Accents and Adverse Conditions: Investigating the Effects of Semantic and Phonemic Information on Accented Speech Comprehension

by

Timothy Andrew Woerle

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

Master of Science

Department of Psychology
University of Alberta

© Timothy Andrew Woerle, 2024

Abstract

Under normal conversational conditions, listeners are typically able to adapt to a speaker's accent or idiosyncrasies with relative ease. While this is a well-established phenomenon, the exact mechanisms which allow fast adaptation are not yet entirely understood, and there are a variety of models which offer competing accounts of the underlying process. In this study, I examined a few of these models by subjecting listeners to artificially accented speech under different adverse circumstances which specifically degraded phonemic information or semantic information to examine how adaptation to a given speaker relies on these respective sources of information. The analysis revealed likely interactions between accent condition and semantic information, and between accent condition and phonemic information, with little support for an interaction between semantic information and phonemic information. The implications of these findings as they relate to existing models are discussed.

Preface

This thesis is an original work by Timothy Woerle. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board, Project Name "Investigating Accented Speech", No. 00114318, August 29 2023.

Acknowledgements

I would like to thank my supervisor, Dr. Pete Dixon, for his assistance and guidance throughout the research process. I would also like to thank Dr. Juhani Järvikivi for his insight and advice in psycholinguistic research. Finally, I would like to thank the members of my committee, Dr. Jeremy Caplan, Dr. Alona Fyshe, and Dr. Chris Westbury, for their feedback and insight on this research project.

Contents

Introduction 1
Creating and Norming Stimuli 7
Method 9
Materials. 9
Participants. 12
Procedure. 12
Analysis. 13
Results 13
Judgements of Accented Speech 15
Method 15
Materials. 15
Participants. 17
Design. 17
Procedure. 17
Analysis. 19
Results 21
Discussion 26
Conclusion 30
References 32

Appendix A 37

List of Tables

Table 1: Examples of Full Sentences, Respective Target Words, and Phoneme Swaps 11

Table 2: Cell Means for Each Combination of Accent, Context Condition, and Noise Condition 24

Table 3: Marginal Means for the Main Effect of Time 25

List of Figures

Figure 1: Histogram of Target Word Accuracies and Quantile Cutoffs 14

Figure 2: Histogram of Target Word Accuracies and Quantile Cutoffs 22

Figure 3: Log Odds Correct as a Function of Accent and Context Conditions 23

Introduction

When engaging in conversation, we are usually able to parse the meaning in our conversational partner's speech without any effort. Occasionally, however, this automatic processing of meaning can break down when listening to a speaker with an unfamiliar idiosyncrasy in their speech. Perhaps the most noticeable type of idiosyncrasy is an unfamiliar accent, either of a speaker of a foreign language the listener has not encountered before or of a speaker of an unfamiliar dialect of the listener's own language. Usually listeners are able to adapt to an unfamiliar accent fairly quickly (Bradlow & Bent, 2008; Clarke & Garret, 2004; Sidaras et al., 2009) and are able to carry out the conversation normally, although not without some residual processing cost (Brown et al. 2020). While in most cases this type of adaptation occurs in the span of just a few minutes, the process can be hampered by a degradation of the incoming information, whether it be phonemic information (the acoustic features of the speech), lexical information (such as familiarity with a particular spoken word), or semantic information (understanding of the broader context of the speech).

While accented speech itself could be considered a form of degradation mainly at the phonemic level, and to a lesser extent the lexical and semantic levels as a result of dialectical jargon and grammatical errors respectively, it is unique among other sources of degradation in that the listener can readily adapt to it (Mattys et al., 2012). Compared to other types of degradation which involve unpredictable differences between the expected and received speech signal (e.g., listening to someone speak in a noisy room), idiosyncratic accents involves systematic changes which allow for rapid adaptation. By examining how a listener is able to correctly perceive idiosyncratic speech, we can investigate the processes behind speech perception more generally, identify what types of information are utilized in speech perception, and examine how that information is related to prior information.

Early models of speech perception were static in that they would always produce the same output when presented with the same stimuli. For example, Massaro's Fuzzy Logical Model of Perception (FLMP; Massaro, 1989; Massaro & Cohen, 1987) posits a feature analysis process in which a speech perception system analyzes various features of the speech (including both visual and phonemic features), integrates them, and arrives at a decision on what is being spoken. Other early speech perception models such as the TRACE model (McClelland & Elman, 1986) instead utilized interactive activation processes in which speech stimuli causes activation or inhibition of corresponding nodes, which themselves cause activation or inhibition in adjacent nodes similar to other neural network models. Notably, McClelland and Elman highlighted that any model of speech perception must be able to take into account the fact that human listeners are often able to comprehend speech in adverse listening conditions with little difficulty. The models themselves, therefore, must also be able to arrive at the correct perception of even significantly degraded or dysfluent speech stimuli. While this is certainly true, these early models neglected the fact that in addition to taking degraded and dysfluent speech into account, they must also consider the fact that perception of idiosyncratic speech improves over time with exposure to the stimuli. This adaptive behaviour necessitates a more flexible model, in which repeated exposure to idiosyncratic stimuli results in improved performance over time. For a neural network model such as TRACE, this might require an online mechanism to alter the weights of the excitatory and inhibitory connections as the stimuli are being perceived. Alternatively, for a feature analysis model such as FLMP, the prototypical features would need to be expanded or altered based on feedback to allow them to match with the perceived stimuli. Despite this limitation, interactive activation models such as TRACE do make good predictions of results and are still in common use (e.g., Chen & Mirman, 2012; Gao et al., 2019; Luthra et al., 2021). However, to further our understanding of speech perception, our models must include the ability to adapt to idiosyncratic speech as well as other types of adverse listening conditions.

Although less prevalent, such adaptive models do exist. The active processing framework (Heald & Nusbaum, 2014) is one such approach, created specifically to address the lack of adaptability in the traditional “passive” models of speech perception. “Passive” in this context refers to models such as those previously discussed, which will invariably produce the same lexical output given the same auditory input, regardless of context, goals, or hypotheses. (See Diehl et al., 2004 for a more in-depth review of traditional “passive” models.) In contrast, active processing refers to processes which incorporate some degree of cognitive processing in the form of top-down or bottom-up feedback during speech processing. It should be noted that the type of feedback described in the active processing framework differs from that of something like the TRACE model. While TRACE and other neural network models do incorporate feedback between levels of processing, the connections themselves are static and will produce the same output given the same input and, by that token, are considered passive models under the active processing framework. An active process, in contrast, involves attentionally driven processes to make either top-down or bottom-up corrections to auditory input based on context, goals, or other information not contained solely in the speech signal.

As an example of how an active processing framework may work, consider a listener who hears a speaker with an unfamiliar accent make an utterance which could be perceived as either "rain" or "lane" depending on how the listener categorizes the first phoneme. Under the active processing framework, the listener may know that the may know that they are listening to a non-native speaker whose articulation of the /r/ phoneme differs from that of a native speaker, and subconsciously attend to auditory features which make the "rain" interpretation more likely. By incorporating this type of feedback into a model, a simulated listener would be able to adjust their perception on-line based on the current conditions, as well as update their knowledge of said conditions, such as a speaker's accent, with further exposure. Note that the active processing approach does not propose a specific model. Instead, it offers a framework by which cognitive resources, particularly attention, can be employed to provide feedback on speech stimuli and allow a listener to adapt to distorted, degraded, or novel stimuli.

Interestingly, Elman and McClelland (1986) did propose a version of the TRACE model which incorporates active processing at the featural level. Specifically, the updated version employs connections not only between feature and phoneme nodes, but also between phoneme nodes and the existing connections between feature and phoneme nodes. These additional connections allow the immediate phonemic context (i.e., the phonemes surrounding the one which is currently being perceived) to modulate the weights between connections based on predictable patterns of variability within a phoneme based on its immediate phonemic context. This is, essentially, adaptation to noise within phonemic features over time. However, the time frame over which adaptation would occur in this model is very short, as it is concerned only with incorporating surrounding feature information rather than information encountered earlier in speech. As a result, this proposal focused primarily on resolving noise within phonemic categories (e.g., categorizing the /b/ in “ball” and “crib” as the same phoneme despite their acoustic differences), rather than sources of environmental noise or source noise to which listeners are also readily able to adapt. Notably, the updated model relies on systematic variability in acoustic characteristics within a given language or dialect, which does not explain how listeners are able to adapt to noise which is completely novel, such as listening to a speaker with an unfamiliar accent.

