual increase in N100 amplitude for the

Low Experience group as Rating increases. The shading represents non-significant regions, showing that the increase is significant for items in the native-like portion of the continuum, as well as items up to rating 6 in the earliest portion of the time window.

P200: Figure 3.5 presents Amplitude from 200 to 300 ms across Accent Rating for both high and low experience groups at electrode Fz. For the High Experience group (left panel), the model indicates little effect of Accent Rating with only a slight decrease in Amplitude for items at the low end of the continuum. Looking at the Low Experience group (right panel), again, the pattern is rather different. Here we see that moving from lowest accent rating to highest, Amplitude first rises, then falls, producing a peak in the middle of the continuum.

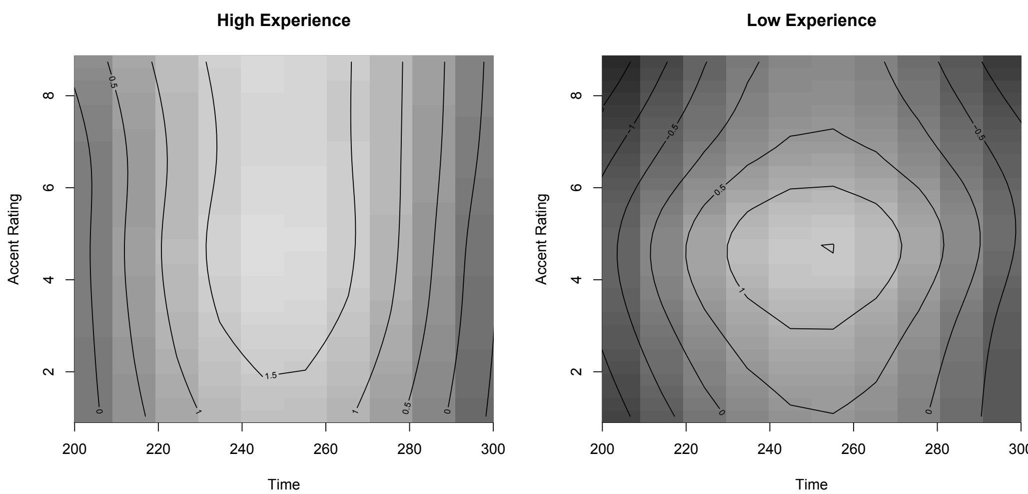

Figure 3.5: Contour plot of effect of Accent Rating over Time by Experience at electrode Fz. Darker shading indicates decreased amplitude while, lighter shading indicates increased amplitude.

Figure 3.6 presents a difference surface (High experience minus Low experience) for the predicted values at Fz through time along the accentedness continuum. As can be seen, there is a significant difference in performance between the two groups along the continuum of accentedness and time, with High Experience group exhibiting greater P200 amplitude as Rating increases. However, this difference narrows in the midrange of the ratings continuum. The shading represents non-significant regions, showing that it is the midrange and in the most native-like items, that the difference is significant. The pattern in this significant region

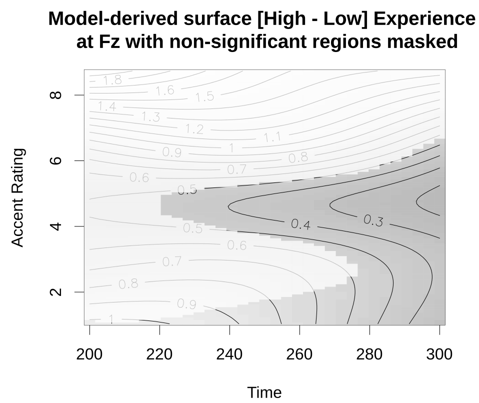

Figure 3.6: Model derived difference surface at Fz.

is an effect of time; that is to say, the High Experience group has greater amplitudes earlier in time, and the difference decreases moving through the time window.

PMN: Figure 3.7 presents Amplitude from 300 to 400 ms across Accent Rating for both high and low experience groups at electrode Cz. For the High Experience group (left panel), the model indicates an effect of Accent Rating such that Amplitude is relatively constant for rating values up to 4. Beyond this, Amplitude gradually decreases. Looking at the Low Experience group (right panel), an opposite pattern emerges, such that amplitude gradually increases moving from the most native-like items up to a rating value of 4. Beyond this point, Amplitude remains relatively constant, with a slight decrease around a rating value of 7.

Figure 3.8 presents a difference surface (High experience minus Low experience) for the predicted values at Cz through time along the accentedness continuum. As can be seen, there is a significant difference in performance between the two groups along the continuum of accentedness and time. The shading represents non-significant regions, showing that the majority of the surface is significant. Differences in PMN amplitude are greatest at the edges of the rating continuum showing gradual shifts moving toward the midrange. For items with high ratings, the Low Experience group has greater amplitudes, while for items with low ratings, the High Experience group displays greater amplitudes. It is in the midrange that the difference is an effect of time, rather than rating. Here, the amplitude of the Low Experience group falls away faster than that of the High experience group, though this difference does not remain significant moving farther in time.

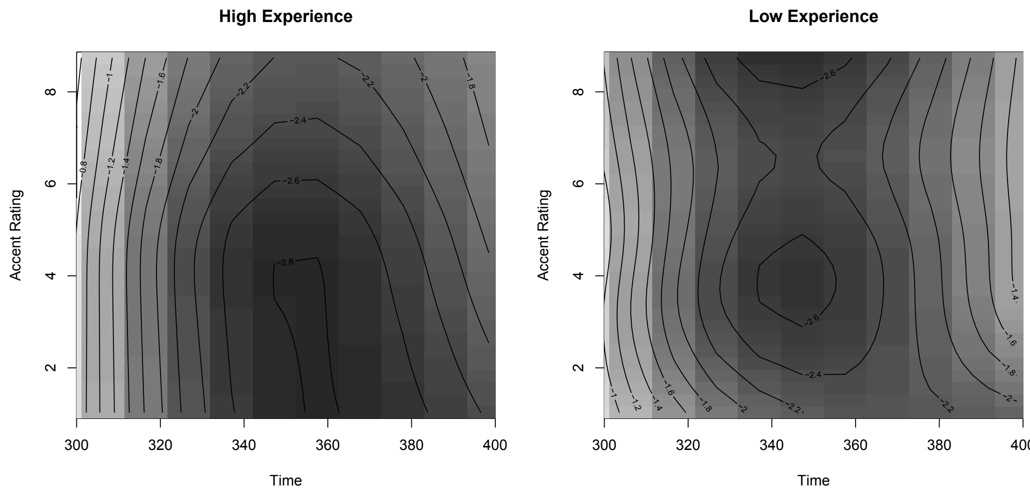

Figure 3.7: Contour plot of effect of Accent Rating over Time by Experience at electrode Cz. Darker shading indicates decreased amplitude while, lighter shading indicates increased amplitude.

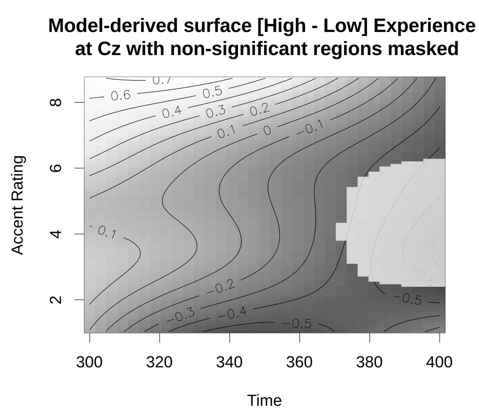

Figure 3.8: Model derived difference surface at Cz.

3.4 Discussion

The results which emerge from this experiment indicate that early auditory processing of spoken words varying in degree of foreign accentedness differs both along the continuum of accentedness and between groups with different levels of experience interacting with Chinese-accented speakers. The High Experience group (i.e., comprising listeners with moderate-to-considerable experience interacting with Chinese-accented speakers) shows a rather different pattern of processing than the Low Experience group (i.e., comprising listeners with little-to-no experience interacting with Chinese-accented speakers). These differences are persistent and extend over the entire window of analysis from 100 to 400 ms post-stimulus onset, with component-specific differences evident in three separate sub-windows at electrode sites corresponding to component peaks. The differences between the groups' response to the continuum indicates differing allocation of processing resources in the face of variable auditory input.

The N100-P200 auditory-evoked complex thought to represent pre-attentive and bottom-up perceptual processing. While evoked together, the two components may represent different processes. The N100 component is thought to reflect a number of processes including attention to auditory stimuli and perceptual processing of acoustic features. An increase in N100 amplitude has been shown to occur when attention is focused on the stimulus (Hillyard et al., 1973) and during heightened states of arousal (Näätänen & Picton, 1987). It has also been shown that N100 amplitude decreases over time as a result of habituation to stimuli (Näätänen & Picton, 1987; Ross & K. L. Tremblay, 2009). Additionally, the N100 has been linked to language specific function as it is sensitive to stress patterns for word segmentation (Sanders & Neville, 2003), voice onset time (Steinschneider et al., 1999; K. L. Tremblay et al., 2001; K. L. Tremblay & Kraus, 2002; K. L. Tremblay et al., 2009), segmental context of assimilated speech sounds (Bien & Zwitserlood, 2013), and expectation-driven phonological mismatch (Brunellière & Soto-Faraco, 2014).

Here, the effect of the accentedness continuum on the N100 is quite different for the two groups. The High Experience group displays a gradual decline in amplitude across the continuum while the Low Experience group shows increased amplitude for tokens outside the most native-like. Modulation of this component indicates that the groups may allocate perceptual and attentional processing resources differently for variable phonetic productions. The N100 for the High Experience group attenuates, perhaps do to habituation during the listening session. Conversely, the N100 for the Low Experience group increases in amplitude for all but the most native-like tokens, perhaps as a result of heightened attention focused on processing the variability present in the stimuli. These group differences are significant primarily between ratings 1 and 4 which comprise the region of the continuum in which the native talker's rating values fall. Therefore processing demands differ even for native-like tokens.

The P200 component of the N100-P200 auditory-evoked complex is associated with a number of processes similar to those of the N100. This component, however, can be dissociated from the N100 and has been shown to index language related perceptual processes which are sensitive to repetition, experience, and discrimination training (Atienza et al., 2002; Ross & K. L. Tremblay, 2009; Shahin et al., 2003; Shahin et al., 2005; Sheehan et al., 2005; Tong et al., 2009; K. L. Tremblay et al., 2009). The results presented here indicate that the effect of the accentedness continuum on the P200 is quite different between the the High and Low Experience groups. The High Experience group displayed a rather uniform amplitude across the continuum with a slight decrease in amplitude for the most native-like items. The Low Experience group on the other hand shows a similar increase moving up the ratings continuum, however this does not persist.

The patterns presented by the two groups seem to align with previous findings of both repetition and experience. Given that different tokens of the same lexical items were repeated as the experiment progressed and that the productions of the native talker were always presented first, it stands to reason that the amplitude of the P200 would increase outside of the most native productions. For the High Experience group, this indicates that tokens presented later were considered repetitions. However, for the Low Experience group, only items through the midrange of the continuum produced an effect of repetition, indicating that the more highly accented items were not registered as repetitions. Repetition effects have primarily been seen for visually presented words (Curran & Dien, 2003; Evans & Federmeier, 2007; Misra & Holcomb, 2003) and for spoken syllables (Ross & K. L. Tremblay, 2009; Sheehan et al., 2005). However, Curran and Dien (2003) found that in a visual recognition task, the amplitude of the P200 was not increased when the word had been previously presented in auditory form, but was when the word had been previously presented in visual form. This may indicate that repetition effects are modality consistent.

Here it appears that the High Experience group has learned over time that the acoustic variability associated with more highly accented tokens can still be mapped onto perceptual memory representations. Alternatively, real-world experience interacting with Chinese-accented speakers has resulted in implicit learning for the discrimination of acoustic features. While the differences between the groups is not significant across the entirety of the accentedness continuum, the fact that a difference exists, underscores previous findings that experience with complex acoustic stimuli can modulate the P200 response (cf. Shahin et al., 2003; Shahin et al., 2005).

The PMN is an auditory-evoked potential linked to early acoustic-phonetic mechanisms in speech comprehension and more specifically to an early, phonological stage of word processing that maps phonetic input onto phonological representations (Connolly & Phillips, 1994; Newman et al., 2003; Newman & Connolly, 2009; Steinhauer & Connolly, 2008). It has also been likened to a "goodness of fit" measure for bottom-up auditory speech processing (Goslin et al., 2012; Newman & Connolly, 2009). The negativity increases in magnitude when mismatch occurs between the expected phonological form of a word and the acoustic input.

The present results seem to agree with this notion but only for listeners in the Low Experience group. Having little-to-no experience interacting with Chinese-accented speakers results in the negativity becoming more pronounced for productions moving away from the native-like end of the accentedness continuum. Higher ratings then evoke a relatively stable negativity. This is in line with previous descriptions of the PMN which have shown that a single segmental mismatch is enough to elicit the PMN and that multiple mismatches do not increase its amplitude (Newman et al., 2003).

The results for the High Experience group seem to point to a different interpretation of the functional significance of the PMN. The response of the PMN across the accentedness continuum is rather opposite that of the Low Experience group, as is clear from the pervasive significant difference between the two surfaces. Having moderate-to-considerable experience interacting with with Chinese-accented speakers results in a stable negativity up to approximately halfway through the continuum, after which point the amplitude of the negativity gradually decreases. If the PMN indexes mismatch, for this group we might expect a similar pattern as that seen for the Low Experience group, though shifted up the continuum so that an increase in amplitude would be seen only for the most accented items. It has been suggested that this negativity may be reduced in the presence of noise which diminishes discriminability (Martin, Sigal, Kurtzberg, & Stapells, 1997), which in this case might reflect the influence of accented-related “noise" on disciminabilty of segments (cf. Goslin et al., 2012). However, this does not explain the pattern seen for the Low Experience group.

The vast majority of research on the PMN has focused its elicitation in sentential or priming contexts. Thus, its presence is highly driven by contextual or experimental expectations of phonological form, both of which can affect its presence and behaviour. Diaz and Swaab (2007) found that manipulating the phonological and semantic constrains of word lists altered the presence of the PMN. When strong phonological constraints were violated, an early negative shift is obtained. However, when phonological constraints were met or when no phonological expectancy could be generated, then no early effect emerged. When elicited in sentential context, an early negative shift was observed in all conditions, though it differed for high and low predictability of the final word.

In this experiment, there was neither sentential context, nor a primed phonological context. Therefore, the results provide additional perspective on the functional significance of the PMN. The negativity can be elicited during single word passive listening; however, it appears to be a measure of expectation rather than an index of bottom-up acoustic-phonetic mapping. Here, it seems that the two Experience groups have very different expectations about the variability inherent to gradiently accented words. Where the Low Experience group expects non-native-like productions to mismatch their representation, the High Experience group seems to relax expectations as words become more highly accented. If the PMN were to index goodness-of-fit during bottom-up auditory processing, we would expect to see a gradual increase in amplitude as accentedness increases. Additionally, we would expect to see the same pattern, albeit shifted, in both Experience groups.

In a recent study, Goslin et al. (2012) examined the PMN as an index of goodness-of-fit between three accent groups: home, regional, and foreign. The authors found that in comparison to speech presented in the home accent only regional-accented speech produced an increase in PMN amplitude while foreign-accented speech elicited no PMN response at all. They interpreted this as a difference in the speech normalization mechanisms employed by listeners. They framed their results around the Different Processes Hypothesis which states that native and non-native speech will be processed by different mechanisms as non-native speech cannot be prelexically normalized in the same way that native speech (accented or not) is.

While it is difficult to make comparisons between Goslin et al.'s study and the present one due to methodological differences, the lack of PMN to foreign-accented speech found in their study may be due to listeners relaxing phonological expectations when presented with sentences that were clearly spoken by a non-native talker. Another ERP study on the processing of foreign-accented speech showed that when listeners process errors in grammatical gender produced by non-native talkers, the P600 component (which indexes grammatical violations) is not elicited (Hanulíková et al., 2012). Hanulíková et al.'s results indicate that listeners are highly adept at adjusting expectations about the speech produced by non-native talkers. The present study adds to this by showing that both degree of accentedness and listener experience with the accent in question may also drive earlier expectations about phonological form. The results from the current study thus present a more complex view of the early stages of foreign-accented speech processing. It appears that the acoustic variation inherent in foreign-accented speech interacts with listener experience and the representations that result from that experience.

The current results do not coincide strongly with either of the hypotheses presented by Goslin et al. (2012) and seem to indicate that experience may in fact play a larger role in how listeners process accent variability than previously thought. There appears to be some support for the Perceptual Distance Hypothesis in that degree of foreign accentedness does in fact modulate the amplitude of the three ERP components examined. However, in the case of both the N100 and the PMN, the patterns of gradience along the continuum are not consistent for both levels of Experience. Additionally, these components may actually be indexing cognitive processes other than those directly related to the evaluation of the acoustics contained in the signal.

The differences that emerge between the two Experience groups might lead one to favor the Different Processes Hypothesis. However, this hypothesis is related specifically to the normalization of variability contained in the speech signal and in particular draws a clear line between native and non-native speech. The present results do not indicate clear delineation of the processing of native and non-native tokens. In fact, tokens produced by the native talker span a range of values at the lower end of the continuum, which overlap with values assigned to some non-native tokens. Therefore, we do not see differentiation in either Experience group for an effect of native vs. non-native tokens. If these components were to index a general speech normalization process, one would expect see no effect of listener experience at all. This is clearly not the case in the present data.

The differences that we see between the Experience groups across the continuum, are perhaps more likely due to general cognitive processes. For example, the High Experience group appears to display effects of attention and habituation (as seen in the N100), repetition (as seen in the P200), and relaxation of expectation (as seen in the PMN). The Low Experience group displays rather different patterns, perhaps due to their lack of previous exposure to the variability inherent to Chinese-accented English.

With regard to the numerous studies reporting behavioral results for foreign-accented speech, the present study provides some additional perspective on how this variation is processed and represented in listeners. Previous studies have shown that listeners can adapt to foreign-accented speech over short- and long-term exposures (Gass & Varonis, 1984; Bradlow & Bent, 2003; Clarke & Garrett, 2004; Bradlow & Bent, 2008; Witteman et al., 2013) and the results of this study show that this ability to adapt can take place outside of an experimental setting. Listeners with greater experience interacting with Chinese-accented speakers in a natural environment display a drastically different pattern in electrophysiological response, which may be an indication of long-term adaptation to accent-specific variability. However, further investigation in this regard is still needed to understand how involuntary brain activity correlates with this type of adaptation.

So, it seems that experience with a particular accent affects how listeners to deal with fluctuating levels of variation in the speech signal. This would indicate that High Experience listeners have developed representations that change the way in which accent-specific variation is processed, while Low Experience listeners have not. One possible way in which to examine this in greater detail would be to hold accent strength constant and examine the effect of experience as a continuous predictor.

3.4.1 Conclusions

This study examined the effects of degree of foreign accentedness and listener experience on early auditory processing. Clear differences emerged across the accentedness continuum between levels of experience with Chinese-accented speakers. These results taken together point to a change in online auditory processing through long-term exposure to accent variation. The trajectory and locus of this change, however, remains and open issue.

References

Arnold, D., Wagner, P., & Baayen, R. H. (2013). Using generalized additive models and random forests to model German prosodic prominence. In *Proceedings of Interspeech 2013* (pp. 272–276). Lyon, France.

