Accent Perception and Adaptation During Lexical Access: Expand, Shift, or Pathway

by

Lauren R. Franklin

Sc.M., Brown University, 2017

B.A., The University of Texas at Austin, 2013

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the Department of Cognitive, Linguistic, and Psychological Sciences at Brown University

Providence, Rhode Island

July 2021

This dissertation by Lauren R. Franklin is accepted in its present form by the Department of Cognitive, Linguistic, and Psychological Sciences as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

Date _____ _____
Dr. James L. Morgan, Advisor

Recommended to the Graduate Council

Date _____ _____
Dr. Uriel Cohen Priva, Reader

Date _____ _____
Dr. Katherine White, Reader

Approved by the Graduate Council

Date _____ _____
Dr. Andrew G. Campbell, Dean of the Graduate School

Curriculum Vitae

Lauren R. Franklin
Department of Cognitive, Linguistic, and Psychological Sciences
Brown University, Providence, RI 02912

Education

Brown University, Providence, RI — PhD in Cognitive Science, July 2021
Department of Cognitive, Linguistic, and Psychological Sciences
Dissertation title: "Accent perception and adaptation during lexical access: expand, shift, or pathway"

Brown University, Providence, RI — MS in Cognitive Science, May 2017
Department of Cognitive, Linguistic, and Psychological Sciences
Thesis title: "On the nature of vocalic representation during lexical access"

University of Texas at Austin, Austin, TX — BA, December 2013
Linguistics, Plan II Honors, Middle Eastern Languages and Cultures, Magna Cum Laude University Honors, Phi Beta Kappa, National Society of Collegiate Scholars

Research Experience

Graduate Student Researcher, Brown University
Metcalf Infant Research Lab
Primary Investigator: Dr. James L. Morgan

Undergraduate Research Assistant, The University of Texas at Austin
UT Phonetics Lab
Primary Investigator: Dr. Rajka Smiljanic

Work Experience

Early Childhood Educator — 2014
Beth Yeshurun Day School, Houston TX

Administrative Assistant — 2015
Children's Aid Society of Ontario, Toronto, ON, Canada

Grants and Fellowships

Edgar Lewis Marston Scholarship Fund, Hardin-Simmons University
Peter D. Eimas Graduate Fund
Brown University CLPS Peter D. Eimas Graduate Research Award

Teaching Experience

Graduate Teaching Assistant, Brown University
Statistical Methods, Fall 2017
Teacher: Dr. Kathryn Spoehr
Enrollment: 52
Evaluation: 1.70 (Effective)
Laboratory in Psycholinguistics, Spring 2018
Teacher: Dr. James L. Morgan
Enrollment: 12
Evaluation: 1.57 (Very Effective)
Mind, Brain & Behavior, An Interdisciplinary Approach, Fall 2018
Teacher: Dr. Elena Festa
Enrollment: 213
Evaluation: 1.92 (Effective)
Language & the Mind, Spring 2019
Teacher: Dr. James L. Morgan
Enrollment: 36
Evaluation: 1.77 (Effective)
Human Cognition, Spring 2020
Teacher: Dr. Kathryn Spoehr
Enrollment: 125
Evaluation: 4.48 (Very Effective)

Sheridan Center Teaching Seminar Facilitator, Brown University, 2017-2019

Leadership, Service, and Professional Development

Sheridan Center Liaison — 2016 - 2021
Brown University

Teaching Material Developer — Summer 2019
Mind, Brain & Behavior, An Interdisciplinary Approach, CLPS Dept., Brown University

LingLangLunch Talk Series Coordinator — 2016 - 2019
CLPS Department, Brown University
Colloquium Committee Graduate Student Representative — 2017 - 2019
CLPS Department, Brown University
Sheridan Center Teaching Consultant — 2017 - 2019
Brown University

Publications, Presentations, & Posters

Franklin, L. R., Tin, S., Gasdaska, B., & Morgan, J. L. (in preparation). *Consonantal and vocalic representation during lexical access*, Brown University.

Franklin, L. R. (April 2021). *Toddlers show adult-like sensitivity to consonants and vowels in early word representations.* LingLangLunch Brown Bag Presentation, Brown University, Providence, RI.

Franklin, L. & Morgan, J. (2020). *Toddlers show adult-like sensitivity to consonants and vowels in early word representations.* International Conference of Infant Studies, July 6-9, 2020 meeting [poster], virtual conference.

Franklin, L. R. (March 2020). *Phonological representation of accented speech*. LingLangLunch Brown Bag Presentation, Brown University, Providence, RI.