The ideal adapter framework (Kleinschmidt & Jaeger, 2015) provides a well-defined model of adaptation to idiosyncratic speech. Broadly, this model integrates probabilistic learning which allows listeners to recognize stimuli which are similar to those they have encountered before, generalize previous situations to presently encountered stimuli, and, finally, when these first two strategies fail, to adapt to novel situations. Crucially, this model explains the ability of listeners to generalize learning across speakers of a particular group (e.g., Bradlow & Bent, 2008; Kraljic & Samuel, 2007; Xie & Myers, 2017) as well as adaptation to particular speakers (e.g., Creel & Bregman, 2011; Trude & Brown-Schmidt, 2013) within a single framework. The framework operates under a belief-updating approach, in which listeners employ an internal generative model which predicts the likely acoustic features which would make up a given word. Listeners compare how well the categories within their current model predict the incoming speech information and proceed accordingly by either recognizing that the speech is familiar and can be predicted from an existing generative model, recognizing that the speech is similar to a previously encountered situation or group and updating their model accordingly, or recognizing that the speech is novel and must be adapted to by updating the generative model. It is the constant updating of these generative models which allows for adaptation to novel speech under the ideal adapter framework.

Critically, most past attempts to explain idiosyncratic speech adaptation have focused solely on the featural and phonemic level. This is true both in terms of the source of the feedback for the change which is occurring, as well as the nature of the adaptation itself. It is, of course, necessary that the final outcome of adaptation must involve some change at the featural and/or phonemic levels, whether this involves changes to signal normalization, phoneme categorization, decision criteria, or some combination of these three systems (Xie et al., 2023). While we do not yet know which of these systems (or combination thereof) is being altered when adapting to idiosyncratic speech (Xie et al., 2023), some change must be occurring at this level in order for listeners to improve their recognition of idiosyncratic speech over time because phonemic differences are one of the defining features of such speech. However, one must consider not only the final outcome of adaptation but also the types of information which are taken into account as adaptation is occurring. In this regard, much of the research surrounding speech adaptation has neglected other key sources of information in favour of focusing on featural and phonemic input, despite the fact that understanding speech entails far more than simply parsing the acoustic properties of the speech signal (Winn, 2018). Indeed, much of the past research has demonstrated the importance of semantic/lexical information (Onifer & Swinney, 1981; Swinney, 1979; Van de Ven et al., 2009), prior knowledge (Bradlow & Pisoni, 1999), and even visual information (Green et al., 1991; McGurk & MacDonald, 1976) when processing and comprehending speech.

Incorporating these additional factors beyond the acoustic properties of speech is therefore crucial to a complete model of idiosyncratic speech adaptation and, by extension, speech perception more broadly. Apart from the recent computational model put forward by Xie et al. (2023), which groups external factors into a single input at the decision-making level called "response biases," there has been little attention paid to the role of non-acoustic factors in idiosyncratic speech adaptation. In this thesis, I intend to investigate the contribution of semantic context to speech adaptation, the recognition of idiosyncratic speech more generally, and the relationship between semantic context and phonemic information during the speech recognition process.

To examine the relationship among these variables, I created novel stimuli that varied in idiosyncratic phonemic usage, semantic constraint, and masking noise. Participants then completed a listening transcription task with these stimuli. Accuracy scores from this listening task were used to assess the relationship between context, noise, and accent, as well has how performance changed over time. In this thesis, I will first describe how the sentences were created and normed for semantic constraint. Then, I will describe the manipulation of noise, the procedures for eliciting the responses, and the results. Finally, conclusions and implications are discussed.

Creating and Norming Stimuli

The first consideration when generating the stimuli was the matter of how to operationalize idiosyncratic speech as a factor. I chose to investigate a simple artificial analogue to naturally accented speech which contained variation from General American English in only a single major factor. While much research in the past has used sentences spoken by non-native English speakers, or native English speakers with accents not common to Western Canada (e.g., Bradlow & Bent, 2008; Hau et al., 2020; Paquette-Smith et al., 2021), this method has some notable drawbacks which would be detrimental to the goals of the present study. First among these problems is that using real speakers and naturally accented speech introduces multiple factors simultaneously which can be difficult to disentangle. This difficulty arises partly from the fact that there is no formal definition of what specific features of speech are involved in accented speech. For instance, Munro and Derwing (1995, p. 289) posit that foreign-accented speech is "non-pathological speech that differs in some noticeable respects from native speaker pronunciation norm," which they go on to specify typically includes segmental and prosodic differences. More recently, Baese-Berk et al. (2020) discuss several features, both segmental and suprasegmental, which may influence perceptions of accented speech. However, these features vary between the particular native and foreign languages being compared, the particular dialect being spoken, and even between different speakers and listeners. While this variability makes it effectively impossible to create a robust definition of accented speech, there is a general consensus that accents are comprised of a wide variety of speech features such as shifts in both the boundaries and categories of multiple vowels and consonants, differences in prosodic features such as pitch and volume, and, in the case of unguided speech, vocabulary and grammatical signifiers. These specific changes may occur in varying combinations and degrees, depending on the particular speaker and listener. While this degree of variability is acceptable, or even desirable in some research contexts, it conflicts with the current goal of investigating the effect of phonemic degradation separately from accent.

There are three types of alterations which could reasonably be made to create a synthetic accented speech analogue: phonemic alteration, phonetic alteration, and prosodic alteration. Though there are also specific grammatical markers of non-native speech at both the morphological and syntactic level (Franceschina, 2001), this type of alteration was rejected as it would likely cause interference with the semantic context of a given sentence which would make it difficult to draw conclusions about the relative role of semantics when adapting to idiosyncratic speech.

I chose to work with phonemic alterations, mainly due to the fact that I wished to create a large enough difference between the incoming speech and the listeners' expected input that they would be likely to make a mistake in what word was spoken. In this scenario where only a single change is being made to the speech, phonemic and prosodic alterations were judged to be too subtle to produce a noticeable degree of uncertainty over what was being said. The specific phonemic alterations chosen were simple phoneme swaps:/l/ ➔ /r/, /w/ ➔ /v/, /v/ ➔ /b/, /θ/ ➔

/t/, and /ð/ ➔ /d/. These particular swaps were selected as they are analogous to existing real-world accents, albeit highly simplified. The /l/ ➔ /r/ and /v/ ➔ /b/ swaps, for instance, can be seen in native speakers of Japanese, while the /w/ ➔ /v/ can be seen in native German speakers. In order to test for the effects of varying levels of context, the artificial accent needed to be placed in a full sentence context, with the sentences themselves offering varying degrees of semantically useful information to the listener. It was also important that the sentences themselves contained at least a few instances of the altered phoneme so that listeners would have the opportunity to adapt to the idiosyncratic speech. The final criteria was that any instances of a changed phoneme should result in a non-word or, where creating a real word could not be reasonably avoided, the resulting word should be relatively uncommon and inappropriate for the given sentence in which it appears. This was to ensure that listeners would be aware that a phoneme change had taken place, rather than simply thinking an entire word had changed. With this criteria in mind, I generated novel sentences which will be discussed in the following method section. The sentences were initially written with deliberately varying levels of contextual constraint for the final target word of each sentence based on the author's intuition. To determine the degree of restraint empirically, however, a brief norming experiment was conducted, which will be explained in the following section.

Method

Materials.

A set of target sentences were created to satisfy the various needs of the experiment. A total of 12 sentences were generated for each of the five phoneme swaps (/l/ ➔ /r/, /w/ ➔ /v/, /v/ ➔ /b/, /θ/ ➔ /t/, and /ð/ ➔ /d/), for a total of 60 sentences. In each sentence, the target word was the final word of the sentence and was two syllables in length; it contained one instance of the swapped phoneme for that particular sentence. Further, each sentence contained at least three prior instances of the swapped phoneme to provide listeners with prior phonemic information on the swap which would occur in the target word. Table 1 shows five example sentences, one for each type of phoneme swap, along with the specific phoneme being swapped. A table of the full set of sentences can be found in Appendix A.