Atienza, M., Cantero, J. L., & Dominguez-Marin, E. (2002). The time course of neural changes underlying auditory perceptual learning. *Learning & Memory*, *9*(3), 138–150.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. H. (2010a). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon*, *5*(3), 436–561.

Baayen, R. H. (2010b). The directed compound graph of English. An exploration of lexical connectivity and its processing consequences. In S. Olson (Ed.), *New impulses in word-formation (Linguistische Berichte Sonderheft 17)* (pp. 383–402). Hamburg: Buske.

Baker, R. E., Baese-Berk, M., Bonnasse-Gahot, L., Kim, M., Van Engen, K. J., & Bradlow, A. R. (2011). Word durations in non-native English. *Journal of Phonetics*, *39*(1), 1–17.

Bernal, J., Harmony, T., Rodríguez, M., Reyes, A., Yáñez, G., Fernández, T., Galán, L., Silva, J., Fernández- Bouzas, A., Rodríguez, H., Guerrero, V., & Marosi, E. (2000). Auditory event-related potentials in poor readers. *International Journal of Psychophysiology*, *36*(1), 11–23.

Bien, H. & Zwitserlood, P. (2013). Processing nasals with and without consecutive context phonemes: Evidence from explicit categorization and the N100. *Frontiers in Psychology*, *4*(21), 1–12.

Bradlow, A. R. & Bent, T. (2003). Listener adaptation to foreign accented English. In M. Sole, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003)* (pp. 2881–2884). Barcelona, Spain.

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

Brunellière, A. & Soto-Faraco, S. (2014). The interplay between semantic and phonological constraints during spoken-word comprehension. *Psychophysiology*, *52*(1), 46–58.

Carretié, L., Mercado, F., Tapia, M., & Hinojosa, J. A. (2001). Emotion, attention and the 'negativity bias', studied through event-related potentials. *International Journal of Psychophysiology*, *41*(1), 75–85.

Clarke, C. M. & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, *116*(6), 3647–3658.

Connolly, J. F. & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6*(3), 256–266.

Connolly, J. F., Service, E., D'Arcy, R. C., Kujala, A., & Alho, K. (2001). Phonological aspects of word recognition as revealed by high-resolution spatio-temporal brain mapping. *Cognitive Neuroscience and Neuropsychology*, *12*, 237–243.

Crowley, K. E. & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology*, *115*(4), 732–744.

Curran, T. & Dien, J. (2003). Differentiating amodal familiarity from modality-specific memory processes: An ERP study. *Psychophysiology*, *40*(6), 979–988.

Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience*, *21*(10), 1893–1906.

Diaz, M. T. & Swaab, T. Y. (2007). Electrophysiological differentiation of phonological and semantic integration in word and sentence contexts. *Brain Research*, *1146*, 85–100.

Evans, K. M. & Federmeier, K. D. (2007). The memory that's right and the memory that's left: Event-related potentials reveal hemispheric asymmetries in the encoding and retention of verbal information. *Neuropsychologia*, *45*(8), 1777–1790.

Flege, J. E. (1980). Phonetic approximation in second language acquisition. *Language Learning*, *30*(1), 117–134.

Flege, J. E. (1984). The detection of French accent by American listeners. *Journal of the Acoustical Society of America*, *76*(3), 692–707.

Flege, J. E. & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America*, *76*(3), 708–721.

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.

Gass, S. & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning*, *34*(4), 65–87.

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*, *122*(2), 92–102.

Hagoort, P. & Brown, C. M. (2000). ERP effects of listening to speech: Semantic ERP effects. *Neuropsychologia*, *38*(11), 1518–1530.

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887.

Hastie, T. & Tibshirani, R. (1990). *Generalized additive models*. Monographs on Statistics & Applied Probability 43. Chapman and Hall/CRC.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*. New Series, *182*(4108), 177–180.

Kryuchkova, T., Tucker, B. V., Wurm, L. H., & Baayen, R. H. (2012). Danger and usefulness are detected early in auditory lexical processing: Evidence from electroencephalography. *Brain and Language*, *122*(2), 81–91.

Lee, J. Y., Harkrider, A. W., & Hedrick, M. S. (2012). Electrophysiological and behavioral measures of phonological processing of auditory nonsense V–CV–VCV stimuli. *Neuropsychologia*, *50*(5), 666–673.

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2007). Speaker-independent factors affecting the perception of foreign accent in a second language. *Journal of the Acoustical Society of America*, *121*(4), 2327–2338.

Martin, B. A., Sigal, A., Kurtzberg, D., & Stapells, D. R. (1997). The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. *Journal of the Acoustical Society of America*, *101*(3), 1585–1599.

Martin, B. A., Tremblay, K. L., & Stapells, D. R. (2007). Principles and applications of cortical auditory evoked potentials. In R. Burkard, J. Eggermont, & M. Don (Eds.), *Auditory evoked potentials: Basic principles and clinical application*. Point (Lippincott Williams & Wilkins). Lippincott Williams & Wilkins.

Misra, M. & Holcomb, P. J. (2003). Event-related potential indices of masked repetition priming. *Psychophysiology*, *40*(1), 115–130.

Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, *36*(1), 39–66.

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

Näätänen, R. & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375–425.

Newman, R. L. & Connolly, J. F. (2009). Electrophysiological markers of pre-lexical speech processing: Evidence for bottom–up and top–down effects on spoken word processing. *Biological Psychology*, *80*(1), 114–121.

Newman, R. L., Connolly, J. F., Service, E., & McIvor, K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event-related brain potentials. *Psychophysiology*, *40*(4), 640–647.

Porretta, V., Kyröläinen, A.-J., & Tucker, B. V. (2013). Influences on perceived foreign accentedness: Acoustic distances and lexical neighborhoods, June 23–28, 2013. International Cognitive Linguistics Conference (ICLC) 12. Edmonton, AB.

Porretta, V. & Tucker, B. V. (2012). Predicting accentedness: Acoustic measurements of Chinese-accented English. *Canadian Acoustics*, *40*(3), 34–35.

R Development Core Team. (2014). *R: A language and environment for statistical computing*. Version 3.1.0. R Foundation for Statistical Computing. Vienna, Austria.

Ross, B. & Tremblay, K. L. (2009). Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. *Hearing Research*, *248*(1–2), 48–59.

Sanders, L. D. & Neville, H. J. (2003). An ERP study of continuous speech processing II. Segmentation, semantics, and syntax in non-native speakers. *Cognitive Brain Research*, *15*(3), 214–227.

Shahin, A. J., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *The Journal of Neuroscience*, *23*(13), 5545–5552.

Shahin, A. J., Roberts, L. E., Pantev, C., Trainor, L. J., & Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport*, *16*(16), 1781–1785.

Sheehan, K. A., McArthur, G. M., & Bishop, D. V. (2005). Is discrimination training necessary to cause changes in the P2 auditory event-related brain potential to speech sounds? *Cognitive Brain Research*, *25*(2), 547–553.

Steinhauer, K. & Connolly, J. F. (2008). Event-related potentials in the study of language. In B. Stemmer & H. A. Whitaker (Eds.), *Handbook of the neuroscience of language* (pp. 91–104). New York: Elsevier.

Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., & Howard, M. A. (1999). Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *Journal of Neurophysiology*, *82*(5), 2346–2357.

Stelmack, R. M., Saxe, B. J., Noldy-cullum, N., Campbell, K. B., & Armitage, R. (1988). Recognition memory for words and event-related potentials: A comparison of normal and disabled readers. *Journal of Clinical and Experimental Neuropsychology*, *10*(2), 185–200.

Tong, Y., Melara, R. D., & Rao, A. (2009). P2 enhancement from auditory discrimination training is associated with improved reaction times. *Brain Research*, *1297*, 80–88.

Tremblay, A. (2011). *icaOcularCorrection: Independent Components Analysis (ICA) based eye-movement correction*. R package version 1.3.

Tremblay, A. & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151–173). London: The Continuum International Publishing Group.

Tremblay, A., Baayen, R. H., & Hendrix, P. (2013). *eRp: Pre-processing, generalized additive mixed-effects modeling, and visualization of event-related brain potential and field (ERP/ERF) data.* R package version 0.9.4.2.

Tremblay, K. L. & Kraus, N. (2002). Auditory training induces asymmetrical changes in cortical neural activity. *Journal of Speech, Language, and Hearing Research*, *45*(3), 564–572.

Tremblay, K. L., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training. *Ear and Hearing*, *22*(2), 79–90.

Tremblay, K. L., Shahin, A. J., Picton, T., & Ross, B. (2009). Auditory training alters the physiological detection of stimulus-specific cues in humans. *Clinical Neurophysiology*, *120*(1), 128–135.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, *53*(4), 510–540.

Vaughan Jr., H. G. & Ritter, W. (1970). The sources of auditory evoked responses recorded from the human scalp. *Electroencephalography and Clinical Neurophysiology*, *28*(4), 360–367.

Wayland, R. (1997). Non-native production of Thai: Acoustic measurements and accentedness ratings. *Applied Linguistics*, *18*(3), 345–373.

Wieling, M., Montemagni, S., Nerbonne, J., & Baayen, R. H. (2014). Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and sociodemographic variation using generalized additive mixed modeling. *Language*, *90*(3), 669–692.

Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLoS ONE*, *6*(9), 1–14.

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, *75*(3), 537–556.

Wolach, I. & Pratt, H. (2001). The mode of short-term memory encoding as indicated by event-related potentials in a memory scanning task with distractions. *Clinical Neurophysiology*, *112*(1), 186–197.

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Boca Raton: Chapman & Hall/CRC Press.

Wood, S. N. (2014). *mgcv: Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation*. R package version 1.7-29.

Zuur, A., Ieno, E., Walker, N., Saveliev, A., & Smith, G. (2009). *Mixed effects models and extensions in ecology with R*. New York: Springer.

Chapter 4

Gradience in word recognition

4.1 Introduction

Most people have had the experience of conversing with a non-native speaker and generally have the sense that some non-native speakers have heavier accents than others. Despite this, communication can be successful and it has even been shown that with increased exposure, native listeners find non-native speakers more intelligible (Bradlow & Bent, 2003; Bradlow & Bent, 2008; Gass & Varonis, 1984). However, it remains unclear exactly to what extent gradience in perceived foreign accent strength affects lexical processing. The present study employs native- and Mandarin-accented English words to examine the effect of gradient foreign accentedness on lexical activation strength and the time-course of word recognition.

Existing behavioral studies, mostly looking at adaptation to accented speech, suggest that foreign-accented speech requires more effortful processing in terms of both comprehensibility and intelligibility. Work on comprehensibility has shown that when listening to foreign-accented speech response latencies are longer in comparison to native speech (Clarke & Garrett, 2004; Floccia et al., 2009; Munro & Derwing, 1995). Munro and Derwing (1995) found that listeners took significantly longer to evaluate the truthfulness of sentences spoken in Mandarin-accented English. Additionally, Floccia et al. (2009) found that in a lexical decision task (in sentential context) the introduction of a foreign-accented talker led to a large increase in reaction time, which then attenuates, but does not return to the native talker baseline. They also determined that this difference in reaction time to foreign-accented speech does not improve (i.e., return to native baseline) with repeated exposure. Similarly, Adank and McQueen (2007) failed to find short-term adaptation to an unfamiliar regional accent even after 20 minutes of exposure. These results are in direct conflict with the results of Clarke and Garrett (2004) who found that adaptation, in relation to baseline reaction time, can occur in as little as 2 to 4 sentences.

Work examining intelligibility of foreign-accented speech, and subsequent adaptation to it, has also shown differences between native- and foreign-accented speech. Typically measured by transcription accuracy, intelligibility of non-native speech is initially poor, but improves with repeated exposure (Bradlow & Bent, 2003; Bradlow & Bent, 2008; Gass & Varonis, 1984). Furthermore, listeners benefit from exposure to multiple speakers of the same accent type, showing a generalized, speaker-independent adaptation for a particular accent (Bradlow & Bent, 2008; Sidaras et al., 2009).

One potentially problematic point is that these behavioral studies tend to dichotomize speech into accented and unaccented groups. This ultimately neglects the apparent gradience of foreign accent strength and fails to explore how processing might be affected along the continuum of foreign accentedness. One recent study has begun to fill this gap. Using cross-modal priming, Witteman et al. (2013) found that the speed of perceptual adaptation to German-accented Dutch is dependent on two factors, namely the strength of foreign accent and listener familiarity with the accent. Three levels of accent strength (high, mid, and low) were used and this strength was manipulated by including words with three Dutch vowels which native German speakers generally produce with varying degrees of accuracy. The strongly accented words contained a diphthong which does not exist in German (i.e., /œy/) and that German speakers typically produce with a perceptually and acoustically different vowel (i.e., /ɔɪ/). The medium accented words also contained a diphthong which does not exist in German (i.e., /ɛɪ/), but that German speakers produce with a vowel that is phonetically very similar (i.e., /aɪ/). The weakly accented words contain vowels that are shared between the two languages (e.g., /ɛ/ and /ɪ/). Participants were grouped for familiarity with the accent based on the university which they attended. Participants at a Dutch university near the German border were taken as the high familiarity group and participants at a central Dutch university who reported hearing German-accented Dutch less than once per week from fewer than two speakers were taken as the low familiarity group. Participants with low experience were facilitated by words spoken with a medium and weak accent, but not by words that were strongly accented. However, participants with high experience were primed by all levels of accent. The magnitude of the priming effect also depended on the type of exposure participants received before completing the task. While this study used relatively coarse-grained groupings for strength and familiarity, it seems that both degree of foreign accentedness and listener experience with a particular accent are involved in processing and merit further investigation.

The Visual World (Word) eye-tracking paradigm has been used to investigate the processing of specific pronunciation (segment-based) variants, one regional (Dahan, Drucker, & Scarborough, 2008) and one non-native (Hanulíková & Weber, 2012). This paradigm has been used extensively to investigate word recognition in native listeners (Allopenna, Magnuson, & Tanenhaus, 1998; Dahan et al., 2001; Huettig & McQueen, 2007; McQueen & Viebahn, 2007; Tanenhaus et al., 1995), by examining the proportion of looks to a particular target (either a picture or an orthographic form) shown on a computer screen after the onset of an auditory stimulus. Typically this method has been used to investigate effects of lexical competition, comparing the looks between the target and a phonologically related competitor. However, as this paradigm measures looks to target over time, it can also be used to investigate the effect of different variables on the time-course of word recognition (Hanulíková & Weber, 2012).

With regard to regional (or dialectal) variants, Dahan et al. (2008) investigated adaptation to a systematic dialectal variant which raises the vowel /æ/ before /g/, but not before /k/. For example, the word *sack* would contain the vowel /æ/, while the word *sag* would contain the raised variant. Native listeners' recognition of words ending in /k/ was facilitated by previous exposure to the dialectal variant, indicating that competition with words ending in /g/ is mitigated by the exposure. Listeners who obtained experience with the variation were thus sensitive to the reduced probability of the word ending in /g/ if the vowel was not raised.

They concluded that this demonstrates local (short-term) adaptation to the dialect-related accent and that this involves modification to representations of words (or their component parts) rather than simply normalizing the input signal. Because listeners' processing of words unaffected by the dialect was influenced by previous experience with the talker's dialectal productions, this indicates that this knowledge is made available when processing speech from a given talker (or, in a given context), independently of whether or not the signal itself contains the dialectal feature. These modifications to representations are likely to occur based on specific experiences listeners have with a particular context and/or speaker characteristics. However, the authors note that this raises questions regarding the generalizability of this experience to new words or talkers.

To the knowledge of the authors, one study has employed the visual world eye-tracking methodology to investigate the recognition of non-native pronunciation variants (segment-based accent) in L2 listeners (Hanulíková & Weber, 2012). In their study, Hanulíková and Weber (2012) presented English words with three *TH* (/θ/) pronunciation variants to native English, Dutch, and German speakers. The three variants were /f/, /t/, and /s/, each corresponding to typical realizations produced by English, Dutch and German speakers, respectively. The stimuli were produced by both a native Dutch speaker and a native German speaker, roughly matched for accent strength in English. Each produced target words with the three segmental variants. The hypothesis was that Dutch speakers would process the /t/ variant more easily, while German speakers would process the /s/ variant more easily, due to increased experience with those realizations. Thus, each group would more easily process the variant that is consistent with their own accent, despite the fact that /f/ is the most perceptually similar to /θ/. This was indeed found to be the case; each group more easily processed the variant with which they had the most experience. However, a control group of native English participants showed no main effect of substitution type, though the time-course indicated a slight preference for /f/. It may be the case that the native listeners had a sufficient amount of experience encountering these variants; however, this was not tested. Hanulíková and Weber (2012) suggest that the results of the Dutch and German participants are in line with a representation-based encoding of words and their variants in the L2 lexicon.

Two hypotheses have been put forth as a means of understanding online processing of foreign-accented speech. First, is the Perceptual Distance Hypothesis which relates the concept that accentedness is a degree of magnitude from what is considered native (Clarke & Garrett, 2004; Floccia et al., 2006; Goslin et al., 2012). The hypothesis states that processing of speech which is more perceptually similar to native speech (e.g., regional native or highly proficient non-native speech) will be an attenuated version of that for speech which is perceptually dissimilar (e.g., less proficient non-native speech). Second, the Different Processes Hypothesis (Goslin et al., 2012) states that speech which is consistent with the native language (e.g., home or regional accent) recruits a pre-lexical processing mechanism different from that which handles inconsistent speech (e.g., foreign accent).

Foreign-accented speech represents a form of variability that does not fully align with native distributions of the acoustic-phonetic dimensions of speech. There are two broad sets of accounts for how variability is handled during lexical processing, and possibly represented in the mind: representational accounts and processing-based accounts. Representational accounts maintain that variation is encoded in lexical representations and that perception can take place without speaker normalization or the conversion of phonetic instantiations to abstract categories before being mapped onto the lexicon. Within this set of accounts, episodic (exemplar-based) models assume that phonetic experiences are stored and become part of the representation (Goldinger, 1996; Goldinger, 1998; Johnson, 1997; Johnson, 2006; Pierrehumbert, 2001; Pierrehumbert, 2003a; Pisoni, 1997; Walsh et al., 2010). Other representational accounts claim that either multiple abstract representations for variant forms exist in addition to the canonical representation (Ranbom & Connine, 2007), or that phonetic variation is stored in memory with the lexical prototype (Klatt, 1979). These accounts support the idea that linguistic experience builds up over time. These representations can also include indexical information (e.g., gender, race, and dialect) about specific talkers or groups of talkers (Johnson, 2006).

Processing-based accounts, on the other hand, claim that the lexicon contains a single abstract, canonical representation for each word and that variability in the speech stream is handled by prelexical processing involving sublexical phonological abstraction (Gaskell & W. D. Marslen-Wilson, 1998; McQueen et al., 2006). Research supporting this claim suggests that variability is not part of the lexical representation because perceptual adaptation through experience can influence recognition of previously unheard words containing the deviation. This then points to a processing mechanism by which the signal is first mapped for individual sound segments before accessing the lexicon. Any effect of experience thus affects the process, not the representation.

Previous studies on variability have looked exclusively at variation of specific, individual phonemes, rather than variation along multiple dimensions in the entire word. The present work examines the effects of gradient foreign-accented speech on lexical activation strength and word recognition and contains two experiments, one a cross-modal identity priming task and the other a visual word eye-tracking task. These allow for an investigation of both the extent to which foreign-accented speech activates lexical representations and the time course of online access to those representations within the lexicon. Additionally, it is possible to examine the effect of listener experience with foreign-accented speech on lexical processing.

Here, we address a number of questions. Does the variability inherent to gradient foreign-accented speech modulate activation of lexical representations as evidenced through lexical priming? If so, is the effect also gradient? Does the variability of gradient foreign accentedness affect the time-course of online spoken word recognition? Again, if so, is the effect gradient? Lastly, does listener experience with Chinese-accented speakers affect patterns of lexical activation strength and spoken word recognition? And, what might this say about mental representations of variation and long-term, generalized adaptation to accent variability? We predict that increased foreign accentedness will reduce lexical activation through priming as well as slow the time-course of word recognition in a gradient fashion. We also predict that listener experience with Chinese-accented English will lead to facilitation in both.

4.2 Experiment 1

4.2.1 Methodology

The first experiment employs a cross-modal identity priming task. Cross-modal priming involves the presentation of an auditory prime word followed by a written target. After hearing the prime, participants must then make a lexical decision for the written target. Recently this paradigm has been used to investigate the effects of accented speech (Floccia et al., 2006; Floccia et al., 2009; Witteman et al., 2013).

In the present study, spoken English words with varying degrees of foreign accentedness will serve as the primes while the targets will be written English words or pseudowords. The auditory primes will consist of the same 40 monosyllabic words used in Chapters 2 and 3, though a reduced set of four speakers is used due to the constraints of this design. However, the speakers chosen for inclusion still represent the full range of foreign accentedness. Additionally, by including the same speakers and words from the previous two studies we allow for continuity across the studies and the added possibility (and benefit) of including the acoustic measures as predictors in the experiment. We created three types of prime-target combinations: 1) prime followed by the intended English word; 2) prime followed by a phonologically and semantically unrelated word; and 3) prime followed by a pseudoword. Given that the stimulus words were restricted in number and the same across all speakers it was necessary to create four counterbalanced lists. The counterbalancing ensured that no participant heard all 40 words from any speaker in the Prime-Identity group, the critical items in the experiment.

Examining participants' reaction times to targets in the Identity group it is possible to observe any differences in the strength with which words spoken with varying degrees of accentedness prime their intended targets. It was expected that words with stronger accents will serve as poorer (or weaker) primes. It was also expected that participant experience with Chinese-accented English will modulate the priming effect such that more experienced listeners will be equally facilitated by strongly accented words as by weakly accented words.

Participants

Forty-eight native speakers of North American English (33 female, 15 male) were recruited from the University of Alberta campus area and ranged in age from 18 to 29 years old (M = 20.38, SD = 2.79). All had normal or corrected-to-normal vision and reported having normal hearing.

Stimuli

Auditory primes. Word list recordings of 3 male native Chinese speakers and 1 male native English speaker were retrieved from the NU Wildcat Corpus of native- and foreign-accented English (Van Engen et al., 2010). Each recording contained a word list read three times by a single talker and a subset of 40 monosyllabic words from the first repetition was used for this study. Individual sound files for each word by talker were extracted and then normalized for amplitude. In a study of perceived foreign accentedness (see Chapter 2), ratings for foreign accentedness were collected for these tokens using a scale of 1 to 9 (1 corresponding to no foreign accent and 9 corresponding to a very strong foreign accent). It was found that the tokens produced by these talkers covered the range of the accentedness and talker means were spread approximately evenly across the scale.

Written targets. Each auditory prime word was paired with three types of written targets. The first type was the standard orthographic form of the prime word (Identity). The second type was a semantically and phonologically unrelated English word (Unrelated). For example, *sash* is unrelated to *beet*. These unrelated words were matched to the prime for number of phonemes (extracted from the English Lexicon Project, Balota et al. (2007)) and lexical frequency (extracted from the Corpus of Contemporary America English, COCA, Davies (2008)). A Welch Two Sample t-test revealed that the frequencies of the prime and unrelated words were not significantly different ($t(77.83) = -0.04$, $p = 0.97$). Every attempt was made to avoid the presence of overlap of phonemic segments between the prime and the unrelated target. However, for four primes (each containing five phonemes), the corresponding unrelated words overlapped by one phonemic segment, though the overlapping phonemes were in different positions between the two words. The third type was a phonologically licit pseudoword. Two lists of pseudowords were created for each prime to ensure there was no repetition due to the counterbalancing. Pseudowords were taken from the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002). Again, pseudowords were matched to primes for number of phonemes and again every attempt was made to avoid the presence of overlap of phonemic segments between the prime and the pseudoword target. As with the unrelated words, four pseudowords overlapped with their prime by one phonemic segment. The written targets can be found in Table A.3 located in the Appendix.

Counterbalancing Given that each talker produced the same 40 auditory prime words, four counterbalanced lists were created. This ensured that any participant would only hear 10 words in the Prime-Identity pairing from each of the four talkers. Thus for a given talker, a participant would hear 10 Prime-Identity pairings, 10 Prime-Unrelated pairings, and 20 Prime-Pseudoword pairings. Stimuli were blocked by talker, beginning with the native English talker. The presentation of the non-native talkers was arranged so that accentedness increased based on their mean accentedness rating. This was done so that the prime words would generally become more difficult to understand as the experiment progressed and counteract any effect of repetition of tokens produced by the previous talker. Within each block stimuli were randomized.

Procedures

Participants completed the task seated at a computer in a sound-attenuated booth. Stimuli were presented binaurally in E-Prime 2.0 (Psychology Software Tools, Inc.) using over-the-ear headphones (MB Quart QP 805) adjusted to a comfortable level. Participants were instructed to listen to the auditory stimulus and decide if the following written target word was an existing English word or not. Responses were made on a Serial Response button box (200A, Psychology Software Tools, Inc.). Visual targets were presented in the center of a 19-inch LCD computer monitor (1440×900 resolution) in black lowercase 36-point Tahoma font and appeared on-screen 500 ms after the acoustic offset of the auditory prime, following Witteman et al. (2013). If no response was made, targets remained on-screen for 2000 ms, before moving on to the next trial. Four practice items (from the Native English talker) were provided prior to the experimental items so that participants could familiarize themselves with the task. Between blocks, these practice items were repeated as fillers. The purpose of these fillers was to serve as an "accent reset" to minimize the effect of the previous accented talker on the subsequent one. Additionally, short, self-paced breaks were allowed between blocks.

After the priming experiment was completed, participants responded to a series of questions (see Table A.5 in the Appendix) which elicited demographic information as well as details of language experience. Language experience information included experience interacting with non-native English speakers, specifically Chinese-accented speakers. In total, the experiment lasted approximately 30 minutes.

4.2.2 Analysis and results

Reaction times were modeled using generalized additive mixed modeling (GAMM) (Hastie & Tibshirani, 1990; Wood, 2006). GAMM does not assume a linear relationship between predictors as does ANOVA or linear regression models and is capable of handling non-linearities in the data. This point is particularly important for the present study due to the working conception of gradient foreign accentedness. GAMM has been previously applied successfully to model a variety of linguistic data. In particular, it has been used for investigating dialectal variation (Wieling et al., 2011; Wieling et al., 2014), event-related potentials (Kryuchkova et al., 2012; A. Tremblay & Baayen, 2010), prosodic prominence (Arnold et al., 2013), and reaction times (Baayen, 2010a; Baayen, 2010b). Outside of language, it has been used extensively in the field of ecology (Zuur et al., 2009).

Here we present two models. The first model is used to examine the expected facilitation effect of items of type Prime-Identity in comparison to items of either the Prime-Unrelated or Prime-Pseudo types. The second model then specifically examines the role of degree of foreign accentedness on reaction times to those Prime-Identity items. The analysis was performed in the statistical environment R, version 3.1.0 (R Development Core Team, 2014) using the package mgcv (Wood, 2014), version 1.7-29. Only items to which participants responded correctly were included in the models. Missing values (n = 15) and incorrect responses (n = 489) were removed, resulting in the loss of 6.56% of the total data (N = 7680). However, before proceeding to the models, we first examine the overall accuracy of the participant responses along with descriptive statistics of the reaction times using the full data set.

Accuracy

The accuracy of the participant responses was examined. Missing values (i.e., items to which participants did not respond) were taken as incorrect responses. Accuracy by subject ranged from 75% to 100% correct (M = 93.44, SD = 5.87). Given that no participant scored below 75%, all participants were included in the analysis. Accuracy by condition, was as follows: Identity M = 98.18%, SD = 7.37; Pseudoword M = 92.76%, SD = 2.67; and Unrelated M = 90.05%, SD = 11.95.

Summary of reaction times

Mean reaction times were calculated for each target type (i.e., Identity, Unrelated, and Pseudoword) across all participants. For targets matching the identity of the prime the mean reaction time was 497.55 ms (SD = 177.79). For targets which were existing words, but unrelated to the prime, the mean reaction time was 575.04 ms (SD = 182.02). And, for targets which were pseudowords the mean reaction time was 622.52 ms (SD = 229.51). Thus, the expected effect of target type is already visible in the means; however, we further confirm this by fitting a model to the reaction times.

RT model - The effect of target type

Variables The input variables to the model were as follows. The primary variable of interest was Target Type with three levels (Identity, Unrelated, and Pseudoword), Identity taken as the baseline. In addition to this, we also included List (i.e., counterbalance with four levels) and Trial (i.e., presentation order of stimuli), along with random effects for Subject and Item. These variables help to control for a number of things: 1) List controls for which items were presented to which participants; 2) Trial controls for the effects of learning and fatigue throughout the course of the experiment; 3) Subject controls for variable performance in the sample; and 4) Item controls for variability within the sample. The response variable of the model was Reaction Time, which was log-transformed prior to analysis (Baayen, 2008).

Model fitting and evaluation The input variables described above were fit to the response variable (Log Reaction Time) with by-Subject and by-Item random intercepts. For Trial, a nonlinear functional relation with the response variable was allowed for by using a thin plate regression spline with restricted maximum likelihood to optimize the number of smoothing parameters. Target Type, a factor with three levels, and List, a factor with four levels, were included as parametric components.

The model was fit using the backwards step-wise elimination procedure set out in Zuur et al. (2009). All predictors were first included as main effects with method set to ML (Maximum Likelihood). Their contribution to the model was evaluated using two criteria. The first criterion was the estimated *p*-value of the smoothing parameter or parametric component. The estimated *p*-value of the smoothing parameter or parametric component indicates whether or not the functional form of the predictor is different from zero. If greater than the conventional alpha level of 0.05, the predictor was considered for removal. The second criterion was the Akaike information criterion (AIC) value (Akaike, 1998). The use of AIC is an information-theoretic approach which supplies information on the strength of evidence for a particular model given the data when a particular smoothing parameter is removed. Lower AIC values indicate increased evidence for that model. This approach is particularly advantageous as it is not affected by the order of variables entered into the model.

Upon removal of the predictor, the AIC value of that model was compared to that of the model containing the predictor in question. If the difference in AIC was less than 2, the predictor was removed and we continued with the model fitting procedure (Burnham & Anderson, 2002). Through this process, only List was eliminated from the model. Thus, the model consists of the following predictors: Random intercepts for Subject and Item, main effect for Trial, and main effect of Target Type.

This model was trimmed, refitting for data within -/+2.5 standard deviations of the residuals of the model (1.9% removed, 134 datapoints) (Baayen, 2008). Re-inspection indicated that the residuals were approximately normally distributed and this was taken as the final model.

To evaluate the model we calculated ΔAIC values for main effects in the model (see Table 4.1). This allows us to rank the predictors in terms of strength of evidence. This is done by subtracting the AIC of the model without the predictor from the AIC of the model including the predictor (i.e., the full model). As a rule of thumb, a $\Delta < 2$ suggests substantial evidence for the model not containing the predictor in question. Δ values between 3 and 7 indicate considerably less support for the removal, while $\Delta > 10$ indicates that the new model is very unlikely (Burnham & Anderson, 2002). Looking at the ΔAIC values in Table 4.1, it can be seen that the smooth function for Trial is playing an important role in explaining the data. Additionally, the variable of interest here (Target Type), has a sizable impact on the likelihood of the model.

Table 4.1: Model AIC and ΔAIC for Target Type RT model.

Model	df	AIC	ΔAIC
Full model	483.76	−2673.1	NA
w/o Target Type	580.22	−2594.02	79.08
w/o Trial	495.89	−2540.66	132.44

After evaluation the full model was refit using method set to REML (Restricted Maximum Likelihood) (Zuur et al., 2009) and trimmed as above. This final model, reported below, accounts for 56.2% of deviance explained showing that the model is able to capture important facets of variation in reaction times.

Results The results of the final model are reported below along with visualizations of the effects. The table of coefficients is found in Table 4.2.

As expected, when the auditory prime preceded its identity, reaction times to the target were significantly faster than the other target types (Figure 4.1), with unrelated targets producing less perturbation in reaction time than pseudowords. This confirms that participants were performing as one might expect in a cross-modal priming task. With this confirmation, we examine specifically the effect of foreign accentedness rating on priming (as measured by reaction time). Additionally, a significant, non-linear effect of Trial was present. This effect

Table 4.2: Generalized additive mixed model reporting parametric coefficient (Part A) and estimated degrees of freedom (edf), reference degrees of freedom (Ref.df), F and p values for the tensor products and random effects (Part B) for the Target Type RT model.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
intercept	6.1512	0.0241	255.7597	< 0.0001
Type (Pseudo)	0.2353	0.0100	23.5992	< 0.0001
Type (Unrelated)	0.1674	0.0116	14.4500	< 0.0001
B. smooth terms	edf	Ref.df	F-value	p-value
smooth for Trial	6.0705	7.1496	25.1034	< 0.0001
smooth for Subjects	46.4463	47.0000	111.8032	< 0.0001
smooth for Item	428.0250	637.0000	2.2307	< 0.0001

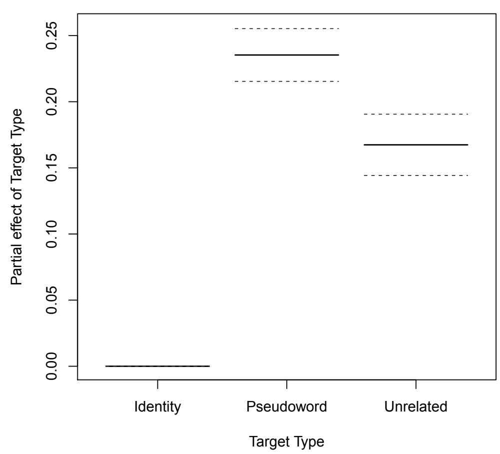

Figure 4.1: Partial effect of Target Type (Identity as baseline) with 95% confidence intervals.

indicates that overall participants' reaction times decreased as they progressed through the experiment.

RT model - The effect of accentedness and experience

Subsetting the dataset to include only the Prime-Identity trials allows us to examine how the effect of the identity prime demonstrated in Model 1 is influenced by both Accentedness Rating and Listener Experience with Chinese-accented speakers.

Input variables The primary variables of interest were Accentedness Rating of the prime and Listener Experience with Chinese-accented speakers. However, in addition to these, List (i.e., counterbalance with four levels), Trial (i.e., presentation order of stimuli), Log frequency of the target word (from COCA, Davies (2008)) were included, along with random effects for Subject and Item. These additional variables help to control for a number of things: 1) List controls for which items were presented to which participants; 2) Trial controls for the effects of learning and fatigue throughout the course of the experiment; 3) Frequency controls for the known effect that higher frequency words are recognized faster; 4) Subject controls for variable performance in the sample; and 5) Item controls for variability within the sample.

Prior to the present experiment, accentedness ratings were collected from 30 native English-speaking participants on the 40 words spoken by each of the four talkers in the stimulus set (total of 160 items). This ratings task was performed separately from the present study by different listeners. The ratings are based on a scale from 1 to 9 in which 1 represents no foreign accent (i.e., completely native production) and 9 represents a strong foreign accent (i.e., highly non-native production). These ratings serve as a measure of accentedness associated not only with each talker but with each experimental item allowing for a more fine-grained investigation of the effect of accentedness. For more information about the task and ratings, please refer to Chapter 2.

Responses to two items on the questionnaire were used to obtain a measure of listener experience interacting with Chinese-accented speakers. The first question was: "On a weekly basis, how often do you interact with non-native speakers of English?" Participants responded on a scale of 0 (Never) to 10 (Daily). The second question was: "What percentage of those interactions include speakers with a Chinese accent?" Participants provided a percentage.

To calculate Experience, the response to the first question was converted to a proportion by dividing by ten. This was then multiplied by the response to the second, also converted to a proportion by dividing by 100. The final value was then multiplied by 100 to obtain a value ranging between zero and 100 (see left panel of Figure A.2 in the Appendix). This variable was split into quantiles and then grouped (quantile 1 and quantiles 2–4), producing a factor (Experience) with two levels which we refer to as Low and High. This allowed for an empirical definition of Experience level based on the sample rather than an predefined criterion for determining High versus Low.

Response variable The response variable of the model was Reaction Time, which was log-transformed prior to analysis (Baayen, 2008).

Model fitting and evaluation The input variables described above were fit to the response variable (Log Reaction Time) with by-Subject and by-Item random intercepts. By-Subject random intercepts allow for the possibility that some participants may be faster than others to respond to the target. By-Item random intercepts allow for the possibility that some items may be more difficult than others. For Trial, Log Frequency and Rating, nonlinear functional relations with the response variable were allowed for by using a thin plate regression spline with restricted maximum likelihood to optimize the number of smoothing parameters. For the smoothing parameter of Rating, an interaction was included for the factor variable of Experience. List, a factor with four levels, was included as a parametric component.

Here, the model fitting procedure is the same as in the previous model. During the fitting process, only List was removed from the model. Thus, the model consists of the following predictors: Random intercepts for Subject and Item, main effect for Trial, main effect of Log Frequency of the target, and an interaction between Rating and Experience. The model was refitted for data within -/+2.5 standard deviations of the residuals of the previous model (2.3% removed, 44 datapoints) (Baayen, 2008). Re-inspection indicated that the residuals were then approximately normally distributed.

Additionally, the model evaluation process mimicked that of Model 1, examining AIC and ΔAIC values of the model variants. As seen in Table 4.3, the smooth function for

Trial is playing an important role in explaining the data. The variables of interest are also contributing significantly to the likelihood of the model. Lastly, Log Target Frequency is impacting model likelihood.

Table 4.3: Model AIC and ΔAIC for Accentedness RT model.

Model	df	AIC	ΔAIC
Full model	122.54	−697.55	NA
w/o Log Target Frequency	131.56	−690.25	7.3
w/o Experience	119.46	−687.27	10.28
w/o Rating	138.08	−670.77	26.77
w/o Trial	129.58	−655.37	42.17

After evaluation the full model was refit using REML (Zuur et al., 2009) and trimmed as above. The final model reported below accounts for 50.6% of deviance explained showing that the model is able to capture important facets of variation in reaction times.

Results The results of the final model are presented and visualized below. The coefficients of the model are presented in Table 4.4.

Table 4.4: Generalized additive mixed model reporting parametric coefficient (Part A) and estimated degrees of freedom (edf), reference degrees of freedom (Ref.df), F and p values for the tensor products and random effects (Part B) for the Accentedness RT model.