Table 1

Examples of Full Sentences, Respective Target Words, and Phoneme Swaps

Target Word	Phoneme Swap	Sentence (Unaccented)	Sentence (Accented)
planted	/l/ ➔ /r/	*Joe had decided to take a new hobby for the summer since he had lots of space in his yard, and his first step was to plan out where everything would be planted*	*Joe had decided to take up a new hobby for the summer since he had rots of space in his yard, and his first step was to pran out where everything would be praanted*
whistle	/w/ ➔ /v/	*The wind was getting very cold and the hiker knew she was lost, so she unpacked her survival kit and blew as hard as she could on the whistle*	*The vinnd vuz getting very cold and the hiker knew she vuz lost, so she unpacked her survival kit and blew as hard as she could on the vistle*
victor	/v/ ➔ /b/	*Although both fighters were evenly matched, by the end of the fight it was clear who would be declared the victor*	*Although both fighters were eebenly matched, by the end ob the fight it was clear who would be declared the bictor*
thorough	/θ/ ➔ /t/	*With her boss paying close attention to the project, everything needed to go smoothly, so Mandy made sure all the initial checks were totally thorough*	*Wit her boss paying close attention to the project, everyting needed to go smoothly, so Mandy made sure all the initial checks were totally torough*
either	/ð/ ➔ /d/	*Rachel looked to her friends for advice, even though it was clear that they didn't know what to do either*	*Rachel looked to her friends for advice, even dough it was clear dat dey didn't know what to do eider*

Note. The spelling changes shown in the accented version of each sentence have been underlined and reflect the spellings which were used to generate the synthetic speech stimuli in the second experiment. Phoneme swaps are shown in IPA notation.

Participants.

Participants for the norming portion were 84 University of Alberta undergraduate students enrolled in an introductory psychology course. Demographic information such as age, gender identity, and ethnicity were not collected. Although participants were not initially screened for language background, a language background survey was completed at the end of the experiment for the purposes of analysis. Participants received course credit for their participation, regardless of whether or not they completed the task.

Procedure.

The experiment was conducted using the online experiment design software Gorilla (Anwyl-Irvine et al., 2020). As a result, the exact hardware which participants used to complete the task is unknown, although the experiment settings were restricted to completion on a desktop computer or laptop. The experiment began with a brief instruction screen which read:

For this experiment, you will be shown several sentences with the final word missing. We ask that you read each sentence carefully and enter what you think the final word should be in the text box below the sentence. Please limit your answers to a SINGLE WORD. Click the button below to continue to the experiment.

After clicking the “continue,” participants completed a cloze task in which they were shown each sentence with a blank line in place of the final word. A text entry field was located below each sentence. Participants could not continue to the next sentence until they made an entry in the text field. Participants were given no feedback on their responses and could not go back to review previous sentences or change their responses. Sentences were presented in a random order to each participant.

After completing the final sentence, participants were given a brief language background survey in which they were asked to give their first written and spoken languages, separated by commas if listing more than one.

Analysis.

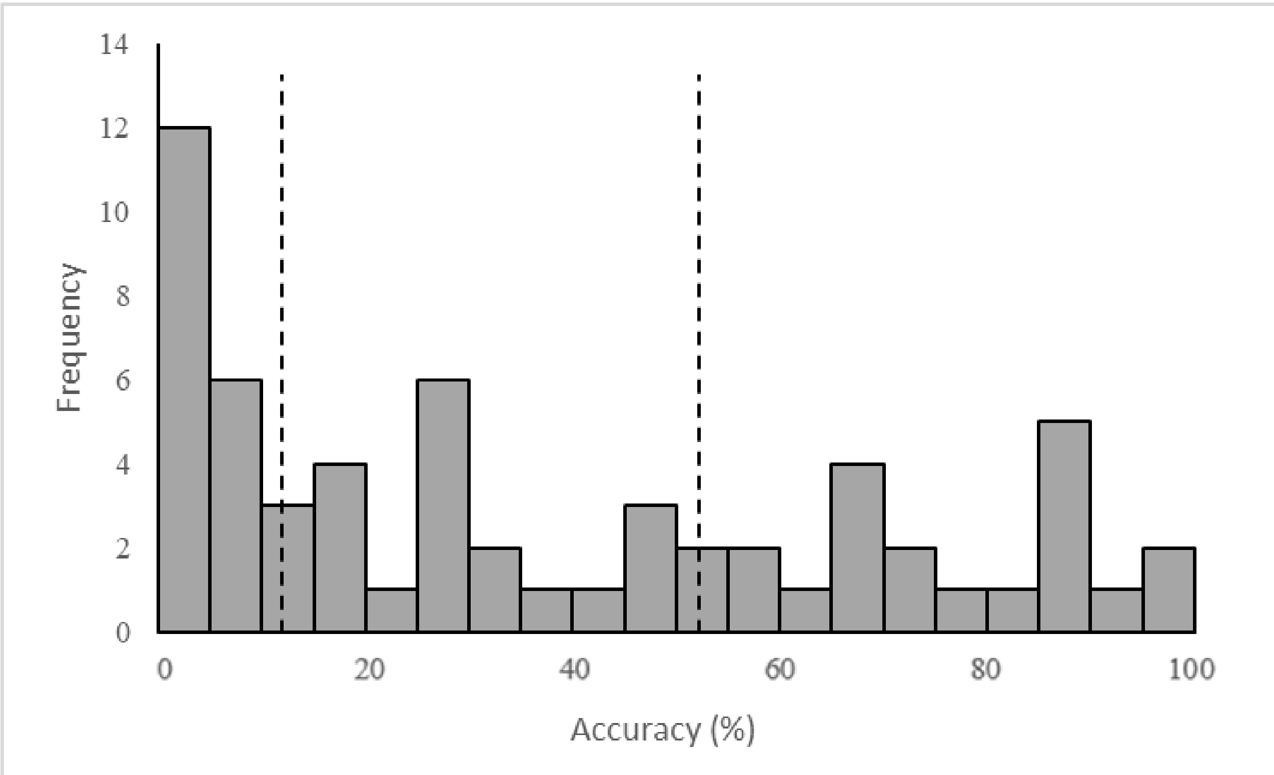
Of the 84 initial participants, 29 did not include English as either their first spoken or written language and were excluded from the analysis, leaving a total of 55 participants. Responses with an exact match to the target word (disregarding letter case) were considered correct. Each target word was given an accuracy score based on the proportion of correct responses.

Results

As shown in Figure 1, scores for target words had a considerable range, with the percent accuracy ranging from 0 to 99 percent (μ = 37.1, *SD* = 31.5), though the distribution was noticeably skewed to the right, with one third of results falling below 11.4% accuracy, one third falling between 11.4% and 52.2% accuracy, and the final third falling above 52.2% accuracy. Target words were placed into three contextual constraint quantiles based on their accuracy scores, yielding 20 low-constraint sentences, 20 medium-constraint sentences, and 20 high-constraint sentences.

Figure 1

Histogram of Target Word Accuracies and Quantile Cutoffs



Note. Dotted lines indicate the quantile threshold values of 11.4% and 52.2%

Judgements of Accented Speech

The primary goal of the second experiment was to investigate the interactions between accent, noise, context, and time. To that end, participants needed to be exposed to varying combinations of accent, noise, and context simultaneously, meaning these three factors needed to be represented within a single sentence. Noise masking was used to reduce phonemic information. While noise masking is an indirect path to reducing the amount of phonetic information in a sentence, past research has found that presenting speech in noise does impact how listeners process speech at a phonemic level (Winn et al., 2013), and the overall effects of noise on speech recognition have been widely studied (see Mattys et al., 2013 for a review). Listeners' perception of these stimuli was assessed with a simple comprehension task in which participants were asked to transcribe the final word of a spoken sentence.

Method

Materials.

For the main comprehension experiment, spoken versions of each sentence were required. The MacOS text-to-speech synthesizer with the "Samantha" voice was used to create spoken versions of each of the target sentences. For this experiment, both unaccented and accented versions of each sentence were required. The unaccented versions of each sentence were simply feeding the unaltered sentences into the speech synthesizer. The accented versions were created using a trial-and-error process, in which the spellings of affected words were altered to produce the desired change. The output of the synthesizer was checked and further alterations were made to the spelling if the pronunciation was incorrect.

Once the spoken versions of the sentences had been generated, noise versions of each accented and unaccented sentence were created using the open source audio editing software

Audacity. Using Audacity's built-in white-noise function, each spoken sentence was layered with white-noise increasing amplitudes. While the exact volume settings of a given participant could not be controlled, the relative relationship between the loudness of each component of the listening experiment was measured using a directional decibel meter placed approximately 20cm away from a laptop speaker. In these tests, the calibration tone had an average peak decibel level of 40dB, the sentences with no noise had an average peak decibel level of 82.9 dB, and each level of noise in isolation had an average peak decibel level of 77.3 dB, 82.7 dB, 84.9 dB, 88.9 dB, and 90.8 dB. The combination of noise level and accent resulted in a total of 12 versions of each sentence: 2 (accent condition) x 6 (no noise plus five increasing levels of noise).