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
intercept	6.1157	0.0274	223.5430	< 0.0001
Experience (Low)	0.1260	0.0502	2.5085	0.0122
B. smooth terms	edf	Ref.df	F-value	p-value
smooth for Subjects	44.1152	46.0000	24.3016	< 0.0001
smooth for Item	59.0837	158.0000	0.6297	< 0.0001
smooth for Rating by Experience (High)	6.3621	8.0000	17.0007	< 0.0001
smooth for Rating by Experience (Low)	3.0322	8.0000	2.1541	0.0021
smooth for Trial	4.5898	5.5516	10.3132	< 0.0001
smooth for Frequency	3.9777	4.5259	4.6111	0.0007

As with the Target Type model, Trial resulted as significant, non-linear effect indicating that participants' reaction times decreased as they progressed through the experiment. Additionally, a significant, non-linear effect of Log Target Frequency was present. This is quite common in lexical decision experiments and shows that reaction times are faster for high frequency words.

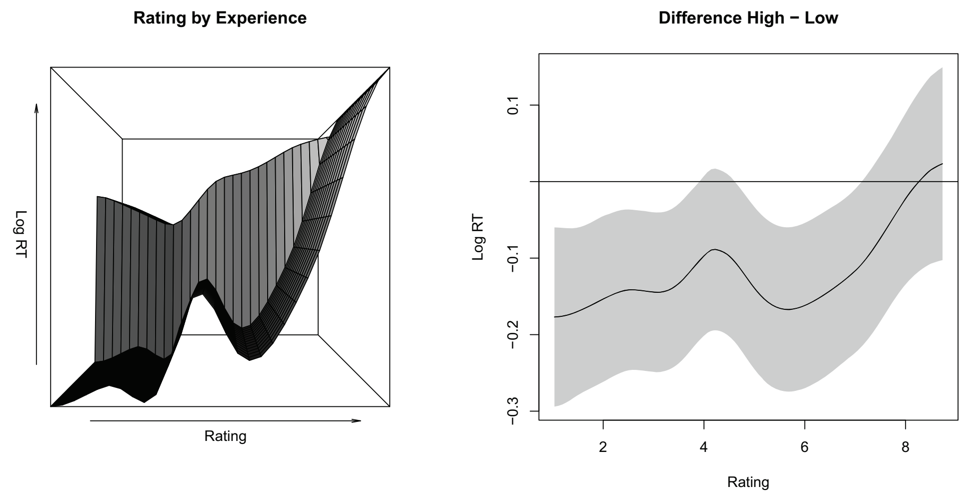

Figure 4.2: Rating and Experience. Left panel: Perspective plot of the interaction between Rating and Experience in which the rear curve represents the Low Experience group and the front curve represents the High Experience group. Right Panel: Plot of the difference between High and Low Experience groups with 95% confidence bands.

As can be seen in the left panel of Figure 4.2 (predicted values), a clear non-linear effect emerges for the High Experience group (front curve) and the Low Experience group (rear curve). Reaction times to the target increase as Accent Rating increases, though this increase is manifested differently between the two experience groups. The High Experience group has generally faster reaction times overall, which spike at the far end of the accentedness continuum. The Low Experience on the other hand have slower reaction times in comparison and display a gradual increase across the continuum.

Interestingly, ratings at the low end of accentedness continuum show relatively flat reaction times in both groups. Reaction times then begin to increase at approximately the same point along the continuum for both groups. However, this increase does not persist for the High Experience group. The right panel of Figure 4.2 presents a difference curve (High experience minus Low experience) for the predicted values along the accentedness continuum. The solid horizontal line represents zero and shading represents 95% confidence bands. As can be seen, there is a significant difference in performance between the two groups along the continuum of accentedness. The High Experience group has faster reaction times overall, with performance in the two groups converging for very strongly accented items.

4.2.3 Discussion

This cross-modal identity priming experiment aimed to illuminate the roles of gradient foreign accentedness and listener experience on lexical activation strength. The first model confirmed that targets in the identity condition received faster responses than either unrelated real word targets or pseudoword targets, an effect seen in previous studies (Clarke & Garrett, 2004; McQueen et al., 2006; Witteman et al., 2013). The second model specifically examined the effect foreign accentedness rating and listener experience on the identity condition. It revealed that the modulation of reaction times is in the direction one might expect for both foreign accentedness rating and listener experience.

This type of priming has been examined using the same paradigm as a means of assessing local adaptation to discretized accented speech (Clarke & Garrett, 2004; Witteman et al., 2013), showing that adaptation to foreign-accented speech can occur and that this adaptation may be affected by listener experience. However, previous studies have not specifically examined the strength of activation using a continuous measure of accentedness. The results of the present model show that the facilitatory effect of this identity condition is in fact modulated by gradient foreign accentedness and listener experience.

This modulation of priming within the identity condition points to variability in the strength of activation of lexical representations. Less accented words prime their identity better than more heavily accented words. This suggests that foreign accent may affect how well the auditory token maps onto the lexical representation. The more deviant a token is from a typical native-like production, the less it activates the representation, ultimately leading to an increase in processing time. Interestingly, however, the functional form of this modulation appears to be affected by the amount of experience a given participant has interacting with Chinese-accented speakers.

Experience plays a significant role in the model, and group differences indicate that gradient foreign accentedness produces different priming effects in the two groups. Low Experience listeners show a gradual increase in reaction times across the continuum while the reaction times of the High Experience listeners spike at the far end of the continuum. This sharp spike is likely due to tokens which are either unintelligible, or which have actually mapped onto a different lexical item than the intended. Noteworthy is that both groups begin an increase between ratings three and four, roughly corresponding with the far end of the rating values assigned to a native English talker. This increase might be characterized as a type of “native range" beyond which performance begins to change. However, the reaction times of the High Experience group decrease again in the mid-range. This may be a compensatory strategy used once it is clear that they are listening to non-native speech and have experience handling this particular type of variability. The Low Experience group does not show this same recovery, instead continuing their upward trajectory. This seems to suggest that listeners in the High Experience group not only map productions differently in order to activate lexical representations, but also appear to use some indexical information about Chinese-accented speakers as a larger group of speakers in order to compensate for the variability. Of additional interest is that differences emerge between the groups even in the native range of the accentedness continuum. This was not anticipated and will be taken up in greater detail in the general discussion.

This experiment gives a clear indication that lexical processing is modulated along the continuum of foreign accentedness with heavier accent resulting in weaker facilitation. It also shows that listener experience plays a role in this modulation. However, it only addressed the extent to which an auditory token will prime the intended lexical representation. In order to understand the time-course of online word recognition of the auditory tokens themselves it is necessary to utilize a method which tracks recognition as the stimulus unfolds. This topic is addressed in more detail in the following eye-tracking study.

4.3 Experiment 2

4.3.1 Methodology

The second experiment is an eye-tracking study using the Visual Word Paradigm. This allows for the determination of location of eye gaze on the screen as the acoustic signal is heard. In this paradigm, fixations reflect the processes involved in the activation of specific referents in visual context (McQueen & Viebahn, 2007; Tanenhaus et al., 1995). Here we employ a listen-and-look design in which participants hear an auditory token (of varying degree of foreign accentedness) and then, using their eyes, must find the written form of the word among four visually-presented options. Thus, no overt, decision-based response is required. This type of design has previously been used in a similar study investigating accented speech (Hanulíková & Weber, 2012). In the current study, the main manipulation is the degree of accentedness of each auditory item. Following Allopenna et al. (1998), the four written words presented on screen consist of the following: 1) the target word, 2) a phonologically similar word sharing the onset, 3) a phonologically similar word varying in the onset, and 4) a phonologically and semantically unrelated distractor.

Experiment 1 showed how degree of foreign accentedness influences lexical processing and the role of listener experience. This experiment expands on this by examining participants' eye gaze to target words on screen while listening to a token of varying accentedness. This makes it possible to investigate whether and how the time-course of online word recognition is affected by gradient foreign accentedness. It is expected that words with stronger accents activate lexical representations more poorly, resulting in increased latency or failure to fixate on the target. It is also expected that participant experience with Chinese-accented English will modulate looks to the target such that increased experience will allow participants to activate lexical representations better across accent strengths.

Participants

Forty-eight native speakers of North American English (36 female, 12 male) were recruited from the University of Alberta campus area and ranged in age from 18 to 27 years old (M = 19.98, SD = 1.99). All had normal or corrected-to-normal vision and reported having normal hearing. None of these participants took part in Experiment 1.

Stimuli

Auditory stimuli. The auditory stimuli were the same as those used in Experiment 1.

Written quadruplets. Quadruplets were created for each auditory stimulus, consisting of the target identical to the auditory stimulus, two competitors, and one distractor. The two competitors were modeled after Allopenna et al. (1998). The first was a rhyme with the target word, while the second shared its onset (simple or cluster) with the target and contained a vowel neighboring that of the target. For example, the target *bet* would be paired with *net* (rhyme) and *ban* (onset). The inclusion of an onset competitor containing the same vowel as the target (as in Allopenna et al. (1998)) was not possible due to the use of an existing set of auditory stimuli. For some, the only existing competitor with the same vowel was already present in the list of auditory stimuli. The distractor was semantically and phonologically unrelated to the target. For example, *zone* is unrelated to *bet*. All words in each quadruplet were matched for number of phonemes and an attempt was made to match for lexical frequency (from COCA). Welch Two Sample t-tests reveal that the frequencies of the target words were not significantly different than those of the rhyme competitors ($t(40.25) = 1.17$, $p = 0.25$), onset competitors ($t(39.25) = 1.28$, $p = 0.21$), or distractors ($t(57.69) = 0.56$, $p = 0.58$). The quadruplets can be found in Table A.4 located in the Appendix.

Counterbalancing As with the previous experiment, each talker produced the same 40 words and therefore it was necessary to create four counterbalanced lists. This ensured that any participant would hear all 40 words, 10 from each talker. Stimuli were blocked by talker, beginning with the native English talker. The presentation of the non-native talkers was arranged in the same way as in Experiment 1. Within each of the four blocks stimuli were randomized.

Procedures

Participants completed the task individually in a sound-attenuated booth in front of a 23-inch LCD computer monitor (1920×1080 resolution) connected to a desktop-mounted Eyelink 1000 eyetracker (SR Research Ltd.). A head-stabilizing chin-rest was fixed to the table 86 cm from the top of the monitor and 58 cm from the top of the eyetracker. The system was calibrated to the participants' right eye. Eye movements were sampled at a rate of 1000 Hz and time-locked to the onset of the auditory stimulus. The auditory and visual stimuli were presented using SR Research Experiment Builder (version 1.10.165). The visual stimuli for each trial consisted of the written quadruplets presented on-screen. The written words were displayed in black lowercase 72-point Tahoma font on a white background. Each word was centered on one of four positions: 610×190 pixels, 610×890 pixels, 1310×190 pixels, and 1310×890 pixels. The positions of the target, competitors, and distractor were balanced across trials.

Prior to beginning the experiment, a 9-point calibration procedure was completed in order to prepare the eye-tracker. Written instructions were provided explaining that the participant would see four printed words on the screen, followed by an auditory word. Participants were instructed to listen to the word, find it's written form among the four words presented on-screen, and look at it. Once the participant's gaze fell with the target interest area for 3 seconds, the trial would then end. Therefore, no explicit response was necessary, apart from locating and looking at the word (Huettig & Altmann, 2005). This listen-and-look task was employed in order to avoid any effects related to additional decision making and motor movements. At the start of each trial, a fixation cross appeared on-screen for 1000 ms followed by the written quadruplet. The auditory stimulus was presented binaurally through ER-1 headphones starting 200 ms after the onset of the visually displayed quadruplet (McQueen & Viebahn, 2007).

A short block of 4 practice items (from the Native English talker) was provided at the beginning to familiarize the participant with the task ahead. Between blocks, these practice items were repeated as fillers. The purpose of these fillers was to serve as an "accent reset" to minimize the effect of the previous accented talker on the subsequent. Additionally, short, self-paced breaks were allowed between blocks. After the experiment was completed, participants responded to the same questionnaire (see Table A.5 in the Appendix) used in Experiment 1. In total, the experiment lasted approximately 30 minutes.

4.3.2 Analysis and results

The data containing each millisecond sample were exported using SR Research Data Viewer (version 1.11.1), relative to the onset of the auditory stimulus (-100 ms to 1200 ms). These samples were then converted to proportion of samples falling within and outside each of the predefined interest areas in 20 ms windows (20 data points). Blinks and saccades were taken as missing data.

Figure 4.3 shows the grand average of these proportions by word type (i.e., interest area) within the visually displayed quadruplet. As expected, looks to the target word grow over time, beginning to diverge from both the onset competitor and the rhyme competitor at approximately 300 ms.

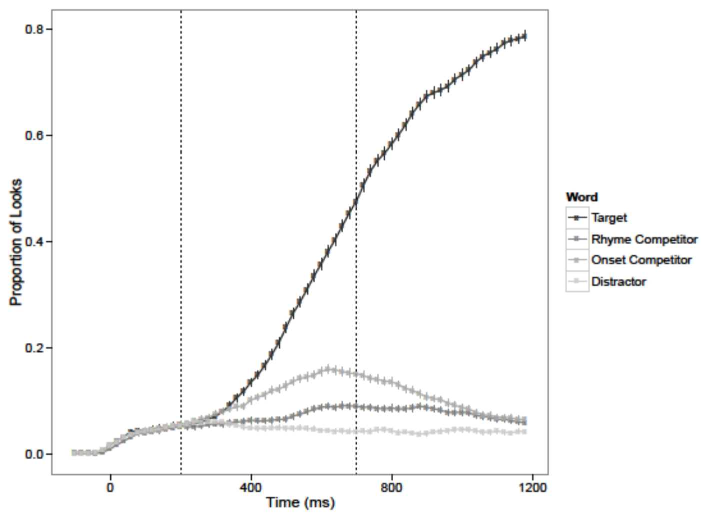

Figure 4.3: Grand average of proportion of looks to the four word types from -100 ms to 1200 ms post-stimulus-onset with standard error bars. Vertical lines at 200 ms and 700 ms indicate the subsequent analysis window.

Programming a saccade takes roughly 200 ms (Fischer, 1992; Matin, Shao, & Boff, 1993), thus it is approximately this point at which fixations will begin to be driven by the acoustic information present in the stimulus. Therefore we take 200–700 ms post stimulus onset as our window of interest for the subsequent analysis. While Figure 4.3 seems to indicate that participants fixated the target most during this window, Welch Paired Two Sample t-tests were performed on average proportion of looks over the window for each combination of word type. All resulted as significantly different with p-values below 0.001 and t-values above 4.20. This indicates that within this window, participants were most likely to fixate on the target word. With this verified, we now examine how degree of foreign accentedness affects looks to the target word over time.

Modeling eye movements

Similar to the analysis of reaction times, looks to the target were modeled using generalized additive mixed modeling. This type of eye-tracking data is non-linear by nature (see Figure 4.3) underlining the importance of the ability to model non-linear relationships. Also, given the time-series nature of visual world data, GAMM allows for the control of autocorrelation in the data. Autocorrelation deals with the correlation between datapoints in a time-series; a measurement at timepoint *t* is correlated to differing degrees with a measurement at timepoint *t-i*, depending on the lag.

Input variables As in Experiment 1, the primary variables of interest were Accentedness Rating of the token and Listener Experience with Chinese-accented speakers. These ratings were the same as those described in Section 4.2.2. Additionally, Experience was again calculated as laid out in the same section, producing a factor with two levels (i.e., High and Low). See the right panel of Figure A.2 in the Appendix for a plot of the raw values.

In addition to these, Time (ms) was included as a co-variate. Log COCA frequency of the target word (Davies, 2008), and Item were also included control for: 1) the known effect that higher frequency words are recognized faster; and 2) variability within the sample of items.

Response variable The response variable of the model was logit-transformed proportion of looks to the target (Barr, 2008). This transformation provides an unbounded measure symmetric around zero, which represents 50%. Additionally, by-item averaging was done within each Experience group, providing a time series, per item, per group. This helped to reduce noise in the signal as well as improve model fit.

Model fitting and evaluation The input variables described above were fit to the response variable (Logit Looks to Target) with by-Item random intercepts. For both Log Frequency and Rating, nonlinear functional relations with the response variable over Time were allowed for by using a thin plate regression spline with restricted maximum likelihood to optimize

An unaveraged model was fitted to the data with similar effects. However, even after trimming the model, the distribution of the residuals still did not approximate a normal distribution.

the number of smoothing parameters. For the smoothing parameter of Rating and Time, an interaction was included for the factor variable of Experience, using a tensor product (Baayen, 2010b; Wood, 2006). Additionally, an AR-1 correlation parameter, $\rho = 0.87$, was estimated from the data and included to control for autocorrelation in the timeseries data.

The model was fit using a backwards step-wise elimination procedure, similar to that used in Experiment 1. The inclusion of predictors in the model was evaluated using two criteria. The first criterion was the estimated p-value of the smoothing parameter or parametric component. The estimated p-value of the smoothing parameter or parametric component indicates whether or not the functional form of the predictor is different from zero. If greater than the conventional alpha level of 0.05, the predictor was considered for removal. The second criterion was Maximum Likelihood (ML) score comparison (Zuur et al., 2009). Here ML comparison is preferred over AIC as AIC becomes less reliable (i.e., anti-conservative) when autocorrelation is taken into account.

Upon removal of the predictor, the ML score of that model was compared to that of the model containing the predictor in question (i.e., the full model), indicating whether or not the inclusion of the predictor significantly improves model likelihood. Through this process, the interaction between Log Frequency and Time was eliminated from the model. Thus the model consists of the following predictors: Random intercepts for Item and an interaction between Time and Rating by Experience. The model was refitted for data within -/+2.5 standard deviations of the residuals of the model (1.5% removed, 122 datapoints) (Baayen, 2008). Re-inspection indicated that the residuals were approximately normally distributed.

Additionally, ML score comparison was done in order to evaluate the terms included in the full model (see Table 4.5). Here, the Chi-square statistic indicates a significant difference in likelihood between the models, given the degrees of freedom, and p-values below the conventional alpha level of 0.05 indicate evidence for the more complex model (i.e., the full model). In all cases, the full model including Item, Time, Rating, and Experience is the most likely.

Again, after evaluation the full model was refit with method set to REML and trimmed as above. This final model, reported below, accounts for 53.9% of deviance explained showing that the model is able to capture important facets of variation in fixations over time.

Table 4.5: Maximum Likelihood comparison between full model and model variant (simpler model) for eye-tracking model.

Model	Score	Edf	Chisq	Difference	*p*-value
Full model	4891.17	13	NA	NA	NA
w/o Experience	4920.55	8	29.37	5	< 0.0001
w/o Rating	4923.08	7	31.9	6	< 0.0001

Results The results of the final model are reported below along with visualizations of the effects (see Figures 4.4 and 4.5). The coefficients of the model are found in Table 4.6.

Table 4.6: Generalized additive mixed model reporting parametric coefficient (Part A) and estimated degrees of freedom (edf), reference degrees of freedom (Ref.df), F and p values for the tensor products and random effects (Part B) for eye-tracking model.

A. parametric coefficients	Estimate	Std. Error	*t*-value	*p*-value
intercept	-1.7043	0.0696	-24.5008	< 0.0001
Experience (Low)	-0.7386	0.0607	-12.1625	< 0.0001
B. smooth terms	edf	Ref.df	*F*-value	*p*-value
random intercept for Item	123.3560	158.0000	3.9149	< 0.0001
tensor product Time by Rating (High Experience)	14.6406	18.3242	62.1472	< 0.0001
tensor product Time by Rating (Low Experience)	13.3119	16.9114	36.9849	< 0.0001

Figure 4.4 displays contour plots of the predicted values for Accent Rating by time at both levels of Experience. In these plots, darker shading indicates decreased looks to target while, lighter shading indicates increased looks. For the High Experience group (left panel), an interesting pattern emerges. Looks to target modulate as Accent Rating increases, manifesting as a decrement in the rate of increase. This modulation starts toward the end of the time window for Accent Ratings beyond the native-like range and begins earlier as non-nativeness increases. The pattern for the Low Experience group (right panel) is somewhat different. For this group, the modulation pattern appears to be shifted further in time, with the first half of the time window showing a depression within the mid-range of Accent Ratings.

The plot on the left for the High Experience group displays a steep growth curve for the most native-like productions, the slope of which gradually decreases as accentedness increases. The plot for the Low Experience group (right panel) displays a similar, albeit flatter, surface, with highly accented tokens producing a minimally sloping curve. Figure 4.5 presents a difference surface (High experience minus Low experience) for the predicted values through time along the accentedness continuum. As can be seen, there is a significant difference in performance between the two groups along the continuum of accentedness and

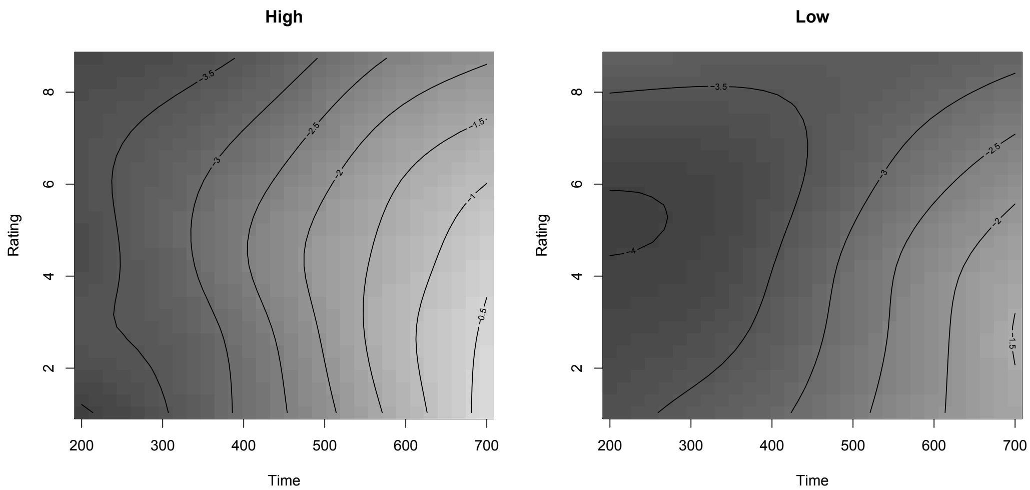

Figure 4.4: Additive effect of Rating and Time by Experience. The right panel displays that of the High Experience group, while the left displays that of the Low Experience group. Darker shading indicates decreased looks to target while, lighter shading indicates increased looks.

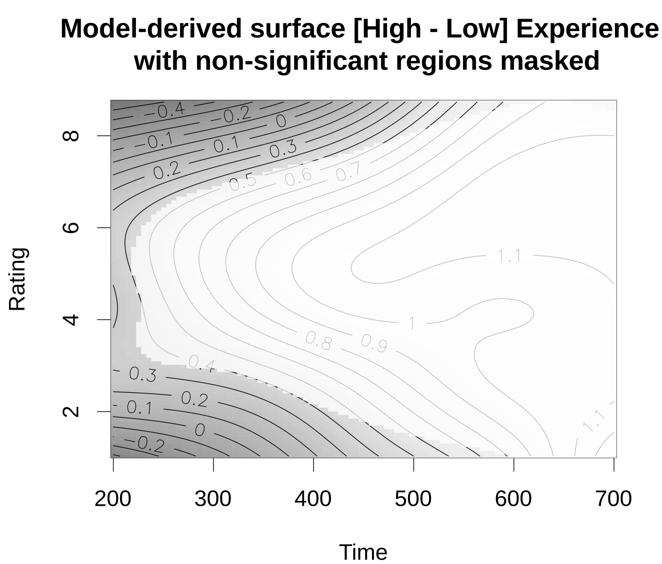

Figure 4.5: Model derived difference surface

time, with the Low Experience group having a flatter surface overall. The shading represents non-significant regions, showing that the difference is significant in the earliest portion of the time window and persist through time at the two edges of the accentedness continuum.

4.3.3 Discussion

This Visual Word eye-tracking experiment aimed to illuminate the roles of gradient foreign accentedness and listener experience on the time-course of auditory word recognition. The model indicates that the time-course of recognition (as evidenced by looks to target) is affected by degree of foreign accentedness. While this has not previously been shown for degree of foreign accentedness, it has been shown for experience with segment-based pronunciation variants (Hanulíková & Weber, 2012). In that study, different second language groups more easily processed the variant with which they had the most experience, not the variant that was most perceptually similar to the target.

The results of the present model are in line with Experiment 1 and show that online word recognition is modulated by gradient foreign accentedness and listener experience. As the auditory stimulus unfolds, certainty of finding the target generally increases. However this certainty appears to decrease as foreign accentedness increases. The pattern of increasing looks shifts further in time indicating that greater accent strength influences the amount of time needed to map the auditory token to the appropriate lexical item. The more deviant a token is from a typical native-like production, the longer it takes to become certain of the target. Interestingly, however, this overall pattern appears to be affected by the amount of experience a given participant has interacting with Chinese-accented speakers.

The modulation across accentedness in both groups appears to begin between ratings three and four. Again, this roughly corresponds with the far end of the rating values assigned to a native English talker and may be considered a nativeness boundary effect. Despite this similarity, experience plays a significant role in the model and group differences indicate that gradient foreign accentedness produces different patterns across time. It appears that in the Low Experience group, the overall pattern is shifted further in time compared to the High Experience group. Additionally, the Low Experience group has a flatter pattern of performance overall. Interestingly, as in the previous experiment, differences emerge between the groups even in the native range of the accentedness continuum. This will be taken up in greater detail in the general discussion.

The results indicate that the time-course of online word recognition is modulated by degree of foreign accentedness of the token as well as listener experience with the accent in question. The fact that the tokens vary from “native" productions along multiple dimensions shows that listeners are attempting to handle the totality of the variability present. High experience listeners are arguably more successful in dealing with it, even from talkers with which they have had no previous experience. This speaks to the accumulation of representations of that variability and generalization over those representations.

4.4 General discussion

Previous research has primarily approached the question of foreign accentedness from a factorial perspective, dividing speech into native, regional, and foreign accent groups (Floccia et al., 2009), or by subdividing foreign accent into high, mid, and low (Witteman et al., 2013). The central contribution of the experiments presented here is the demonstration that gradient foreign accentedness (specifically, Chinese-accented English) influences lexical processing differently along different parts of the continuum of foreign accentedness. Subsequent to this is the demonstration that listener experience with Chinese-accented speakers influences the processing of Chinese-accented English along the same continuum.

The cross-modal identity priming experiment examined the processing of tokens which varied in terms of their degree of foreign accentedness. The results show that gradient foreign-accented words result in differences in the strength with which they prime the intended lexical representation. It has been shown previously that multiple acoustic dimensions and lexical properties influence ratings of foreign accentedness (Levi et al., 2007; Munro, 1993; Porretta et al., 2013). Additionally, these acoustic dimensions correlate with the ratings when they are taken as distances from typical native productions. That the effect of priming is modulated by degree of foreign accentedness indicates that lexical activation strength is dependent on the goodness-of-fit between the token and the lexical representation. Here we see that when accentedness is treated as a continuum rather than discrete factors, gradience in processing emerges.

The visual word eye-tracking experiment, examining the time-course of the effect of accentedness on word recognition, produced similar results. The results show that gradient foreign-accented words result in differences in the certainty with which listeners can identify the intended word. Specifically, as the accentedness of tokens increases, looks to the target decrease in a gradient fashion. This indicates that decreased goodness-of-fit between the token and the lexical representation during this on-line mapping slows word recognition, perhaps by increasing the uncertainty of the message being communicated. While this has not been previously shown with degree of foreign accentedness using native listeners, Hanulíková and Weber (2012) found a similar effect for segmental variants in non-native listeners. They showed that the time-course of non-native listener word recognition was facilitated when presented with a segmental variation consistent with their own foreign accent in comparison to two other possible segmental variants. In effect, this is a result of experience over perceptual similarity.

In both the cross-modal identity priming and eye-tracking experiments presented here, experience with Chinese-accented English produces significantly different patterns in processing. Our measure of experience takes into account participants' current level of interaction with Chinese-accented speakers occurring outside the laboratory. The results indicate that participants who have moderate-to-considerable experience conversing with non-native (in particular Chinese-accented) speakers process accent-related variation differently from those who have little-to-no experience. This point is particularly important as it suggests that these listeners have developed a generalized, speaker-independent adaptation to the variability contained in Chinese-accented English. There has been at least some indication that listeners can develop short-term adaptation to foreign-accented speech with regard to processing time (Clarke & Garrett, 2004; Witteman et al., 2013), as well as an indication that speaker-independent adaptation can occur with regard to intelligibility (Bradlow & Bent, 2008).

Interestingly, in both experiments, the performance of the two experience groups was significantly different within the range of accentedness values corresponding to native talkers.

There is no a priori reason for which differences may occur between the experience groups at the native end of the continuum; however, this occurs in different experimental tasks with different participants. That this pattern emerges may point to an overall influence of experience with variability on the processing of one's native language. The measure of experience utilized in this study may be capturing something in addition to experience with Chinese-accented speakers, for example, overall communicative interaction. That participants in the two groups perform differently may be a result of increased experience with variability in general. That is to say that the more variability a participant has encountered through experience, and subsequently encoded in mental representations of possible instantiations of their native language, the better they perform in tasks related to the native language. Under an episodic account, it would not be unreasonable to think that different participants process even their native language differently, as it allows experience to accumulate over time and modify representations. However, this variation in performance is usually not the subject of study and would be more difficult to accommodate under a processing-based account. Here we have discretized Experience as the primary objective was to study degree of foreign accentedness. That being said, it may very well be that the nature of the differences between these two groups with regard to native productions may be illuminated by specifically examining gradience in experience with language variability at large.

Here we have demonstrated that when accentedness is not discretized, gradience in processing appears and the effect is in the expected direction. What is interesting to note is that, in both experiments, there appears to be some indication of a native range. Degree of accentedness begins to influence processing after a particular point, somewhere around a rating value of four. This roughly corresponds to the upper end of rating values for the native talker in this study. This could be interpreted in favor of the Different Processes Hypothesis (Goslin et al., 2012) which states that native and regional accents recruit processing mechanisms different from those that handle foreign accent-related variation. However, this range of values (i.e., 1 to 4) does not correspond exclusively to the native talker in these experiments. It is particularly important to understand that the distribution of accentedness values for Chinese-accented and native English talkers partially overlap. That is to say, there is no clear boundary by which to delineate native from non-native accentedness, other than a priori knowledge of the talker's origin. Therefore, the results seem to indicate that native-like productions are processed in a more uniform fashion regardless of the nativeness status of the talker, and that productions which fall outside this range of possible native-like variation are processed more gradiently. The Different Processes Hypothesis gives no threshold for determining what constitutes a native versus non-native status. Additionally, it does not predict a gradient response across the continuum of accentedness, nor an effect of listener experience.

Overall, it would seem that the results support a view of processing accent-related variation that is in line with the Perceptual Distance Hypothesis (Clarke & Garrett, 2004; Goslin et al., 2012) which predicts that processing of native, regional, and foreign accents will be a function of their perceptual distance from typical native productions. However, the Perceptual Distance Hypothesis alone cannot explain the effect of listener experience. The concept of perceptual distance implies a static native perceptual reference point. The effect of experience in both the reaction time and eye-tracking data presented here indicate that experience qualitatively changes the likelihood of a given token offering sufficient match as accentedness increases. Additionally, in both cases, a native range emerges, suggesting that tokens most similar to those which they would most likely encounter are processed in a more uniform way. Interestingly, the dip in the reaction times of the High Experience group for mid-range values suggests something about experience. The reaction times to these mid-range values likely represent the distribution of accentedness most commonly experienced. Therefore, if compensation occurs, it can occur for the most typical of productions previously encountered. This suggests that the reference for perceptual distance, for the purposes of processing, is not static. Rather, it appears to incorporate more information through experience.

Representational and processing-based accounts of variation in the speech stream both maintain that listeners attempt to deal with variability to achieve word recognition. How and at which point in time this is done is of crucial difference. Representational accounts would say that the results presented here can be accounted for by mapping the signal directly to the lexical representation and that they align with the claim that phonetic experiences are stored, and ultimately generalized over, allowing for more efficient retrieval. In this way, experience with more variability affects the likelihood of future tokens offering a sufficient match. Processing-based accounts would say that the same results can be accounted for prelexical processing by converting of the phonetic instantiations to abstract phonological categories before being mapped onto the lexical representation. While this study was not designed to conclusively resolve the debate over the nature and extent of lexical representations or the specifics of underlying processes, the results show that foreign-accented speech is in fact processed in a gradient manner, something that has not been previously shown. Additionally the results show that listener experience with the accent in question ultimately modulates processing of foreign-accented speech.

The representational view seems to better account for the effects seen here. Modulation in processing across the continuum of accentedness appears to be an issue of goodness-of-fit to mental representations (including fit to phonemic categories contained in the word). This mapping and evaluation of goodness-of-fit seems to happen rapidly and along multiple dimensions simultaneously. For the processing-based view to account for this gradience, it would require the prelexical processing stage to evaluate the fit of each phoneme, output these weighted measures of fit, and send them to the lexicon in order to activate the closest matched candidate. A representational view would do this on the fly as the potential target representations are activated as the component phonemic information is also available in the lexical representation. How these dimensions in the signal match the probabilistic information in the representation determine how strongly it will be activated in real-time. The evidence from the eye-tracking experiment shows that this is done rapidly, even before the end of the token is heard. While episodic models posit that specific previously-encountered phonetic instantiations of words become encoded, this does not preclude abstraction over those exemplars. In fact, this would be necessary to account for the fact that the participants in these studies have never heard these particular talkers before. The effect of experience here, particularly the facilitation seen for mid-range tokens in the High Experience participants in the cross-modal priming experiment, indicates that group-specific generalization can occur. The processing-based view is less able to account for these group-related effects, which would obligatorily have to be handled prelexically at a segmental level. Therefore, the results seem to generally favor a representational account of variability.

4.5 Conclusions

The results of this study provide converging and complementary evidence regarding the effect of gradient foreign accentedness on lexical processing. Namely, degree of foreign accentedness influences the strength with which words activate the corresponding representations, suggesting that multiple dimensions of the signal map as a constellation of properties onto the representation and the goodness-of-fit of this mapping thus influences the strength of the priming effect. Additionally, the degree of foreign accentedness influences the time-course of word recognition, suggesting that the online mapping of a signal which deviates from a typical production increases the uncertainty of the recognition process, though not ultimately preventing recognition. Both experiments also indicate a significant role of listener experience with Chinese-accented speakers in processing. The measure of listener experience presented here can be considered an indication of long-term, generalized adaptation to Chinese-accented talkers. This supports representational accounts of the mental lexicon and points to a long-term encoding of accent-related variability which may accumulate over time.

References

Adank, P. & McQueen, J. M. (2007). The effect of an unfamiliar regional accent on spoken-word comprehension. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)* (pp. 1925–1928). Dudweiler: Pirrot.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of Hirotugu Akaike* (pp. 199–213). New York: Springer. (Reprinted from Proceedings of the Second International Symposium on Information Theory, B.N. Petrov and F. Caski, eds., Akademiai Kiado, Budapest, 1973, 267–281)

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.

Arnold, D., Wagner, P., & Baayen, R. H. (2013). Using generalized additive models and random forests to model German prosodic prominence. In *Proceedings of Interspeech 2013* (pp. 272–276). Lyon, France.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. H. (2010a). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon*, *5*(3), 436–561.

Baayen, R. H. (2010b). The directed compound graph of English. An exploration of lexical connectivity and its processing consequences. In S. Olson (Ed.), *New impulses in word-formation (Linguistische Berichte Sonderheft 17)* (pp. 383–402). Hamburg: Buske.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474.

Bradlow, A. R. & Bent, T. (2003). Listener adaptation to foreign accented English. In M. Sole, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003)* (pp. 2881–2884). Barcelona, Spain.

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

Burnham, K. P. & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: Springer.

Clarke, C. M. & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, *116*(6), 3647–3658.

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

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*(4), 317–367.

Davies, M. (2008). The Corpus of Contemporary American English (COCA): 400+ million words, 1990–present.

Fischer, B. (1992). Saccadic reaction time: Implications for reading, dyslexia, and visual cognition. In K. Rayner (Ed.), *Eye movements and visual cognition* (pp. 31–45). Springer Series in Neuropsychology. New York: Springer.

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

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

Gaskell, M. G. & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 380–396.

Gass, S. & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning*, *34*(4), 65–87.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1166–1183.

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

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*, *122*(2), 92–102.

Hanulíková, A. & Weber, A. (2012). Sink positive: Linguistic experience with *th* substitutions influences nonnative word recognition. *Attention, Perception, & Psychophysics*, *74*(3), 613–629.

Hastie, T. & Tibshirani, R. (1990). *Generalized additive models*. Monographs on Statistics & Applied Probability 43. Chapman and Hall/CRC.

Huettig, F. & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23–B32.

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

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mulinnex (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego: Academic Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, *7*(3), 279–312.

Kryuchkova, T., Tucker, B. V., Wurm, L. H., & Baayen, R. H. (2012). Danger and usefulness are detected early in auditory lexical processing: Evidence from electroencephalography. *Brain and Language*, *122*(2), 81–91.

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2007). Speaker-independent factors affecting the perception of foreign accent in a second language. *Journal of the Acoustical Society of America*, *121*(4), 2327–2338.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*(4), 372–380.

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

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

Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, *36*(1), 39–66.

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

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. L. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam: John Benjamins.

Pierrehumbert, J. B. (2003a). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2–3), 115–154.

Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mulinnex (Eds.), *Talker variability in speech processing* (pp. 9–32). San Diego: Academic Press.

Porretta, V., Kyröläinen, A.-J., & Tucker, B. V. (2013). Influences on perceived foreign accentedness: Acoustic distances and lexical neighborhoods, June 23–28, 2013. International Cognitive Linguistics Conference (ICLC) 12. Edmonton, AB.

R Development Core Team. (2014). *R: A language and environment for statistical computing*. Version 3.1.0. R Foundation for Statistical Computing. Vienna, Austria.

Ranbom, L. J. & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*(2), 273–298.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *The Quarterly Journal of Experimental Psychology*, *55A*, 1339–1362.

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in spanish-accented speech. *Journal of the Acoustical Society of America*, *125*(5), 3306–3316.

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

Tremblay, A. & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151–173). London: The Continuum International Publishing Group.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, *53*(4), 510–540.

Walsh, M., Möbius, B., Wade, T., & Schütze, H. (2010). Multilevel exemplar theory. *Cognitive Science*, *34*(4), 537–582.

Wieling, M., Montemagni, S., Nerbonne, J., & Baayen, R. H. (2014). Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and sociodemographic variation using generalized additive mixed modeling. *Language*, *90*(3), 669–692.

Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLoS ONE*, *6*(9), 1–14.

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, *75*(3), 537–556.

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Boca Raton: Chapman & Hall/CRC Press.

Wood, S. N. (2014). *mgcv: Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation*. R package version 1.7-29.

Zuur, A., Ieno, E., Walker, N., Saveliev, A., & Smith, G. (2009). *Mixed effects models and extensions in ecology with R*. New York: Springer.