Additionally, an error in the initial construction of the sentences led to one of the target words in the high-constraint condition not meeting the necessary phoneme criteria (specifically, the sentence was intended to have a target word with a /l/ ➔ /r/ swap, but instead the target word involved a /r/ ➔ /l/ swap). This sentence was excluded from the comprehension portion of the experiment, leaving a total of 59 sentences; 20 low constraint, 20 medium constraint, and 19 high constraint.

In addition to the sentences described above, an additional four practice sentences were created to introduce participants to the sort of stimuli they would be encountering during the main experiment. These sentences included one sentence with no accent and no noise, one sentence with an accent and no noise, one sentence with no accent and noise, and one sentence with both accent and noise. The noise level used in the practice sentences corresponds to noise level 2 used in the experimental stimuli. A full list of the practice sentences can be found in Appendix A. The phoneme swap used for the practice sentences was a */f/* ➔ */p/* swap; this was unique to the practice sentences and did not appear in the experimental portion.

The experiment was conducted using the online experiment design software Gorilla. As a result, the exact hardware which participants used to complete the task is unknown, although the experiment settings were restricted to completion on a desktop computer or laptop.

Participants.

Participants for the comprehension portion were University of Alberta undergraduate students enrolled in an introductory psychology course. Ninety-six students participated in the experiment. Demographic information such as age, gender identity, and ethnicity were not collected. Although participants were not initially screened for language background, a language background survey was completed at the beginning of the experiment for the purposes of analysis. Participants received course credit for their participation, regardless of whether or not they completed the task.

Design.

The comprehension portion followed a within-subjects design, in which each participant was exposed to both accent conditions, all noise levels, and all contextual constraint levels. Participants would hear only one version of each sentence. For example, if they heard the unaccented version of a sentence with a noise level of 0.5, they would not hear any other versions of that particular sentence. This necessitated 12 counterbalanced versions to allow all versions of each sentence to appear once.

Procedure.

Participants began by completing a brief language background survey in which they were asked to enter their first written and spoken languages in a text entry field. Following the language background survey, participants completed a calibration and practice segment. Before beginning the calibration, participants were given the following instructions:

> *For this experiment you will be listening to audio recordings of sentences. If possible, please use headphones or earbuds and complete the experiment in a quiet room. The experiment will begin with a few practice sentences where you will be given feedback on your answers. After this practice segment, you can continue to the experimental portion where you will not receive feedback. When you are ready, click the button below to continue.*

After clicking continue, participants were played a 500 Hz tone and asked to adjust the volume of their audio device until the tone was "barely audible." As participants were completing the task on their own computers, this was a necessary step to ensure that regardless of what audio equipment a participant was using, they would experience approximately the same perceived loudness. Once participants had adjusted their volume, they began a practice segment in which they heard the four practice sentences described in the materials section and were asked to write what they thought the final word was. In the practice section, participants were given feedback in the form of being shown the correct answer after their response, regardless of whether their response was correct or not.

Once participants had completed the practice segment they were shown a final instruction screen which read:

> *During the experimental portion, you will only hear each sentence once and you will not receive feedback on your answers. If you are not sure what the final word was, please enter your best guess. Click the button below to begin.*

Participants were assigned to one of the 12 counterbalance conditions such that each combination of accent condition and noise level for each sentence was seen by an approximately equal number of participants. Once participants continued from the instruction screen, they were brought to a screen with a button labeled “play” which allowed them to hear the sentence when they were ready. After hearing the sentence, they were automatically brought to a text entry field and a prompt asking them what they thought the final word was. Participants were not able to listen to the sentence a second time and could not go back to change their answers. Additionally, participants could not advance to the next sentence until they had given a response. This was to ensure that participants did not simply skip over sentences which they found too difficult to understand. This process repeated, with sentences were presented to participants in a random order until they had responded to the full set, at which point they were taken to a debriefing screen.

Analysis.

Of the 96 participants who completed the task, 20 did not have English as either their first written or spoken language and were removed from the data set, leaving a total of 76 participants in the final analysis. The overall score for each participant was then calculated, and one participant with a score of less than 30 percent correct was removed under the assumption that they did not understand or were not properly attending to the task. Scoring was binary, with an exact match between target and response (disregarding letter case) considered correct and any other response considered incorrect.

The resulting data were analysed using R Statistical Software (Version 4.3.1, R Core Team, 2023) and the R packages, lme4 (Version 1.1-33, Bates et al., 2023) and emmeans (Version 1.8.7, Lenth, 2023). Because the accuracy data were binomial, the results were fit with a generalized linear model with a logistic link function (cf. Dixon, 2008). The outcome variable was a simple binary score based on whether the response exactly matched the target word, with contextual constraint (low, medium, high), noise category (no-noise baseline, and levels one through five), accent condition (accented or unaccented), and time (first quarter, second quarter, third quarter, and fourth quarter) as factors. Participant and stimulus were included as random intercepts in the models. To analyze the effect of noise, two orthogonal contrasts were created: one contrast separated the baseline condition from trials in which any amount of noise was present and a second contrast treated noise levels one through five as a linear effect. The rationale for these contrasts was that the exact amount of ambient noise in each participant's environment was variable and unknown, so an exact signal-to-noise ratio could not be determined for the baseline condition. However, the effect of added noise should systematically increase from level 1 to 5. Both contrasts were included to represent the effects of noise as a whole.

To investigate the main effects and interactions of each of the factors, nested model were compared with (adjusted) likelihood ratios. (For a review of likelihood ratios and their application, see Glover & Dixon, 2004.) Likelihood ratios were adjusted for the varying number of model parameters based on the Akaike Information Criterion (AIC); thus, a comparison of adjusted likelihood ratios is equivalent to comparing model AIC values. For comparison, in some prototypical hypothesis testing situations, a significant effect would correspond to an adjusted likelihood ratio of about 3. An initial null model was created which included no fixed predictors, but participant and item were included as random factors. The main effect of each factor was added to the model in sequence and compared to the previous iteration of the model using likelihood ratios. Following the addition of main effects, two-way interactions were added to the model in sequence and tested in the same manner. Finally, a full model which included all main effects and interactions was generated and compared to previous models. To generate descriptive statistics, the estimated marginal means of the logits were estimated from the full model using the emmeans function within R.

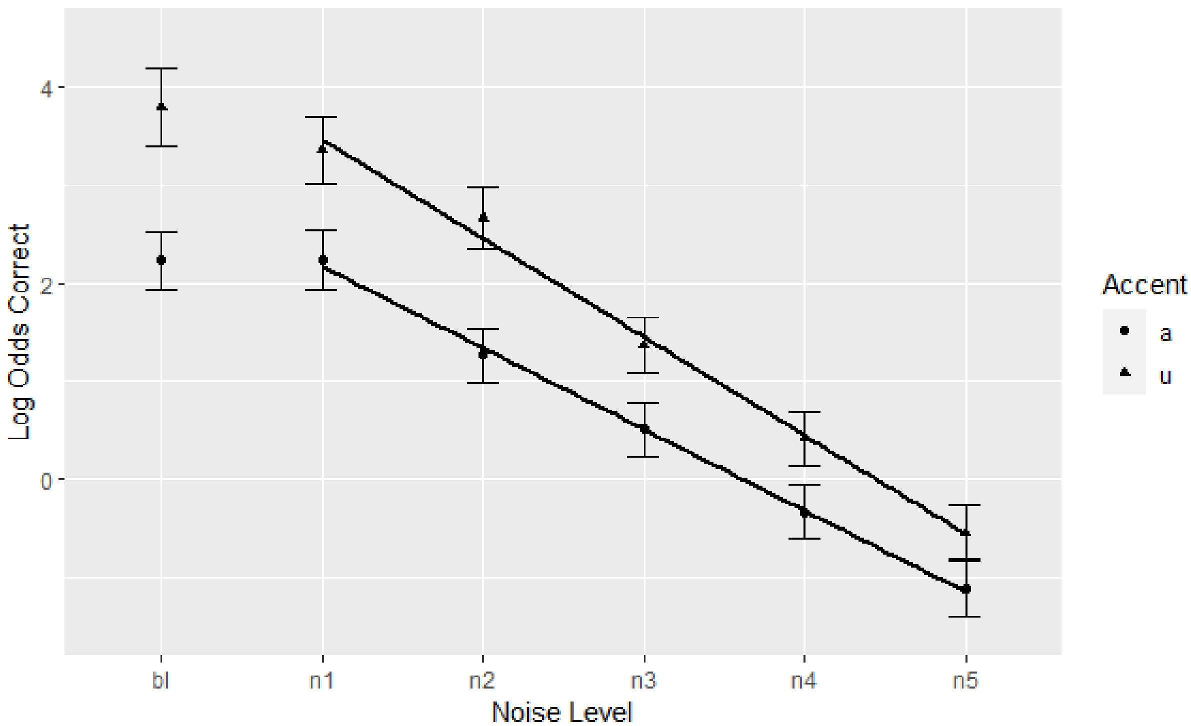
Results

The analysis showed evidence of main effects of context, noise, and accent, and no effect of time. Additionally, there were notable interactions between accent and noise (fig. 2), as well as accent and context (fig. 3). Specifically, at higher noise levels and with less contextual constraint, accent had less of an effect on performance.