Chapter 5

General discussion and conclusions

Throughout this dissertation, I have explored gradient foreign accentedness and its effect on lexical processing using a variety of psycholinguistic methods. Across three studies, I have found that both the perception and processing of foreign-accented speech reflect the gradience of accentedness. In the first study regarding the perception of foreign accentedness, this gradience appears to reflect the goodness-of-fit between the acoustic properties of a particular token and an average native production as well as the relationship among sounds and words within the mental lexicon. In the subsequent two studies examining online processing, converging and complementary evidence indicates that when perceived foreign accentedness is treated as a continuum, gradience emerges. As tokens become less native-like, processing is impacted; however, this impact is not the same for all listeners. Experience interacting with accented speakers changes the way that the variability inherent to foreign-accented speech influences processing. This has consequences for the way in which variability might be represented within the mental lexicon and subsequently utilized during processing.

In this chapter I first summarize the results of the studies presented in this dissertation. Then I discuss the implications of those results for both our understanding of the phenomenon of foreign accentedness as well as mental representations of variability. Following the discussion, I acknowledge the limitations of the studies and propose some possible directions for future research.

5.1 Summary of results

5.1.1 Study 1

The degree of perceived foreign accentedness was explored through ratings of native- and Chinese-accented English words. In particular, the influences of both acoustic (speaker-dependent) and lexical (speaker-independent) variables on these ratings were investigated. In this study, acoustic distances were employed as speaker-dependent variables, measuring the magnitude difference between the talker and the native acoustic reference point for vowel formants as well as the relationship between vowel and word durations. Here, the native reference point was taken to be a typical native value (i.e., the mean of native speaker values) for a given dimension. speaker-independent variables were measures related to the lexical items themselves such as lexical frequency, phonotactic probability, and phonological neighborhood density.

The results indicate that, indeed, both acoustic and lexical variables affect ratings of perceived foreign accentedness suggesting that both are involved in the matching of a particular token to the representation of a typical native production. Specifically, as the distance of both F1 and F2 from the native acoustic reference (i.e., typical native value) increases, accent strength is judged to be greater. Additionally, the relationship between temporal measures of vowel and word duration also positively correlates with the strength of perceived accentedness. As talkers move closer to the durational template of the word, they are perceived as less accented. These results provide support for the view that foreign accentedness is affected by perceptual distance, the magnitude difference between a particular production and a native acoustic reference point.

With regard to lexical properties, the model indicates the influence of both neighborhood density and phonotactic probability, such that as both neighborhood density and phonotactic probability increase, perceived accentedness decreases. For example, *bit*, which has a high phonotactic probability and many neighbors, is rated as less accented than *through* which has a low phonotactic probability and few neighbors. This suggests that denser neighborhoods may facilitate the matching process especially when the phonemic sequence is probable. In addition, it is interesting to note that lexical frequency fell out of this model during the fitting process; therefore it was not possible to confirm the effect seen in previous research (Levi et al., 2007).

The results of this experiment suggest that the goodness-of-fit of a given token, appears to be driven by perceptual distance of the acoustic signal from a typical native-like production, but is not dependent solely on these properties. This fit appears to also be affected in part by probabilistic information about a particular lexical item and its relationship to other items in the lexicon. This is similar to work by Vitevitch et al. (1997) who found that ratings of well-formedness of English nonse words reflect the probability of phonotactic sequences and to that of Hay et al. (2004) who found that similar results are gradient with this probability. The matching process that takes place appears to be multi-faceted in nature, such that the shorter the acoustic distances and the better the fit to probable lexical properties, the less accented a token is judged to be. So, perceived foreign accentedness may reflect the matching that occurs across both acoustic and lexical properties, rather than strict perceptual distance. These results, taken together, point to a lexicon that encodes multidimensional and distributional properties of acoustic realizations of words in the native accent as well as the probabilistic relationship between sounds and words which might affect the judgment of accentedness.

5.1.2 Study 2

The influence of gradient foreign accentedness on early auditory processing was investigated by examining electrophysiological response. Specifically, the study examined the brain's early and involuntary response to native- and Chinese-accented English words. The results suggest that both degree of accentedness and listener experience with Chinese-accented speakers affect signatures of early processing when listening to foreign-accented speech.

The results of the model indicate modulation in the three ERP components under consideration (i.e., N100, P200, and Phonological Mapping Negativity) across the accentedness continuum with differences between listener experience groups. With regard to the N100, the amplitude of the High Experience group gradually declines as accentedness increases, while that of the Low Experience group increases. These patterns of modulation indicate differences in the allocation of perceptual and attentional processing resources, suggesting habituation in the High Experience group, and heightened attention in the Low Experience group. As for the P200, the amplitude of the High Experience group displays a slight decrease for the most native-like items, while that of the Low Experience group displays a similar increase moving up the ratings continuum, though this increase does not persist. These results indicate differences between the two groups in language related perceptual processes which are sensitive to repetition and experience. This suggests that for the High Experience group, tokens across the continuum could be registered as repetitions, while this is only the case for midrange items in the Low Experience group. Thus, the Low Experience group has learned over time to map accent-related variability onto perceptual memory representations, but not for the most accented items as in the High Experience group.

The results of the Phonological Mapping Negativity (PMN) indicate that for the High Experience group amplitude begins to gradually decrease approximately halfway through the continuum, while for the Low Experience group amplitude gradually increases for productions moving away from the native end of the continuum, then remaining relatively constant. While the PMN has been likened to a goodness-of-fit component, these results suggest something different. The PMN is also sensitive to expectations about sounds and the results here suggest that the two Experience groups have very different expectations about variability. The Low Experience group appears to expect non-native-like productions to mismatch their representation, while the High Experience group seems to relax expectations as words become more highly accented. Thus the PMN might not index goodness-of-fit during bottom-up auditory processing per se, but rather top-down expectations about the incoming signal.

The results of this experiment suggest that long-term exposure to accent variability can change involuntary, non-behavioral response during online auditory processing. This may be due to enhanced representations or more developed processing strategies in listeners with greater experience interacting with a specific accent. Furthermore, it highlights the importance of listener experience when examining electrophysiological response to foreign-accented speech.

5.1.3 Study 3

Two experiments were presented to investigate the influence of gradient foreign accentedness on the priming of lexical representations and spoken word recognition, respectively. This was accomplished using cross-modal identity priming and visual word eye-tracking. The two experiments provide converging and complementary results.

The results of the reaction time model indicate that the facilitatory effect of the priming is modulated by degree of foreign accentedness. In general, tokens at the low end of accentedness continuum show relatively flat reaction times and as degree of foreign accentedness increases, facilitation of subsequent written targets decreases. This suggests that lexical activation strength is dependent on the goodness-of-fit between the token and the target lexical representation. Interestingly, listener experience with Chinese-accented speakers significantly influences this effect of facilitation. Specifically, Low Experience listeners show longer reaction times across the board, with the High Experience group slowing to meet them only at the most accented end of the continuum. While the two groups begin to show an increase in reaction times at approximately the same point along the continuum, the patterns diverge. The Low Experience group displays a gradual increase in reaction times across the continuum while the reaction times of the High Experience listeners begin to rise, then fall in the midrange of the continuum and finally spike at the far end. It seems that experience with accent-specific variability facilitates lexical activation, however it appears to do so in a non-linear fashion.

The results of the eye-tracking model indicate that gradiently foreign-accented words result in differences in the time-course of word recognition. Specifically, as accentedness increases, looks to the target decrease gradiently. This modulation of slope begins sooner in time as accentedness increases. As in the reaction time model, listener experience has a significant effect. In particular, the Low Experience group produces a pattern that is shifted further in time, with the first half of the time window (200–450 ms) showing a depression within the mid-range of Accent Ratings. The High Experience group on the other hand displays a steep curve for the most native-like productions, which gradually decreases in slope as accentedness increases. In comparison to the High Experience group, the Low Experience group displays a flatter response across the board. This can be considered an effect of the certainty with which listeners can identify the intended word. As uncertainty grows, the slope of looks to the target becomes more shallow. This is seen in both groups, but is more pronounced within the Low Experience group. So, it seems that decreased goodness-of-fit between the token and the lexical representation during this on-line mapping slows word recognition, perhaps by increasing the uncertainty of the message.

Taken together, the results of these two experiments provide evidence for the influence of gradient foreign accentedness on lexical processing along with listener experience. This suggests two things. First, during the mapping of a token to a lexical representation, the overall goodness-of-fit (along multiple dimensions) affects activation in a gradient manner. Second, increased deviation from typical native productions increases uncertainty gradiently during word recognition. Both of these are facilitated by listener experience, suggesting that listeners can develop, at least to some degree, a generalized adaptation to Chinese-accented English.

5.2 General discussion

5.2.1 Gradience

It has generally been acknowledged that some non-native speakers have stronger foreign accents than others. Some previous work has examined the acoustic properties of foreign-accented speech (Flege & Hillenbrand, 1984; Munro, 1993; Wayland, 1997; Baker et al., 2011) and possible influences on perceived foreign accentedness (Levi et al., 2007; Munro, 1993; Wayland, 1997). The work presented here, corroborates previous results and indicates in a single model that perceived gradient accentedness is driven in part by acoustic properties of the token and in part by properties of the lexical item itself.

Multiple dimensions of the acoustic signal work in tandem to influence perceived foreign accentedness as evidenced in the ratings study. Ratings appear to be a judgment of the overall goodness-of-fit of a given token to a typical native production, instantiated here as a native acoustic reference value for each dimension. Because each acoustic dimension is considered as a distance measure from this native reference, it suggests that the magnitude of deviation is the driving force for the magnitude of accentedness. The judgment is then based on how well the token fits the representation of what is considered an acceptable native production across multiple dimensions. In addition, it appears that judgments are affected by the relationship between the dimensions. This matching process is directly in line with the concept of perceptual distance (cf. Clarke & Garrett, 2004; Floccia et al., 2006; Goslin et al., 2012) as a way of understanding the phenomenon of gradient foreign accentedness. In order for this to occur, listeners would likely need representations that richly define what constitutes a typical (or at least, acceptable) native-like production along and between multiple dimensions (cf. Pierrehumbert, 2001).

Beyond this, lexical variables are general properties of the mental lexicon (Pierrehumbert, 2003b) and influence the matching process. Probabilistic and distributional properties of words affect the judgment of the fit. These properties lead to a reduction in perceived accentedness perhaps by making the matching process easier, most likely through increased probability of the phonemic sequence and activation within dense lexical neighborhoods. Thus, acoustic fit and general properties of the lexicon work together in determining degree of foreign accentedness. It may be that perceived foreign accentedness is in fact an index of the ease of the match.

The gradience in accentedness, evident from the ratings, has not been previously investigated in regard to lexical processing. In all three processing experiments presented above, gradience in processing emerges when foreign accentedness is taken to be a continuum. This variability modulates processing, sometimes in a non-linear fashion. As seen in the reaction time and eye-tracking experiments, increased perceived accentedness affects processing in the expected direction. That is, the effect of facilitatory priming and the time-course of word recognition are negatively affected as the accent becomes stronger. This is a rather clear indication that perceptual distance is at work in lexical processing. Thus the mapping of the acoustic signal to lexical representation is driven by perceptual distance along many dimensions, suggesting a process by which the fit ultimately affects activation of the lexical representation.

The evidence from the ERP experiment, however, provides a slightly different view on the matter. While indicating that gradience emerges along the continuum, it is not entirely clear that the early auditory-evoked components should index mapping and mismatch per se. These are likely indices of other general perceptual and attentional processes, which also vary with increased accentedness. The Phonological Mapping Negativity has been described as a component which indexes goodness-of-fit for bottom-up auditory speech processing (Goslin et al., 2012; Newman & Connolly, 2009). However, in the ERP data presented here, it does not appear to consistently index goodness-of-fit. Instead, the PMN more likely indexes expectation about the signal rather than strict bottom-up mapping to phonological categories. Similar effects of expectation in language processing have been shown previously. Language users can rapidly adapt their expectations to structural statistics, overcoming processing disadvantages (Fine, Jaeger, Farmer, & Qian, 2013). This type of expectation-based processing has also been reported for the P600 component in response to foreign-accented speech (cf. Hanulíková et al., 2012). It was shown that listeners can adjust their expectations about the speech produced by non-native talkers, such that when they hear errors in grammatical gender produced by non-native talkers, no P600 effect is found for the violation. A similar relaxation of expectations may be at play in the PMN for different listeners when processing gradiently accented speech.

Overall, the results of the studies contained in this dissertation clearly point to the role of gradience in the perception of foreign accentedness and the processing of the accentedness continuum. An investigation of processing along this continuum of variability represents the first of its kind and is a significant contribution to our understanding of how listeners process non-native productions. It has been proposed that foreign-accented speech recruits different processing mechanisms and this view has garnered some support in the literature (Floccia et al., 2006; Goslin et al., 2012). The Different Processes Hypothesis, states that native and non-native speech will be processed by different mechanisms which allow for speech that is consistent with the native language to be normalized without additional intervention. This suggests a clear delineation between what is native and what is not, resulting in processing mechanisms that are “all-or-nothing". One potentially problematic issue in the studies which support this view is that listener experience is often confounded with perceptual similarity.

In both Floccia et al. (2006) and Goslin et al. (2012), listeners, by nature of geography and media, have far more exposure to perceptually similar regional accents than to perceptually dissimilar regional or foreign accents. The results of the studies presented here provide little evidence for the Different Processes Hypothesis, insofar as a clear distinction between native and non-native speech processing for all listeners. Instead, when accentedness is viewed as a continuum, a different picture emerges. In particular, patterns of processing are not discontinuous between native and non-native, but rather, gradient.

The concept of perceptual distance (cf. Clarke & Garrett, 2004; Floccia et al., 2006; Goslin et al., 2012) has been put forth as a means of understanding the phenomenon of gradient foreign accent and has been related to processing through the Perceptual Distance Hypothesis. The hypothesis states that processing of speech which is more perceptually similar to native speech (e.g., regional or highly proficient non-native speech) will be an attenuated version of that for speech which is perceptually dissimilar (e.g., less proficient non-native speech). This idea is generally in line with the data presented here; productions that are less native-like are processed slower, with a gradient slope for intermediate productions. However, the concept of perceptual distance implies a static and singular native perceptual reference point. The data presented here show that listener experience with Chinese-accented English modulates the effect of the accentedness continuum. Experience changes the likelihood of a given token offering sufficient match as accentedness increases. This suggests that the reference for perceptual distance, for the purposes of processing, is not static. Rather, it appears to incorporate more information through experience.

5.2.2 Experience

Apart from the gradience that emerges across the accentedness continuum, differences in processing arise between levels of listener experience interacting with Chinese-accented speakers. This appears to fall in line with the relatively few studies which have examined listener experience with regard to processing of non-native productions. Weber et al. (2011) found that accented tokens consistent with a non-native listener's own accent facilitate L2 word processing, while no facilitation was seen for tokens in a different accent. Additionally, Hanulíková and Weber (2012) obtained a similar result during visual word eye-tracking in which participants most easily processed pronunciation variants consistent with their own accent, in effect the one with which they had the most experience. Thus, linguistic experience with a foreign accent affects the ability to recognize words carrying this accent. In a study of native listeners' ability to adapt to foreign-accented speech, Witteman et al. (2013) showed that the speed with which listeners adapt perceptually to foreign-accented speech is dependent upon listener familiarity with the accent and the strength of that accent.

The effect of both short and long term exposure has been shown to influence intelligibility of foreign-accented speech as well. Within this adaptation, listeners benefit from exposure to multiple speakers of the same accent type; both Bradlow and Bent (2008) and Sidaras et al. (2009) have shown that talker-independent adaptation for the specific accent can occur with laboratory-based training. What is seen in the present studies is that even when exposure takes place outside of a laboratory setting, some degree of accent generalization seems to have occurred. This generalization may be the result of learning across experience helping listeners to facilitate processing of variability by making it unnecessary to repeat adaptation, talker by talker. While it remains unclear exactly how listeners get from experience with individual talkers to group-level generalizations, this overarching experience clearly affects processing. It may be that non-native accent generalization occurs in the same or similar manner as generalization within the native accent. Experience appears to qualitatively change the likelihood of a given token offering sufficient match as accentedness increases. Here, this effect on processing has been shown across multiple methods. We see that experience changes not only the pattern of behavioral performance, but also the pattern of involuntary electrophysiological response, indicating an underlying change in neural activity, rather than simply a change in performance-based measures. It should be noted, however, that experience does not entirely undo the effect of gradient accentedness. Gradience in processing remains to some extent even in listeners with a high level of experience. What this suggests is that generalizations may be drawn which help to mitigate the effect of accentedness, but may not completely nullify it. This stands to reason. Because listeners have the greatest amount of contact with native-like productions in their daily life, these would be most strongly represented. Non-native productions would therefore always constitute the minority of encountered productions.

A curious finding in the present set of studies is that experience with Chinese-accented speakers appears to influence the processing of tokens which were rated to be most native-like. This is an interesting result as it was seen across different paradigms (i.e., electrophysiological, behavioral, and eye-tracking). In each experiment, there were different participants, suggesting that this is not simply something inherent to a particular sample. So, there appears to be something about experience with Chinese-accented speakers that changes even native processing. It may be that the measure of experience utilized here captures something in addition to experience with Chinese-accented speakers, such as overall communicative interaction (regardless of who the interlocutor is). This result of experience with non-native speakers might be similar to how native processing can change with the acquisition of another language (van Hell & Dijkstra, 2002). Changes of this sort may also result from interaction with non-native variability. This suggests that merely encountering greater variability may modify the structure and accessibility of even the most native of representations. It should be noted that this result would not be predicted by either the Different Processes Hypothesis or the Perceptual Distance Hypothesis, which do not take experience (of any sort) into account. Further investigation is needed to begin to understand exactly how experience at large might impact processing in the native accent.

The results presented in this dissertation provide supporting evidence for the role of listener experience in the processing of non-native variability. This effect of experience suggest that listeners subsequently utilized language experience to facilitate processing. This is perhaps achieved through decreased uncertainty about how the signal might map to stored representations; greater experience with a specific type of variability increases the likelihood of a given token offering sufficient match to the intended word. Additionally, experience may allow listeners to alter their expectations about the speech produced by different talkers and make generalizations over variable input. Because this effect of experience is increasingly seen in the literature, it necessitates not only further treatment in subsequent research, but an understanding of how and where this experience might accumulate and be encoded. Ultimately, this will require that processing models incorporate the learning of variability.

5.2.3 Representation and recognition

Experience-dependent processing comes to bear on our understanding of lexical representation and spoken word recognition. Specifically, how and where is experience with a particular kind of variability represented so that it might influence speech processing? The two general accounts (i.e., Representational and Processing-based) present very different ideas about this. While the experiments presented in this dissertation were not designed to specifically test either account, the results do speak to speech processing in general.

Processing-based accounts maintain that a single, abstract and possibly phonologically underspecified representation is present for each word in the lexicon. Under this view, extraneous noise and irrelevant variability is stripped from the acoustic signal resulting in a normalized representation before matching to the lexicon (Lahiri & W. Marslen-Wilson, 1991; McClelland & Elman, 1986; Norris et al., 2000). In this way, variability in the speech stream must be handled by prelexical processing and any retuning due to perceptual learning occurs at the prelexical phonemic level (McQueen et al., 2006; Norris, McQueen, & Cutler, 2003). More recent models, such as Shortlist B, allow for more within-phoneme variability by having phoneme likelihood as the input rather than a string of discrete phonemes. This calculates the probability of a phoneme given the evidence and multiplies across these probabilities to derive the likelihood of the string.

Representational accounts, and in particular episodic (exemplar-based) models (Goldinger, 1996; Goldinger, 1998; Johnson, 1997; Johnson, 2006), instead hold that lexical representations contain detailed acoustic traces. So, rather than involving a speaker normalization process, accessing lexical entries involves matching the incoming signal to any number of acoustic memories or to a quantized perceptual space. Under this view, similarity to an appropriate set of reference exemplars is calculated and weighted such that exemplar and category activation arise (Johnson, 1997; Johnson, 2006). In this way, experience with foreign-accented speakers encodes a myriad of acoustic and indexical information about each talker, and similarity amongst these talkers and between these talkers and native talkers can be assessed on a range of properties, which may be more or less salient in a particular language.

Within the processing-based account, experience is either not present or relegated to the speech perception system. It is there that the input maps onto a stipulated representation. For perceptual learning to occur, lexical verification must feedback to the category level in order to train categorization (McQueen et al., 2006; Norris et al., 2003). This would allow for category distributions to change, however, it is unclear how broader distributional patterns, for example in word duration, might be handled during recognition. Additionally, this type of approach would not allow other speaker characteristics to cluster and influence recognition. For this type of model to account for the findings in this dissertation, the prelexical level of processing would necessarily have to evaluate the fit of each phoneme (or its probability), output these weighted measures for all the individual component phonemes of the word, and feed them to the lexicon in order to activate the likely candidate as in Norris and McQueen's Shortlist B. Therefore, gradience would have to emerge as direct a result of the phoneme-by-phoneme probability calculation. Any effect of learning would then modify the probability density function for a particular perceptual dimension. This could account for the present findings, but does not necessarily account for how probabilities might co-occur (or not co-occur) across groups of speakers (i.e., a particular accent group). For example, it may be that a particular set of values each have relatively low probabilities overall. However, the co-occurrence of those values may be probable within a particular group.

Models which integrate time, experience/learning, expectation, and generalization into word recognition (and speech processing more generally) more readily account for the effect of experience and gradience in processing. Adaptive Resonance provides a useful framework for incorporating these aspects into speech processing (Goldinger & Azuma, 2003; Grossberg et al., 1997; Grossberg et al., 2004; Grossberg & Myers, 2000; Grossberg & Stone, 1986; Johnson, 2006; McLennan & Luce, 2005). This framework lays out a system which can ultimately learn to make and/or modify categorical distinctions, such that speech units are emergent rather than stipulated (Goldinger & Azuma, 2003). In this way, both speech unit categories and social categories emerge as a natural consequence of experience and similarity. Johnson's resonance-based exemplar model calculates distances from exemplars, sums activation across exemplars, and determines evidence for category membership. This model accounts for gender identity for stereotypical and non-stereotypical voices and shows how ambiguous voices can be drawn into one category or the other based on the distribution of exemplars. By allowing socially constructed representations to influence speech processing, listeners might then develop expectations about what they are hearing. A similar mechanism may be at work for handling the variability inherent to the foreign-accented speech of any given group.

5.2.4 The possible nature of representations and processing

Given the data reported in this dissertation, one might speculate about the nature of linguistic representations stored in the mind and how spoken input is mapped to them during processing. As stated above, processing-based and representation-based accounts provide two perspectives on the matter, each with a unique contribution to the issue and each with its own benefits and drawbacks. Processing-based accounts hold that a lexical representation is singular and abstract and that input is necessarily processed in discrete modular steps, one of which requires the input to map onto sublexical units. Representation-based accounts, such as exemplar models, hold that a lexical representation is a cloud-like entity comprised of numerous acoustic memories and that input is mapped directly to that cloud.

Given previous work as well as the results presented here, it seems that a representation-based account can more readily explain the effects. Taking up this view, it is then possible to speculate about the potential shape and structure of linguistic representations in the mental lexicon. Indeed, representations are likely multi-dimensional entities, the shape of which is determined by the experience of the possessor. These clouds would encode a myriad of information along perceptual and contextual dimensions. For example, listeners appear sensitive to familiar voices (Goldinger, 1996). Similar exemplars of a particular word would congregate together within the space and more frequent exemplars would be more strongly represented. In this way, even rather dissimilar exemplars would be represented and connected in the multi-dimensional spaces. The reason for this is that experience has indicated that those exemplars still constitute possible instantiations of the word in question. For example, highly accented exemplars could be represented in the cloud at a particular distance from the more strongly represented native exemplars. These accented exemplars, particularly ones of the same accent type, might then congregate within the space to form their own sub-cluster, once experience starts to accumulate. Thus, the cloud-like representations constitute weighted structures for a particular lexical item along numerous dimensions.

These clusters and sub-clusters might then serve as the basis for generalizations (i.e., abstractions) to emerge over the likelihood of co-occurrence (or non-co-occurrence). So the generalization may correspond to a probable (perhaps, mean) instantiation, which while derived from experience, may not index a specific exemplar. In this sense, representations contain both specified and abstract information at multiple levels of linguistic structure. However, it is key to note that the abstract information is necessarily emergent from experience (Pierrehumbert, 2003a). Then, all the information one obtains from experience in the real world can accumulate and link to linguistic representations. It may be that extra-linguistic contextually- and socially-derived categories are also allowed to form and link to the information contained in the (sub)clusters and generalizations.

With regard to sublexical units, it seems clear that language users have at least some knowledge of the possible sounds that make up their language. Therefore it is important for the representation to account for their existence. Sublexical representations are likely part of the same (or similar) multi-dimensional space which contains lexical representations. The occurrence of perceptual properties relevant to these sounds would be represented and form their own clusters and abstractions. These sublexical representations, like all others, would be emergent and allowed to change over time with experience, thus allowing for the learning of acoustic variability. These representational entities could then link to all other representations, including other sublexical representations (e.g., phonotactic relations), lexical representations (e.g., constituency relations), and larger group representations (e.g., social/contextual relations). In this way, lexical items sharing a sublexical unit might be similarly affected by changes in the distributions of its properties, even if those properties have never been experienced for one of the words in question. Additionally, because they could also link to higher level representations, linguistic and social contextual information could influence expectations about the sublexical representation and its associated generalization (Johnson, 2006).

Here, it is important to mention briefly how processing might take place. Given the interconnected structure described above, processing would happen in a direct, rather than modular, manner. As auditory information begins to arrive, it would activate all representations at all levels of linguistic structure with which it is consistent. This means that the auditory information corresponding to all or part of a phoneme would begin to activate the representation of the phoneme, the representations of phonemes with which it co-occurs, the representations of the lexical items which contain the phoneme in that particular position, the semantic representations of those words, and possibly larger group information related to the particular properties of that instantiation. In this way processing can be viewed as spreading activation (Anderson, 1983) in which all levels of representation begin to be activated and subsequent information narrows the possibility of the lexical target.

Thus, in the case of variability in the signal, rich representations which have been shaped by experience can better predict the intended message. Additionally, because all levels are involved, co-occurrence probabilities as well as listener expectations can influence activation and ultimately the outcome. In this way, there are no modules, nor any point in time at which lexical access occurs (McLennan, Luce, & Charles-Luce, 2003). This view takes a stance that is consistent with studies which show that semantic information and frequency distributions are available almost instantaneously (Kryuchkova et al., 2012; MacGregor & Shtyrov, 2013), something that one would not expect under a modular structure. Lastly, this type of structure and processing is highly parsimonious as it would provide the same (or similar) account for processing at all levels of linguistic structure (Walsh et al., 2010). This is in line with the adaptive resonance framework which is beginning to provide neurologically-based and plausible models of language processing. In this framework representations come in the form of chunks, which are learned sets of associated features that vary in size ranging from sublexical unit to lexical word. When input is consistent with a chunk, it begins to resonate and it is this resonance that then constitutes the percept (Grossberg, 1986). The network that allows for resonance is emergent and continually developing through learning. Therefore, in the view of the author, a representation-based account employing adaptive resonance theoretic principles seems to be the most reasonable and consistent with the effects explored in this work.

5.3 Limitations

While the results of the three studies provide similar and converging evidence for the processing of gradiently accented non-native speech, there are some limitations which should be acknowledged. First, it is important to note that in these studies I have dealt with monosyllabic words, from the same set of talkers, presented in isolation. Therefore I have only examined lexical processing without the influence of context. This was done partially for greater control; however, it is known that the presence of context can affect spoken word recognition and may do so immediately (Dahan & Tanenhaus, 2004; Revill, Tanenhaus, & Aslin, 2008). It is possible that semantic context or even additional spoken input from the talker may mitigate the effect of foreign accent by establishing sufficiently large constraints on listener expectations.

Second, here I have only examined native- and Mandarin-accented English. It is possible that results might look different if a more or less familiar accent were used. Additionally, I have not examined the effect of other native varieties of English (e.g., New Zealand or Scottish English) within the continuum of accentedness. It is certainly conceivable that other native varieties of English may be rated higher than some non-native accents or that listeners index that a particular talker speaks a native variety.

Third, across the studies I have employed a rather course-grained measure of listener experience. The measure was a self-reported estimation of current interaction with Chinese-accented speakers which was used to group experience levels. Therefore, rather than having a predefined criterion for experience, the definition of experience level was based on the sample. While the distributions of the measure are similar across the studies (see Figures A.1 and A.2 in the Appendix), in order to begin to understand the effect of listener experience in a more nuanced manner, it would be necessary to examine experience as a continuous predictor.

5.4 Future directions

The research contained in this dissertation, while helping to elucidate some aspects of foreign-accented speech, also opens new avenues for future research. In particular, two main areas emerge as potential topics for further investigation: context and experience. First is an investigation of the role of context in the processing of foreign-accented speech. By comparing words in isolation to those in sentential (or possibly discourse) context, it is possible to further our understanding of the influences of top-down information. Additionally, context could be defined more broadly to include extra-linguistic information (e.g., beliefs about the talker based on visual information or communicative objective) which might alter listeners' expectations about or motivations for processing.

Second is a deeper investigation of the effect of experience. Given the results of the experiments presented here, it would be interesting to explore in greater detail the effect of experience on the processing of both non-native and native speech. Specifically, an investigation of experience with non-native speakers as a continuous measure would provide evidence for how the accumulation of experience affects processing of foreign-accented speech. It could also be fruitful to explore additional ways of estimating listener experience with non-native speakers. Additionally, by examining the effect of experience (at large), it is possible to investigate how the processing of native speech might be affected by experience with variability, for example by the amount of overall communicative interaction with all types of speakers (both native and non-native).

5.5 Conclusion

The three studies presented in this dissertation aimed to explore the phenomenon of gradient foreign accentedness and its influence on lexical processing. The ratings experiment shows that perceived foreign-accent strength is influenced by both goodness-of-fit along multiple acoustic dimensions and the structure of the lexicon itself. The ERP, behavioral, and eye-tracking experiments provide converging evidence that when accentedness is treated as a continuum, gradience in processing emerges. Gradience in this regard suggests that goodness-of-fit to the representation influences processing. This supports a view of foreign accent processing which acknowledges degree of magnitude difference between the production and the representation and the role of listener experience in shaping representations. Specifically, foreign accent processing is likely driven by perceptual distance; but this distance is ultimately modulated by listener experience with the accent in question. This suggests that the reference point is not static and that experience in mapping variant productions may accumulate over time and change expectations about incoming signals in order to learn to accommodate the variability. These findings as a whole suggest that lexical representations are robust and can evolve to incorporate, at least to some extent, long-term encoding of accent variation for talker-independent generalization.

References

Anderson, D. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 261–295.

Baker, R. E., Baese-Berk, M., Bonnasse-Gahot, L., Kim, M., Van Engen, K. J., & Bradlow, A. R. (2011). Word durations in non-native English. *Journal of Phonetics*, *39*(1), 1–17.

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

Clarke, C. M. & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, *116*(6), 3647–3658.

Dahan, D. & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 498–513.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, *8*(10), 1–18.

Flege, J. E. & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America*, *76*(3), 708–721.

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

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1166–1183.

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

Goldinger, S. D. & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, *31*(3–4), 305–320.

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*, *122*(2), 92–102.

Grossberg, S. (1986). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines: Vol. 1. Speech perception* (pp. 187–294). New York: Academic Press.

Grossberg, S., Boardman, I., & Cohen, M. (1997). Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(2), 481–503.

Grossberg, S., Govindarajan, K. K., Wyse, L. L., & Cohen, M. A. (2004). ARTSTREAM: A neural network model of auditory scene analysis and source segregation. *Neural Networks*, *17*(4), 511–536.

Grossberg, S. & Myers, C. W. (2000). The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychological Review*, *107*(4), 735–767.

Grossberg, S. & Stone, G. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review*, *93*(1), 46–74.

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887.

Hanulíková, A. & Weber, A. (2012). Sink positive: Linguistic experience with *th* substitutions influences nonnative word recognition. *Attention, Perception, & Psychophysics*, *74*(3), 613–629.

Hay, J., Pierrehumbert, J., & Beckman, M. E. (2004). Speech perception, well-formedness and the statistics of the lexicon. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic interpretation: Papers in laboratory phonology 6* (pp. 58–74). Cambridge, UK: Cambridge University Press.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mulinnex (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego: Academic Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499.

Kryuchkova, T., Tucker, B. V., Wurm, L. H., & Baayen, R. H. (2012). Danger and usefulness are detected early in auditory lexical processing: Evidence from electroencephalography. *Brain and Language*, *122*(2), 81–91.

Lahiri, A. & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, *38*(3), 245–294.

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2007). Speaker-independent factors affecting the perception of foreign accent in a second language. *Journal of the Acoustical Society of America*, *121*(4), 2327–2338.

MacGregor, L. J. & Shtyrov, Y. (2013). Multiple routes for compound word processing in the brain: Evidence from EEG. *Brain and Language*, *126*(2), 217–229.

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

McLennan, C. T. & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 306–321.

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 539–553.

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

Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, *36*(1), 39–66.

Newman, R. L. & Connolly, J. F. (2009). Electrophysiological markers of pre-lexical speech processing: Evidence for bottom–up and top–down effects on spoken word processing. *Biological Psychology*, *80*(1), 114–121.

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

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*(3), 299–370.

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

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. L. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam: John Benjamins.

Pierrehumbert, J. B. (2003a). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2–3), 115–154.

Pierrehumbert, J. B. (2003b). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probability theory in linguistics* (pp. 177–228). Cambridge, MA: The MIT Press.

Revill, K. P., Tanenhaus, M. K., & Aslin, R. N. (2008). Context and spoken word recognition in a novel lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1207–1223.

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in spanish-accented speech. *Journal of the Acoustical Society of America*, *125*(5), 3306–3316.

van Hell, J. G. & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, *9*(4), 780–789.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, *40*(1), 47–62.

Walsh, M., Möbius, B., Wade, T., & Schütze, H. (2010). Multilevel exemplar theory. *Cognitive Science*, *34*(4), 537–582.

Wayland, R. (1997). Non-native production of Thai: Acoustic measurements and accentedness ratings. *Applied Linguistics*, *18*(3), 345–373.

Weber, A., Broersma, M., & Aoyagi, M. (2011). Spoken-word recognition in foreign-accented speech by L2 listeners. *Journal of Phonetics*, *39*(4), 479–491.

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, *75*(3), 537–556.

Bibliography

Adank, P. & McQueen, J. M. (2007). The effect of an unfamiliar regional accent on spoken-word comprehension. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)* (pp. 1925–1928). Dudweiler: Pirrot.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of Hirotugu Akaike* (pp. 199–213). New York: Springer. (Reprinted from Proceedings of the Second International Symposium on Information Theory, B.N. Petrov and F. Caski, eds., Akademiai Kiado, Budapest, 1973, 267–281)

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.

Anderson, D. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 261–295.

Arnold, D., Wagner, P., & Baayen, R. H. (2013). Using generalized additive models and random forests to model German prosodic prominence. In *Proceedings of Interspeech 2013* (pp. 272–276). Lyon, France.

Atienza, M., Cantero, J. L., & Dominguez-Marin, E. (2002). The time course of neural changes underlying auditory perceptual learning. *Learning & Memory*, *9*(3), 138–150.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. H. (2010a). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon*, *5*(3), 436–561.

Baayen, R. H. (2010b). The directed compound graph of English. An exploration of lexical connectivity and its processing consequences. In S. Olson (Ed.), *New impulses in word-formation (Linguistische Berichte Sonderheft 17)* (pp. 383–402). Hamburg: Buske.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Baker, R. E., Baese-Berk, M., Bonnasse-Gahot, L., Kim, M., Van Engen, K. J., & Bradlow, A. R. (2011). Word durations in non-native English. *Journal of Phonetics*, *39*(1), 1–17.

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, *29*(4), 639–647.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474.

Bernal, J., Harmony, T., Rodríguez, M., Reyes, A., Yáñez, G., Fernández, T., Galán, L., Silva, J., Fernández- Bouzas, A., Rodríguez, H., Guerrero, V., & Marosi, E. (2000). Auditory event-related potentials in poor readers. *International Journal of Psychophysiology*, *36*(1), 11–23.

Bien, H. & Zwitserlood, P. (2013). Processing nasals with and without consecutive context phonemes: Evidence from explicit categorization and the N100. *Frontiers in Psychology*, *4*(21), 1–12.

Boersma, P. & Weenink, D. (2011). Praat: Doing phonetics by computer [Version 5.3.61].

Bradlow, A. R. & Bent, T. (2003). Listener adaptation to foreign accented English. In M. Sole, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003)* (pp. 2881–2884). Barcelona, Spain.

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

Brunellière, A. & Soto-Faraco, S. (2014). The interplay between semantic and phonological constraints during spoken-word comprehension. *Psychophysiology*, *52*(1), 46–58.

Burnham, K. P. & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York: Springer.

Carretié, L., Mercado, F., Tapia, M., & Hinojosa, J. A. (2001). Emotion, attention and the 'negativity bias', studied through event-related potentials. *International Journal of Psychophysiology*, *41*(1), 75–85.

Clarke, C. M. & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America*, *116*(6), 3647–3658.

Connolly, J. F. & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6*(3), 256–266.

Connolly, J. F., Service, E., D'Arcy, R. C., Kujala, A., & Alho, K. (2001). Phonological aspects of word recognition as revealed by high-resolution spatio-temporal brain mapping. *Cognitive Neuroscience and Neuropsychology*, *12*, 237–243.

Crowley, K. E. & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology*, *115*(4), 732–744.

Curran, T. & Dien, J. (2003). Differentiating amodal familiarity from modality-specific memory processes: An ERP study. *Psychophysiology*, *40*(6), 979–988.

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

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*(4), 317–367.

Dahan, D. & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 498–513.

Davies, M. (2008). The Corpus of Contemporary American English (COCA): 400+ million words, 1990–present.

Derwing, T. M. & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition*, *20*(1), 1–16.

Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience*, *21*(10), 1893–1906.

Diaz, M. T. & Swaab, T. Y. (2007). Electrophysiological differentiation of phonological and semantic integration in word and sentence contexts. *Brain Research*, *1146*, 85–100.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46.

Dufour, S. & Nguyen, N. (2014). Access to talker-specific representations is dependent on word frequency. *Journal of Cognitive Psychology*, *26*(3), 256–262.

Evans, K. M. & Federmeier, K. D. (2007). The memory that's right and the memory that's left: Event-related potentials reveal hemispheric asymmetries in the encoding and retention of verbal information. *Neuropsychologia*, *45*(8), 1777–1790.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE*, *8*(10), 1–18.