Beginning with the null model, main effects for each factor were added to the model in sequence and compared to the previous iteration of the model. This revealed minimal evidence for an effect of time (λ_{adj} = 1.51), strong evidence for an effect of noise (λ_{adj} > 1000), strong evidence for an effect of accent (λ_{adj} > 1000), and moderate evidence for an effect of context (λ_{adj} = 6.46). Following these results, time was excluded from the remainder of the model comparisons. Two-way interactions were then tested against a model which included main effects of noise, context, and accent. Once again, interactions were added in sequence and tested against the previous iteration, beginning with the interaction between context and noise. This showed strong evidence against an interaction between noise and context (λ_{adj} = 0.027), moderate evidence for an interaction between noise and accent (λ_{adj} = 16.50), and moderate evidence in favour of an interaction between context and accent (λ_{adj} = 5.97). Finally, a comparison of a model which included both accent by noise and accent by context interactions to a full model which included all two-way interactions as well as the three-way accent by noise by context interaction showed strong evidence against the full model (λ_{adj} = 0.01). A table of the cell means for the full model can be seen in Table 2, while the means for the effect of time can be seen in Table 3.

Figure 2

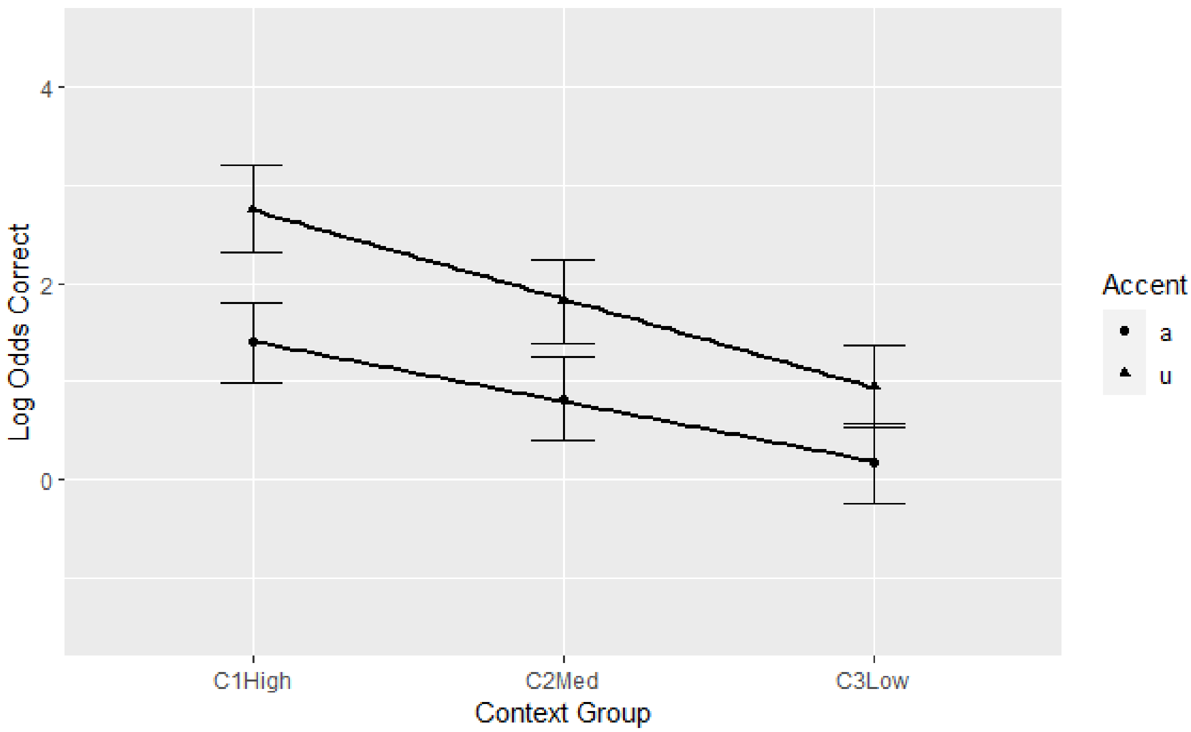
Histogram of Target Word Accuracies and Quantile Cutoffs



Note. The trend line excludes the baseline noise conditions as the model was fit by treating the baseline condition separately from the conditions in which noise was added artificially. Error bars show one standard error.

Figure 3

Log Odds Correct as a Function of Accent and Context Conditions



Note. Error bars show one standard error.

Table 2

Cell Means for Each Combination of Accent, Context Condition, and Noise Condition

Accent Condition	Context Condition	Noise Condition	Mean (Log Odds Correct)	SE
Unaccented	High	Baseline	4.985	0.839
		Noise 1	4.409	0.661
		Noise 2	3.684	0.548
		Noise 3	2.306	0.481
		Noise 4	1.014	0.458
		Noise 5	0.159	0.450
	Medium	Baseline	3.617	0.607
		Noise 1	2.983	0.545
		Noise 2	2.675	0.533
		Noise 3	1.351	0.481
		Noise 4	0.569	0.465
		Noise 5	-0.283	0.460
	Low	Baseline	2.768	0.519
		Noise 1	2.662	0.517
		Noise 2	1.641	0.481
		Noise 3	0.452	0.464
		Noise 4	-0.325	0.455
		Noise 5	-1.499	0.474
Accented	High	Baseline	3.156	0.511
		Noise 1	3.016	0.526
		Noise 2	1.729	0.466
		Noise 3	0.786	0.454
		Noise 4	0.275	0.454
		Noise 5	-0.577	0.464
	Medium	Baseline	1.931	0.502
		Noise 1	2.445	0.538
		Noise 2	1.168	0.477
		Noise 3	0.636	0.471
		Noise 4	-0.402	0.468
		Noise 5	-0.811	0.462
	Low	Baseline	1.617	0.475
		Noise 1	1.241	0.460
		Noise 2	0.900	0.460
		Noise 3	0.102	0.454
		Noise 4	-0.875	0.455
		Noise 5	-1.937	0.472

Table 3

Marginal Means for the Main Effect of Time

Quartile	Mean (Log Odds Correct)	SE
First	0.892	0.204
Second	0.904	0.203
Third	0.946	0.203
Fourth	1.123	0.204

Discussion

The results of the analysis support a model in which accent interacts with both semantic context and noise, while context and noise do not interact with one another. In particular, the presence of an accent has a stronger detrimental effect at lower noise levels and more constrained contexts, that is, the “easier” conditions. The interaction between noise and accent is interesting in that it seems to conflict with some previous findings such as those of Munroe (1998) in which the addition of noise to a transcription task had a much more detrimental effect on the transcription of foreign-accented speech compared to native English speech. It should be noted that those results were based on percent accuracy rather than, for instance, log odds correct, and ceiling effects may have made this interaction more pronounced. Munroe also noted that the effect varied greatly between both speakers and transcribers and that making a general statement on the relationship between noise and accented speech comprehension would be inappropriate. The different pattern of results in the present study may be explained in part by the intelligibility of the stimuli, as the accent condition involved only a single alteration in pronunciation for a given sentence, in comparison to the naturally accented speech used by Munroe (1998). At high noise levels, such a minor difference may become irrelevant to comprehension in the face of highly adverse listening conditions. Alternatively, the presence of noise may mask the altered phoneme entirely such that a listener would not perceive any noticeable difference between the unaccented and accented sentences. Regardless of the precise mechanism, the potential interaction between noise and intelligibility should be explored further.

Although it was not the main intent of their study, Strori et al. (2020) investigated the relationship between intelligibility, noise level, and sentence complexity. While their analysis did not show an interaction between intelligibility and noise level, this may be due to the smaller range of noise used in their experiments (a minimum signal-to-noise ratio of -4 dB and a maximum of 0 dB, in contrast to a minimum signal-to-noise ratio of approximately -7.0 dB and a maximum of approximately 5.6 dB in the present study). The interaction seen in the present study may only be evident across a wider range of noise than the range explored by Strori et al. (2020).