Fischer, B. (1992). Saccadic reaction time: Implications for reading, dyslexia, and visual cognition. In K. Rayner (Ed.), *Eye movements and visual cognition* (pp. 31–45). Springer Series in Neuropsychology. New York: Springer.

Flege, J. E. (1980). Phonetic approximation in second language acquisition. *Language Learning*, *30*(1), 117–134.

Flege, J. E. (1984). The detection of French accent by American listeners. *Journal of the Acoustical Society of America*, *76*(3), 692–707.

Flege, J. E. & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America*, *76*(3), 708–721.

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

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

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

Gaskell, M. G. & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, *12*(5–6), 613–656.

Gaskell, M. G. & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 380–396.

Gaskell, M. G. & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology*, *45*(2), 220–266.

Gass, S. & Varonis, E. M. (1984). The effect of familiarity on the comprehensibility of nonnative speech. *Language Learning*, *34*(4), 65–87.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1166–1183.

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

Goldinger, S. D. & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, *31*(3–4), 305–320.

Goslin, J., Duffy, H., & Floccia, C. (2012). An ERP investigation of regional and foreign accent processing. *Brain and Language*, *122*(2), 92–102.

Grossberg, S. (1986). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines: Vol. 1. Speech perception* (pp. 187–294). New York: Academic Press.

Grossberg, S., Boardman, I., & Cohen, M. (1997). Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(2), 481–503.

Grossberg, S., Govindarajan, K. K., Wyse, L. L., & Cohen, M. A. (2004). ARTSTREAM: A neural network model of auditory scene analysis and source segregation. *Neural Networks*, *17*(4), 511–536.

Grossberg, S. & Myers, C. W. (2000). The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychological Review*, *107*(4), 735–767.

Grossberg, S. & Stone, G. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review*, *93*(1), 46–74.

Hagoort, P. & Brown, C. M. (2000). ERP effects of listening to speech: Semantic ERP effects. *Neuropsychologia*, *38*(11), 1518–1530.

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, *24*(4), 878–887.

Hanulíková, A. & Weber, A. (2012). Sink positive: Linguistic experience with *th* substitutions influences nonnative word recognition. *Attention, Perception, & Psychophysics*, *74*(3), 613–629.

Hasher, L. & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, *39*(12), 1372–1388.

Hastie, T. & Tibshirani, R. (1990). *Generalized additive models*. Monographs on Statistics & Applied Probability 43. Chapman and Hall/CRC.

Hay, J., Pierrehumbert, J., & Beckman, M. E. (2004). Speech perception, well-formedness and the statistics of the lexicon. In J. Local, R. Ogden, & R. Temple (Eds.), *Pho-*

netic interpretation: Papers in laboratory phonology 6 (pp. 58–74). Cambridge, UK: Cambridge University Press.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*. New Series, *182*(4108), 177–180.

Huettig, F. & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23–B32.

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

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. *Journal of the Acoustical Society of America*, *117*(2), 896–907.

Iverson, P. & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, *97*(1), 553–562.

Jaeger, F. T. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*(1), 23–62.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mulinnex (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego: Academic Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*(4), 485–499.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, *7*(3), 279–312.

Kryuchkova, T., Tucker, B. V., Wurm, L. H., & Baayen, R. H. (2012). Danger and usefulness are detected early in auditory lexical processing: Evidence from electroencephalography. *Brain and Language*, *122*(2), 81–91.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990.

Kutas, M. & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647.

Lahiri, A. & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, *38*(3), 245–294.

Lee, J. Y., Harkrider, A. W., & Hedrick, M. S. (2012). Electrophysiological and behavioral measures of phonological processing of auditory nonsense V–CV–VCV stimuli. *Neuropsychologia*, *50*(5), 666–673.

Levi, S. V., Winters, S. J., & Pisoni, D. B. (2007). Speaker-independent factors affecting the perception of foreign accent in a second language. *Journal of the Acoustical Society of America*, *121*(4), 2327–2338.

Luce, P. A. & Lyons, E. A. (1998). Specificity of memory representations for spoken words. *Memory & Cognition*, *26*(4), 708–715.

Luce, P. A. & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36.

MacGregor, L. J. & Shtyrov, Y. (2013). Multiple routes for compound word processing in the brain: Evidence from EEG. *Brain and Language*, *126*(2), 217–229.

Marslen-Wilson, W. & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1–71.

Martin, B. A., Sigal, A., Kurtzberg, D., & Stapells, D. R. (1997). The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds /ba/ and /da/. *Journal of the Acoustical Society of America*, *101*(3), 1585–1599.

Martin, B. A., Tremblay, K. L., & Stapells, D. R. (2007). Principles and applications of cortical auditory evoked potentials. In R. Burkard, J. Eggermont, & M. Don (Eds.), *Auditory evoked potentials: Basic principles and clinical application*. Point (Lippincott Williams & Wilkins). Lippincott Williams & Wilkins.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*(4), 372–380.

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

McLennan, C. T. & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 306–321.

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 539–553.

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

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

Misra, M. & Holcomb, P. J. (2003). Event-related potential indices of masked repetition priming. *Psychophysiology*, *40*(1), 115–130.

Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, *36*(1), 39–66.

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

Näätänen, R. & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375–425.

Newman, R. L. & Connolly, J. F. (2009). Electrophysiological markers of pre-lexical speech processing: Evidence for bottom–up and top–down effects on spoken word processing. *Biological Psychology*, *80*(1), 114–121.

Newman, R. L., Connolly, J. F., Service, E., & McIvor, K. (2003). Influence of phonological expectations during a phoneme deletion task: Evidence from event-related brain potentials. *Psychophysiology*, *40*(4), 640–647.

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

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*(3), 299–370.

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

Osterhout, L. & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806.

Park, H. (2013). Detecting foreign accent in monosyllables: The role of L1 phonotactics. *Journal of Phonetics*, *41*(2), 78–87.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. L. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam: John Benjamins.

Pierrehumbert, J. B. (2003a). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, *46*(2–3), 115–154.

Pierrehumbert, J. B. (2003b). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probability theory in linguistics* (pp. 177–228). Cambridge, MA: The MIT Press.

Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*(2), 191–215.

Pisoni, D. B. (1997). Some thoughts on “normalization” in speech perception. In K. Johnson & J. W. Mulinnex (Eds.), *Talker variability in speech processing* (pp. 9–32). San Diego: Academic Press.

Plaut, D. C. & Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWhinney (Ed.), *The emergence of language* (pp. 381–415). Carnegie Mellon Symposia on Cognition Series. Taylor & Francis.

Porretta, V., Kyröläinen, A.-J., & Tucker, B. V. (2013). Influences on perceived foreign accentedness: Acoustic distances and lexical neighborhoods, June 23–28, 2013. International Cognitive Linguistics Conference (ICLC) 12. Edmonton, AB.

Porretta, V., Kyröläinen, A.-J., & Tucker, B. V. (submitted). Perceived foreign accentedness: Acoustic distances and lexical properties. *Attention, Perception, & Psychophysics*.

Porretta, V. & Tucker, B. V. (2012). Predicting accentedness: Acoustic measurements of Chinese-accented English. *Canadian Acoustics*, *40*(3), 34–35.

R Development Core Team. (2014). *R: A language and environment for statistical computing*. Version 3.1.0. R Foundation for Statistical Computing. Vienna, Austria.

Ranbom, L. J. & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*(2), 273–298.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *The Quarterly Journal of Experimental Psychology*, *55A*, 1339–1362.

Revill, K. P., Tanenhaus, M. K., & Aslin, R. N. (2008). Context and spoken word recognition in a novel lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1207–1223.

Ross, B. & Tremblay, K. L. (2009). Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. *Hearing Research*, *248*(1–2), 48–59.

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

Sanders, L. D. & Neville, H. J. (2003). An ERP study of continuous speech processing II. Segmentation, semantics, and syntax in non-native speakers. *Cognitive Brain Research*, *15*(3), 214–227.

Shahin, A. J., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *The Journal of Neuroscience*, *23*(13), 5545–5552.

Shahin, A. J., Roberts, L. E., Pantev, C., Trainor, L. J., & Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport*, *16*(16), 1781–1785.

Sheehan, K. A., McArthur, G. M., & Bishop, D. V. (2005). Is discrimination training necessary to cause changes in the P2 auditory event-related brain potential to speech sounds? *Cognitive Brain Research*, *25*(2), 547–553.

Shrout, P. E. & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428.

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in spanish-accented speech. *Journal of the Acoustical Society of America*, *125*(5), 3306–3316.

Steinhauer, K. & Connolly, J. F. (2008). Event-related potentials in the study of language. In B. Stemmer & H. A. Whitaker (Eds.), *Handbook of the neuroscience of language* (pp. 91–104). New York: Elsevier.

Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., & Howard, M. A. (1999). Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *Journal of Neurophysiology*, *82*(5), 2346–2357.

Stelmack, R. M., Saxe, B. J., Noldy-cullum, N., Campbell, K. B., & Armitage, R. (1988). Recognition memory for words and event-related potentials: A comparison of normal and disabled readers. *Journal of Clinical and Experimental Neuropsychology*, *10*(2), 185–200.

Strand, E. A. (2000). *Gender stereotype effects in speech processing* (Unpublished doctoral dissertation, Ohio State University, Columbus, Ohio).

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

Tong, Y., Melara, R. D., & Rao, A. (2009). P2 enhancement from auditory discrimination training is associated with improved reaction times. *Brain Research*, *1297*, 80–88.

Tremblay, A. (2011). *icaOcularCorrection: Independent Components Analysis (ICA) based eye-movement correction*. R package version 1.3.

Tremblay, A. & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 151–173). London: The Continuum International Publishing Group.

Tremblay, A., Baayen, R. H., & Hendrix, P. (2013). *eRp: Pre-processing, generalized additive mixed-effects modeling, and visualization of event-related brain potential and field (ERP/ERF) data.* R package version 0.9.4.2.

Tremblay, K. L. & Kraus, N. (2002). Auditory training induces asymmetrical changes in cortical neural activity. *Journal of Speech, Language, and Hearing Research*, *45*(3), 564–572.

Tremblay, K. L., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training. *Ear and Hearing*, *22*(2), 79–90.

Tremblay, K. L., Shahin, A. J., Picton, T., & Ross, B. (2009). Auditory training alters the physiological detection of stimulus-specific cues in humans. *Clinical Neurophysiology*, *120*(1), 128–135.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, *53*(4), 510–540.

van Hell, J. G. & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, *9*(4), 780–789.

Vaughan Jr., H. G. & Ritter, W. (1970). The sources of auditory evoked responses recorded from the human scalp. *Electroencephalography and Clinical Neurophysiology*, *28*(4), 360–367.

Vitevitch, M. S. & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*(3), 374–408.

Vitevitch, M. S. & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers*, *36*, 481–487.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, *40*(1), 47–62.

Walsh, M., Möbius, B., Wade, T., & Schütze, H. (2010). Multilevel exemplar theory. *Cognitive Science*, *34*(4), 537–582.

Wayland, R. (1997). Non-native production of Thai: Acoustic measurements and accentedness ratings. *Applied Linguistics*, *18*(3), 345–373.

Weber, A., Broersma, M., & Aoyagi, M. (2011). Spoken-word recognition in foreign-accented speech by L2 listeners. *Journal of Phonetics*, *39*(4), 479–491.

Weil, S. (2001). Foreign accented speech: Encoding and generalization. *Journal of the Acoustical Society of America*, *109*, 2473–2473.

Weinberger, S. H. (2013). Speech accent archive, George Mason University [http://accent.gmu.edu].

Wieling, M., Montemagni, S., Nerbonne, J., & Baayen, R. H. (2014). Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and sociodemographic variation using generalized additive mixed modeling. *Language*, *90*(3), 669–692.

Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLoS ONE*, *6*(9), 1–14.

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, *75*(3), 537–556.

Wolach, I. & Pratt, H. (2001). The mode of short-term memory encoding as indicated by event-related potentials in a memory scanning task with distractions. *Clinical Neurophysiology*, *112*(1), 186–197.

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Boca Raton: Chapman & Hall/CRC Press.

Wood, S. N. (2014). *mgcv: Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation*. R package version 1.7-29.

Wood, S. N., Scheipl, F., & Faraway, J. J. (2013). Straight forward intermediate rank tensor product smoothing in mixed models. *Statistics and Computing*, *23*(3), 341–360.

Zuur, A., Ieno, E., Walker, N., Saveliev, A., & Smith, G. (2009). *Mixed effects models and extensions in ecology with R*. New York: Springer.

Appendix A

Supplementary information by chapter

A.1 Supplementary information for Chapter 2

Table A.1: Experimental monosyllabic words selected from the NU Wildcat Corpus wordlist.

bait	booze	peep	sheets
bat	boss	poop	shoes
bees	bought	pop	shoots
beet	box	pot	shop
bet	but	put	shots
bit	cheese	risks	soup
bite	chirp	rubbed	splash
block	chop	screen	spring
boat	girl	searched	stripe
boot	legs	sheep	through

A.2 Supplementary information for Chapter 3

Table A.2: Demographic and language experience questionnaire.

	Question	Response type
1	Gender	Male/Female
2	Age	Free response
3	Which is you dominant hand?	Right/Left/Both
4	Do you have normal hearing?	Yes/No
5	What is your dialect of English?	Free response
6	Are you bilingual?	Yes/No
7	Which accent(s) did you hear during the experiment?	Free response
8	Do you have a Chinese background?	Yes/No
9	Have you ever studied Chinese?	Yes/No
10	On a weekly basis, how often do you interact with non-native speakers of English?	0(Never)-10(Daily)
11	What percentage of those interactions include speakers with a Chinese accent?	Free response

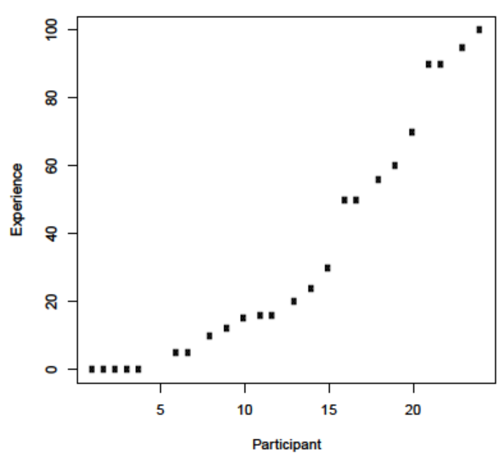

Figure A.1: Distribution of Experience values among participants ($n = 24$).

A.3 Supplementary information for Chapter 4

Table A.3: Cross-modal identity priming stimuli.

Identity	Unrelated	Pseudoword 1	Pseudoword 2
bait	mug	chuth	ciff
bat	loop	fek	cype
bees	vowed	detch	whid
beet	sash	jodd	jush
bet	cap	neech	phiff
bit	soon	gos	yav
bite	noon	joof	hass
block	cross	freaf	spem
boat	thin	viff	kav
boot	dumb	thafe	ziege
booze	lame	sheg	thean
boss	dish	coove	gim
bought	grew	fiece	zean
box	hall	maff	jief
but	his	keme	mome
cheese	ring	voun	voke
chirp	tuft	stuce	blegg
chop	rude	keef	kaif
girl	cost	susk	scoosh
legs	picked	dran	brop
peep	stow	gomb	chutt
poop	fawn	tebe	tetch
pop	cat	tib	kib
pot	gear	deg	fiss
put	came	lail	vev
risks	proved	spleet	splaff
rubbed	pools	blesh	clett
screen	plant	crolt	glomp
searched	shields	blunk	fleant
sheep	thumb	medge	zodd
sheets	cooked	brube	droove
shoes	laid	futh	thabe
shoots	gazed	flome	croal
shop	moon	deace	jife
shots	leaned	braif	danch
soup	shore	jish	lafe
splash	fringe	dronk	brenge
spring	text	claint	brelf
stripe	bland	blinch	phlalk
through	much	zof	sazz

Table A.4: Visual Word Paradigm quadruplets.

Target	Rhyme competitor	Onset competitor	Distractor
bait	mate	bike	loop
bat	chat	bend	mill
bees	seas	bids	dome
beet	feat	bin	doll
bet	net	ban	zone
bit	hit	beach	soon
bite	height	bake	noon
block	shock	blood	dream
boat	coat	bound	ring
boot	root	bud	mess
booze	fuse	bun	leach
boss	sauce	bomb	cat
bought	taught	bond	neck
box	fox	board	heat
but	cut	ball	can
cheese	breeze	chin	loud
chirp	burp	chink	nag
chop	hop	chore	reef
girl	pearl	guess	seem
legs	begs	lands	moon
peep	seep	pin	loot
poop	coupe	pun	shag
pop	cop	port	juice
pot	dot	punk	lane
put	foot	pool	mean
risks	disks	reads	flag
rubbed	dubbed	rocked	jam
screen	dean	script	flat
searched	perched	sucked	mob
sheep	leap	shift	dull
sheets	meets	ships	rod
shoes	jews	shows	fair
shoots	suits	shores	lend
shop	drop	shut	fill
shots	lots	shucks	meal
soup	loop	sum	fame
splash	trash	splurge	crude
spring	king	spread	club
stripe	hype	strain	mug
through	few	throat	name

Table A.5: Demographic and language experience questionnaire.

	Question	Response type
1	Gender	Male/Female
2	Age	Free response
3	Which is you dominant hand?	Right/Left/Both
4	Do you have normal hearing?	Yes/No
5	What is your dialect of English?	Free response
6	Are you bilingual?	Yes/No
7	Which accent(s) did you hear during the experiment?	Free response
8	Do you have a Chinese background?	Yes/No
9	Have you ever studied Chinese?	Yes/No
10	On a weekly basis, how often do you interact with non-native speakers of English?	0(Never)-10(Daily)
11	What percentage of those interactions include speakers with a Chinese accent?	Free response

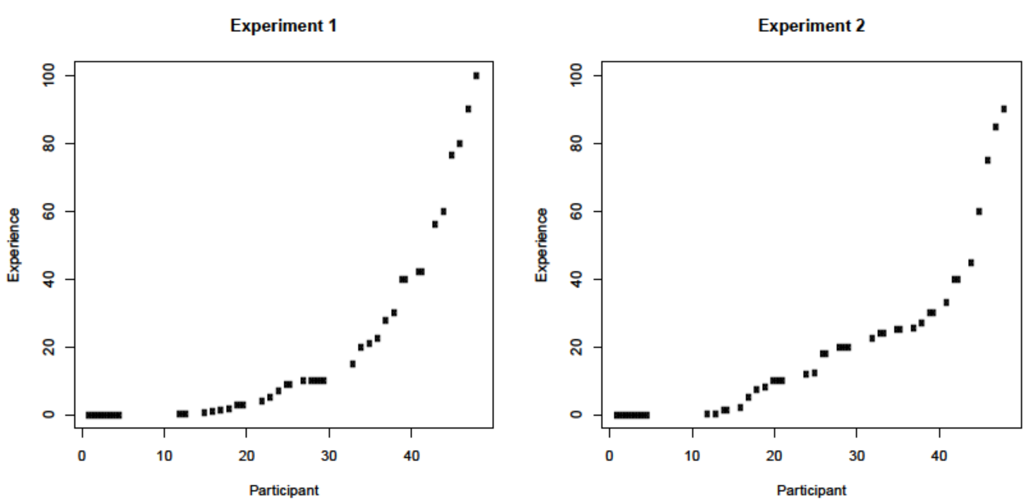

Figure A.2: Distribution of Experience values among participants in Experiments 1 (n = 48) and 2 (n = 48).