One final possible explanation for the interaction between noise an accent is effort and attention. The load theory of attention proposes that in ideal listening conditions, attentional resources are automatically assigned to processing and overcoming distractors, while in adverse conditions these resources are instead allocated to overcoming the noise present within the environment, and the impact of distractors becomes less apparent (Gustafson et al., 2023). In the context of the present study, the differences caused by the accented words may act in a similar manner to a distractor in a listening task, causing listeners to not properly attend to the final word. When large amounts of noise are added, attentional resources are instead devoted to overcoming the noise of the environment, and the accented words become less distracting (and less of an impact on performance) as a result.

A similar pattern of results can be seen in the interaction between semantic context and accent, with accent having a smaller effect in sentences with lower levels of semantic constraint. Once again, these results seem to contradict previous findings which suggest that listeners utilize semantic information to a greater degree when processing accented speech compared to native speech (Sajin & Connine, 2017). Unlike the previously discussed results from Munroe (1998), these results were derived from a generalized linear model which eliminated ceiling or floor effects as a possible explanation. The current pattern of results seems to indicate that when processing accented speech, semantic information is instead less beneficial to recognition than it is when processing unaccented speech. One possible explanation for this finding is similar to that which was suggested to explain the interaction between accent and noise: In the presence of highly adverse conditions (in this case highly open-ended sentences) the relatively minor difficulty imposed by the artificial accent is overwhelmed by the difficulty imposed by the lack of useful semantic information. Another possible factor is the prediction process which listeners employ when listening to highly constrained sentences. When predicting the final word of constrained sentences, listeners show a difference in evoked potentials when listening to a native speaker compared to a foreign-accented speaker (Romero-Rivas et al., 2016). Simply acknowledging that a difference in processing exists, however, is not a suitable basis for predicting the pattern of results found in the current experiment, particularly when such a difference has only been shown in the context of highly constrained sentences. Indeed, it is difficult to speculate on the interaction between semantic context and accent further, given the relatively sparse literature on this particular intersection. Further investigation into this interaction in particular would be a good avenue for future research.

A much more well-explored phenomenon is the adaptation to accent over time. Generally one would expect that in an experiment involving accented stimuli, participants' accuracy would improve over time. The analysis did not find evidence for a main effect of time, nor was there evidence for an interaction between accent condition and time or noise level and time. Although this finding conflicts with the previous literature, this result is somewhat unsurprising given the randomized between-subjects design. As participants were subjected to varying degrees of noise, contextual constraint, and accented or unaccented stimuli, the most likely reason for the lack of adaptation is that they were not given consistent enough stimuli in order to fully grasp the pattern of the accented sentences. This possibility was considered during the initial design stages, but it was thought that the simple nature of the accents (swapping only a single phoneme in a given sentence) would allow participants to adapt quickly. The opposite is also a possibility, that is, the artificial accents were simple enough that participants reached their maximum degree of adaptation almost immediately. Although participants were universally less accurate on accented sentences, this does not necessarily mean that they did not immediately adapt to the artificial accents immediately, as even after a period of adaptation participants in previous studies have still shown a lower accuracy and increased processing time compared to unaccented stimuli (Clarke & Garrett, 2004). Regardless of whether participants adapted too quickly, or did not adapt at all, the lack of any main effects of, or interactions with time should not be taken as evidence against previous findings.

While the initial goal of this thesis was to investigate the adaptation process specifically, the lack of a meaningful change in accuracy over time makes it inappropriate to make general statements on this process. However, there are still several observations that can be drawn from the results which are relevant to the models discussed in the introduction. One such result is the evidence in favour of a model which includes an interaction between noise and accent, as well as semantic constraint and accent, with no interaction between noise and semantic constraint. While the interactions of noise and accent, and semantic constraint and accent have already been discussed, the lack of an interaction between noise and semantic constraint is worth noting. The lack of interaction would suggest that these two factors play similar but independent roles in speech recognition, likely relatively early in the recognition process given that both interact with accent in a similar manner. One model which is particularly relevant here is that of the logogen model (Morton, 1969), in which all relevant semantic, visual, and acoustic information feeds into a single node, termed the logogen, associated with a particular word. Notably, the logogen model predicts additive, rather than multiplicative, effects between semantic and acoustic information, a prediction which aligns with the findings of the current study. Additionally, the model incorporates acoustic and semantic information at the same point early in processing, which is compatible with the current finding of noise and context interacting with accent in a similar manner. Being an early model of word-recognition, the logogen model does not satisfy all of the criteria previously discussed, in particular, the role of accent in speech recognition and adaptation over time. Regardless, it offers a simple, if incomplete, account of the relationship between noise and context and may be worth re-examining more deeply in the context of current models of speech perception.

The early influence is unsurprising with regards to noise, as both accent and noise can be thought of as occurring at the very start of the recognition process by affecting the perceived waveform of the speech. An early influence of semantic information is similar to the active processing framework proposed by Heald and Nusbaum (2014), with semantic information in particular acting in a top down manner to influence early perceptual input. It bears repeating, however, that the observed interaction between semantic constraint and accent is not well-understood, and should be investigated further.

Conclusion

This thesis set out initially to investigate the effects of semantic context and noise on process of accented speech adaptation. While the main goal was not achieved due to a lack of evidence for adaptation over time, the novel approach of investigating accent, noise, and semantic information simultaneously has led to several interesting results. Most notably, increasingly adverse conditions, both noise and low semantic constraint, led to a reduction in the detrimental effect caused by artificially accented speech, contrary to the general findings of the literature. The complex design of this experiment makes it difficult to determine the exact nature of these interactions, and there are several possible explanations (intelligibility, presence of noise and semantic constraint simultaneously, the degree of the adverse conditions) which warrant their own studies. Further, the results of the analysis suggest a model which incorporates semantic context as a main factor influencing speech perception. The results also suggest that such a model would involve contextual information at an early stage of perception which simultaneously allows it to interact with phonetic idiosyncrasies such as accent while acting independently of other sources of noise.

References

Baese-Berk, M. M., McLaughlin, D. J., & McGowan, K. B. (2020). Perception of non-native speech. *Language and Linguistics Compass*, *14*(7), e12375.

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

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America*, *106*(4), 2074-2085.

Brown, V. A., McLaughlin, D. J., Strand, J. F., & Van Engen, K. J. (2020). Rapid adaptation to fully intelligible nonnative-accented speech reduces listening effort. *Quarterly Journal of Experimental Psychology, 73*(9), 1431-1443.

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review, 119*(2), 417.

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

Creel, S. C., & Bregman, M. R. (2011). How talker identity relates to language processing. *Language and Linguistics Compass*, *5*(5), 190-204.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology, 55*, 149-179.

Elman, J. L., & McClelland, J. L. (1986). Exploiting Lawful Variability In the Speech Wave. In invariance and Variability in Speech Processes (pp. 360-385). Psychology Press.

Franceschina, F. (2001). Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals. *Second Language Research*, *17*(3), 213-247.

Gao, X., Yan, T. T., Tang, D. L., Huang, T., Shu, H., Nan, Y., & Zhang, Y. X. (2019). What makes lexical tone special: a reverse accessing model for tonal speech perception. *Frontiers in Psychology, 10*, 2830.

Glover, S., & Dixon, P. (2004). Likelihood ratios: A simple and flexible statistic for empirical psychologists. *Psychonomic Bulletin & Review, 11*(5), 791-806.

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

Gustafson, S. J., Nelson, L., & Silcox, J. W. (2023). Effect of auditory distractors on speech recognition and listening effort. *Ear and Hearing*, 10-1097.

Hau, J. A., Holt, C. M., Finch, S., & Dowell, R. C. (2020). The adaptation to Mandarin-accented English by older, hearing-impaired listeners following brief exposure to the accent. *Journal of Speech, Language, and Hearing Research*, *63*(3), 858-871.

Heald, S. L., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience, 8*, 35.

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

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*(1), 1-15.

Levenshtein, V. I. (1966, February). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet Physics Doklady* (Vol. 10, No. 8, pp. 707-710).

Luthra, S., Li, M. Y., You, H., Brodbeck, C., & Magnuson, J. S. (2021). Does signal reduction imply predictive coding in models of spoken word recognition?. *Psychonomic Bulletin & Review, 28*, 1381-1389.

Massaro, D. W., & Cohen, M. M. (1987). Process and connectionist models of pattern recognition. In *Proceedings of the Ninth Annual Conference of the Cognitive Science Society* (pp. 258-264). Erlbaum Hillsdale, NJ.

Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology, 21*(3), 398-421.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2013). Speech recognition in adverse conditions: A review. *Speech Recognition in Adverse Conditions*, 1-26.

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

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746-748.

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.

Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition*, *9*(3), 225-236.

Paquette-Smith, M., Cooper, A., & Johnson, E. K. (2021). Targeted adaptation in infants following live exposure to an accented talker. *Journal of Child Language*, *48*(2), 325-349.

Romero-Rivas, C., Martin, C. D., & Costa, A. (2016). Foreign-accented speech modulates linguistic anticipatory processes. *Neuropsychologia*, *85*, 245-255.

Sadeghi, M., Marvi, H., & Ali, M. (2018). The effect of different acoustic noise on speech signal formant frequency location. *International Journal of Speech Technology*, *21*, 741-752.

Sajin, S. M., & Connine, C. M. (2017). The influence of speech rate and accent on access and use of semantic information. *Quarterly Journal of Experimental Psychology*, *70*(4), 619-636.

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

Strori, D., Bradlow, A. R., & Souza, P. E. (2020). Recognition of foreign-accented speech in noise: The interplay between talker intelligibility and linguistic structure. *The Journal of the Acoustical Society of America*, *147*(6), 3765-3782.

Swinney, D. A. (1979). Lexical access during sentence comprehension:(Re) consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*(6), 645-659.

Trude, A. M., & Brown-Schmidt, S. (2013). Talker-specific perceptual adaptation during online speech perception. In *Speech Recognition in Adverse Conditions* (pp. 27-49).

Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2013). Roles of voice onset time and F0 in stop consonant voicing perception: Effects of masking noise and low-pass filtering. Psychology Press. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1097–1107.

Xie, X., Jaeger, T. F., & Kurumada, C. (2023). What we do (not) know about the mechanisms underlying adaptive speech perception: A computational framework and review. *Cortex*, *166,* 377-424.

Xie, X., & Myers, E. B. (2017). Learning a talker or learning an accent: Acoustic similarity constrains generalization of foreign accent adaptation to new talkers. *Journal of Memory and Language*, *97*, 30-46.

Appendix A

Table A1. Accented and unaccented sentences

Unaccented Sentence	Accented Sentence	Target Word	Phoneme Swap
Since meteorologists were predicting a short summer, the farmer wanted to plant his crop as soon as possible, meaning the field would need to be ploughed	*Since meteorawrogists were predicting a short summer, the farmer wanted to prant his crop as soon as possibur, meaning the fierd would need to be proud*	*ploughed*	*/l/* ➔ */r/*
Joe had decided to take a new hobby for the summer since he had lots of space in his yard, and his first step was to plan out where eveything would be planted	*Joe had decided to take up a new hobby for the summer since he had rots of space in his yard, and his first step was to pran out where everything would be praanted*	*planted*	*/l/* ➔ */r/*
Finally, after four years of dealing with budget constraints and other problems, the construction project was complete	*Finery, after four years of dearing with budget constraints and other prawbrems, the construction project was comprete*	*complete*	*/l/* ➔ */r/*
After she broke her leg at the job site, the first thing Kate did after going to the emergency clininc was to call a lawyer	*After she broke her reg at the job site, the first thing Kate did after going to the emergency crinic was to call a roier*	*lawyer*	*/l/* ➔ */r/*
The valley was lush and full of plant life, and Erin decided to pick some flowers	*The varrey was rush and fur of prant rife, and Tom decided to pick some frowers*	*flowers*	*/l/* ➔ */r/*

Unaccented Sentence	Accented Sentence	Target Word	Phoneme Swap
Mike and Sally had extremely strict parents, and playing inside the house was not allowed	*Mike and Sarry had extreemery strict parents, and praying in the house was not a roud*	*allowed*	*/l/ ➔ /r/*
Ben liked going to buffets because it let him decide exactly how much food he wanted on his plate	*Ben riked going to buffays because it ret him decide exactry how much food he wanted on his prate*	*plate*	*/l/ ➔ /r/*
The painters first needed to lay out some drop cloths to keep things clean before the first coat of paint could be applied	*The painters first needed to ray out some drop croths to keep things creen before the first coat of paint could be a pried*	*applied*	*/l/ ➔ /r/*
Eventually after walking for a long time, the hikers stumbled upon a quiet clearing	*Evenchooary after walking for a rong time, the hikers stumberd upon a quiet crearing*	*clearing*	*/l/ ➔ /r/*
Even though the seaside village was very peaceful, it was often foggy and overcast so the atmosphere was quite gloomy	*Even though the seaside virrage was very piecefur, it was often foggy and overcast so the atmosphere was quite groomy*	*gloomy*	*/l/ ➔ /r/*
It was the perfect night for telling scary stories because the rain was falling and the wind was blowing	*It was the perfect night for terring scary stories because the rain was farring and the wind was broeing*	*blowing*	*/l/ ➔ /r/*

Unaccented Sentence	Accented Sentence	Target Word	Phoneme Swap
The wind was getting very cold and the hiker knew she was lost, so she unpacked her survival kit and blew as hard as she could on the whistle	*The vinnd vuz getting very cold and the hiker knew she vuz lost, so she unpacked her survival kit and blew as hard as she could on the vistle*	*whistle*	/w/ ➔ /v/
The evening was going well and the diners decided that they wanted some more wine, so they flagged down the waiter	*The evening vuz going vell and the diners decided that they vawnted some more vine, so they flagged down the vayter*	*waiter*	/w/ ➔ /v/
The meteorologists predicted a relatively mild season ahead, but the hunter wanted to be extra cautious and prepped his cabin to withstand a hard winter	*The meteorologists predicted a relatively mild season ahead, but the hunter vawnted to be extra careful and prepped his cabin to vith stand a hard vinter*	*winter*	/w/ ➔ /v/
After arriving at the cabin for the weekend, the campers decided that it was the perfect day to go swimming	*After arriving at the cabin for the veek end, the campers decided that it vuz the perfect day to go svimming*	*swimming*	/w/ ➔ /v/
Everyone around town was preparing for the worst, since the news said they were in for some extremely rough weather	*Every vun around town vuz preparing for the vurst, since the news said they vur in for some extremely rough veather*	*weatherw*	/w/ ➔ /v/
Jess was shocked when she went down to her basement to find it covered in water	*Jess vuz shocked ven she vent down to her basement to find it covered in vawter*	*water*	/w/ ➔ /v/

Unaccented Sentence	Accented Sentence	Target Word	Phoneme Swap
Kate screamed and quickly got out of her car when she saw the wasp	*Kate screamed and kvickly got out of her car ven she saw the vawsp*	*wasp*	/w/ ➔ /v/
Mary decided to go out with friends for the night, but first she had to decide what to wear	*Mary decided to go out vith friends for the night, but first she had to decide vut to vare*	*wear*	/w/ ➔ /v/
The air outside was wonderful and fresh, so the couple decided to go for a walk and try some bird watching	*The air outside vuz vonderful and fresh, so the couple decided to go for a vawk and try some bird vawtching*	*watching*	/w/ ➔ /v/
The man worked hard seven days a week for years to learn how to become a weaver	*The man vorked hard seven days a veek for years to learn how to become a veever*	*weaver*	/w/ ➔ /v/
Instead of planning a route like she usually would, the hiker instead packed a compass and map and went out to the woods to wander	*Instead of planning a route like she usually vood, the hiker instead packed a compass and map and vent out to the voods to vander*	*wander*	/w/ ➔ /v/
Because of the pointy hat, the robes, and the wand, it was very clear that the old man was a wizard	*Because of the pointy hat, the robes, and the vawnd, it vuz very clear that the old man vuz a vizard*	*wizard*	/w/ ➔ /v/

Unaccented Sentence	Accented Sentence	Target Word	Phoneme Swap
Mark decided he should visit an optometrist as he had noticed a very slight blurriness in his vision	*Mark decided he should bisit an optometrist as he had noticed a bery slight bluriness in his bision*	*vision*	*/v/* ➔ */b/*
Although both fighters were evenly matched, by the end of the fight it was clear who would be declared the victor	*Although both fighters were eebenly matched, by the end ob the fight it was clear who would be declared the bictor*	*victor*	*/v/* ➔ */b/*
Since they didn't have any plans for the rest of the evening, the couple decided to drive to the theater to see a movie	*Since they didn't hab any plans for the rest of the eebning, the couple decided to dribe to the theater to see a moobie*	*movie*	*/v/* ➔ */b/*
Rather than try to carry all of their furniture by themselves, Garry and Cole decided it was best to hire some movers	*Rather than try to carry all ob their furniture by themselbes, Garry and Cole decided it was best to hire some moobers*	*movers*	*/v/* ➔ */b/*
The woman owned a variety of pets, and there was one dog in particular that visitors noted as being especially vicious	*The woman owned a bariety of pets, and there was one dog in particular that bisitors noted as being especially biscious*	*vicious*	*/v/* ➔ */b/*
Finally, after the prosecution had tried in vain to come up with a convincing argument, the defendant's innocence was proven	*Finally, after the prosecution had tried in bain to come up with a conbincing argument, the defendant's innocence was prooben*	*proven*	*/v/* ➔ */b/*

Unaccented Sentence	Accented Sentence	Target Word	Phoneme Swap
The first hurdle the detective had to overcome on this case was trying to establish a motive	*The first hurdle the detectib had to obercome on this case was trying to establish a moetib*	*motive*	*/v/ ➔ /b/*
The man had a very sophisticated taste, and it was quite obvious from the fact that most of his furniture was antique, and most of his clothes were vintage	*The man had bery sophistocated taste, and it was quite obious from the fact that most ob his furniture was antique and most ob his clothes were bintage*	*vintage*	*/v/ ➔ /b/*
Even though the director thought his most recent film was his most creative work yet, he was still sad to see the scathing review	*Eeben though the director thought his most recent film was his most creyaytib work yet, he was still sad to see the scathing rebiew*	*review*	*/v/ ➔ /b/*
As he walked around the village, the explorer saw many buildings covered in vines	*As he walked around the billage, the explorer saw many buildings cupboard in bines*	*vines*	*/v/ ➔ /b/*
To everyone's amazement, instead of falling as he stepped off the stage, the magician stood on the air and hovered	*To ebbrywun's amazement, instead ob falling as he stepped off the stage, the magician stood on the air and hubbered*	*hovered*	*/v/ ➔ /b/*
Sam had been putting off his French vocabulary homework all day, but by the time the evening rolled around it was impossible to avoid	*Sam had been puting off his French beau cab yewlarry homework all day, but by the time the eebening rolled around it was impossible to a boid*	*avoid*	*/v/ ➔ /b/*

Unaccented Sentence	Accented Sentence	Target Word	Phoneme Swap
Since Joy was turning three this month, her parents decided to throw a party for her birthday	*Since Joy was turning tree this maunt, her parents decided to trow a party for her birtday*	*birthday*	/θ/ ➔ /t/
The running club decided that they would hold their thirtieth anniversary marathon on the seventh	*The running club decided that they would hold their turtiette anniversary maratawn on the sevent*	*seventh*	/θ/ ➔ /t/
With her boss paying close attention to the project, everything needed to go smoothly, so Mandy made sure all the initial checks were totally thorough	*Wit her boss paying close attention to the project, everyting needed to go smoothly, so Mandy made sure all the initial checks were totally torough*	*thorough*	/θ/ ➔ /t/
Without a second thought, Helen decided to enter the raffle even though her chances of winning with only one ticket were less than one in a thousand	*Witout a second tought, Helen decided to enter the raffle even though her chances of winning wit only one ticket were less than one in a tousand*	*thousand*	/θ/ ➔ /t/
It was a boring day with nothing to do, so Mark thought he would go for a drive to the theater	*It was a boring day wit nutting to do, so Mark taught he would go for a drive to the teeutter*	*theater*	/θ/ ➔ /t/
She had walked for three hours through the desert and by this point she was feeling very thirsty	*She had walked for tree hours true the desert and by this point she was feeling very tirsty*	*thirsty*	/θ/ ➔ /t/

Unaccented Sentence	Accented Sentence	Target Word	Phoneme Swap
Because the remote island could only be accessed by a three hour boat ride through treacherous waters, the local population of birds was thriving	*Because the remote island could only be accessed by a tree hour boat ride true treacherous waters, the local population of birds was triving*	*thriving*	/θ/ ➔ /t/
Even without being able to read minds, it was very obvious through his face what the man was thinking	*Even witout being able to read minds, it was very obvious true his face what the man was tinking*	*thinking*	/θ/ ➔ /t/
As an entymologist, Oliver thought all sorts of bugs were amazing, but his absolute favourite things were moths	*As an entymologist, Oliver taught all sorts of bugs were amazing, but his absolute favourite tings were mawts*	*moths*	/θ/ ➔ /t/
The storm had passed through about an hour ago, but there was still a thin rain and they could still hear the sound of thunder	*The storm had passed true about an hour ago, but there was still a tin rain and they could still hear the sound of tunder*	*thunder*	/θ/ ➔ /t/
They had been stranded without help for three days, and at this point it was becoming obvious that someone would have to do something	*They had been stranded witout help for tree days, and at this point it was becoming obvious that someone would have to do someting*	*something*	/θ/ ➔ /t/
Even though the new fourth grade student had only moved to the country three days ago, the teacher still asked him to stand for the national anthem	*Even though the new fort grade student had only moved to the country tree days ago, the teacher still asked him to stand for the national antem*	*anthem*	/θ/ ➔ /t/

Unaccented Sentence	Accented Sentence	Target Word	Phoneme Swap
The couple thought they heard something outside, but when they opened the door to check, nothing was there	*De couple thought dey heard something outside, but when dey opened de door to check, nothing was dare*	*there*	*/ð/ ➔ /d/*
Isaac decided that he didn't need so many possessions and decided to donate many of them to charity, including a large portion of his clothes	*Isaac decided dat he didn¥'t need so many posessions and decided to donate many of dem to charity, including a large portion of his clodes*	*clothes*	*/ð/ ➔ /d/*
The couple had forgotten to ask their neighbour to water their plants before going on vacation for two weeks, and by the time they got home, every plant had withered	*De couple had forgotten to ask dare neighbour to water dare plants before going on vacation for two weeks, and by the time dey got home every plant had widdered*	*withered*	*/ð/ ➔ /d/*
Although most people found the sound of birds in the early morning quite irritating, Zoey found it oddly soothing	*Aldoe most people found de sound of birds in de early morning quite irritating, James found it oddly sooding*	*soothing*	*/ð/ ➔ /d/*
As the drummer of the band, Tom was responsible for keeping the rhythm	*As de drummer of de band, Tom was responsible for keeping de riddum*	*rhythm*	*/ð/ ➔ /d/*
Out of all his peers in the track club, Albert was able to throw the javelin the farthest	*Out of all his peers in de track club, Albert was able to throw de javelin de fardist*	*farthest*	*/ð/ ➔ /d/*

Unaccented Sentence	**Accented Sentence**	**Target Word**	**Phoneme Swap**
Rachel looked to her friends for advice, even though it was clear that they didn't know what to do either	*Rachel looked to her friends for advice, even dough it was clear dat dey didn¥'t know what to do eider*	*either*	*/ð/ ➔ /d/*
Although Eric was still mad that Will had forgotten his birthday, he still forgave him because it was his brother	*Aldoe Eric was still mad dat Will had forgotten his birthday, he still forgave him because it was his brudder*	*brother*	*/ð/ ➔ /d/*
The gift was clearly something Abby had repackaged from another gift exchange, but even so Sarah was still thankful	*De gift was clearly something Abby had repackaged from anudder gift exchange, but even so Sarah was still dankful*	*thankful*	*/ð/ ➔ /d/*
The new couch cost quite a bit of money, but that was mainly because the upholstery was real leather	*De new couch cost quite a bit of money, but dat was mainly because de upholstery was real ledder*	*leather*	*/ð/ ➔ /d/*
Rebecca had a busy day, and the only thing she had eaten so far was a rather bland smoothie	*Rebecca had a busy day, and de only thing she had eaten so far was a radder bland smoodi*	*smoothie*	*/ð/ ➔ /d/*
It took several minutes, and much coaxing from the parents, but eventually the children started to gather	*It took several minutes, and much coaxing from de parents, but eventually de children started to gadder*	*gather*	*/ð/ ➔ /d/*

Table A2. Sentences presented in the practice portion of the experiment

Practice Sentence	Accented Condition	Noise Condition
The children had been playing outside all day, and as the sun started to set their parents called them in for dinner	Unaccented	No noise
After buying the new recipe book, Joey was eager to crack it open and start cooking	Unaccented	Noise
On her day op, Jenniper decided to walk to the cape down the street to get a cup of coppee	Accented	No noise
Now that the lain had stopped and the clouds had bloken up, they could finally enjoy their summer blake	Accented	Noise