Franklin, L. & Morgan, J. (2017). *For toddlers, like adults, vowel mispronunciations are readily detected but do little to impede lexical access*. 42nd Annual Meeting of Boston University Conference on Language Development, November 2017 meeting [poster], Boston, MA.

Franklin, L. & Morgan, J. (2017). *On the nature of vocalic representation during lexical access.* 173rd Meeting of the Acoustical Society of America, June 2017 meeting [poster], Boston, MA.

Masapollo, M., Zhao, T.C., **Franklin, L**., & Morgan, J.L. (2019). Asymmetric discrimination of non-speech tonal analogues of vowels. Journal of Experimental Psychology: Human Perception and Performance, 45(2), 285-300.

Masapollo, M., Polka, L., Ménard, L., **Franklin, L.**, Tiede, M., & Morgan, J.L. (2018). Asymmetries in unimodal visual vowel perception: The roles of oral-facial kinematics, orientation and configuration. Journal of Experimental Psychology: Human Perception and Performance, 44(7), 1103-1118.

Masapollo, M ., Polka, L., Morgan, J., **Franklin, L.**, & Ménard, L. (2017) *Asymmetric discrimination of phonetically-incongruent audio-visual vowels.* 174th Meeting of the Acoustical Society of America, December 2017 meeting, New Orleans, LA.

Preface and Acknowledgements

This dissertation work has been a long time in the making, and my family and friends have been there with me throughout, through all the ups and downs. Mom, Dad, Geeta, Vijay, Bibi, Baba, thank you for your support, even though my work can be pretty esoteric. To all my friends I met during my time at Brown, thank you for going through all the good and bad times with me, commiserating with me, and celebrating with me. Thank you especially to Boyoung Kim, Shiying Yang, Brittany Baxter, Danny Ullman, Junwen Lee, and Youtao Liu—you have inspired me in so many ways, both professionally and personally. Keli Chang, my bestie, thank you for being a constant presence in my life and for listening to me whenever I need you. You have no idea how much you helped and supported me. Ahad Khattak, my partner in life, you have always been the strongest voice advocating for me, lifting me up, instilling confidence in me when I've worn thin, and being so patient and caring always. This PhD is probably part yours, because I don't think I would have completed it without you. Thank you for everything you've done for me, I am so blessed to have you in my life.

I would like to thank my committee members, Katherine White and Uriel Cohen Priva, who have been so kind and helpful throughout my first year project, prelim and dissertation, always providing guidance and thoughtful feedback and talking through problems with me. Thank you also to Sheila Blumstein for your mentorship in my early years at Brown, to Chelsea Sanker, who has been invaluable in her knowledge, advice, and patience, and to all of the faculty and staff that were always so kind and helpful.

Finally, thank you to my advisor, Jim Morgan, for taking me on as your student. I am so grateful that you were my advisor for the last six years. Through your mentorship, you have made me a better researcher and a better scientist, and I appreciate the understanding and kindness you showed me throughout. I will miss your friendly, jolly attitude, our lab meeting life updates, and of, course, your famous eggnog.

Table of Contents

List of Tables and Figures

Chapter 1: General Introduction

Every day, humans encounter a variety of people speaking in a variety of tones and timbres, using a variety of dialects, in a variety of acoustic environments. Even when the same speaker, at the same place and time, repeats a word, those two utterances of that word will never be exactly the same. Therefore, when listening to speech, listeners encounter endless variability on an endless number of dimensions. Despite this apparent lack of invariance in the physical signal, however, listeners quickly and easily "hear through" variability when processing and understanding speech. It is likely that listeners accomplish this feat at least in part by identifying the phonological units that make up the target word, underlining the importance of phonological representation within the mind of a listener.

The lack of invariance in the speech signal has often, been framed as a problem for listeners to overcome. However, it is likely that listeners don't merely ignore variability, but rather encode it, allowing them to use information about the nature of linguistic variability during speech processing. Indeed, a good deal of variability is systematic, predictable, and contextually grounded. Sociolinguists have long demonstrated that age, gender, sexual orientation, socioeconomic status, social group, familiarity among interlocutors, affect, and many other social factors have effects on aspects of speech like production of phonological forms. For example, male and female speakers of English appear to produce /s/ and /ʃ/ with differences in spectral distribution above and beyond what would result from physiological difference, and listeners likewise are sensitive to these differences, categorizing /s/ and /ʃ/ differently depending on the

perceived sex of the speaker's voice and face (Strand & Johnson, 1996; Strand, 1999). Another example is of pop science's biggest linguistic pet peeve of the twenty-first century: vocal fry, or creaky voice, in American English. Associated first and foremost with young women, vocal fry does not appear to affect the intelligibility of speech despite being quite different acoustically from non-creaky speech. However, it can affect listeners' judgements of qualities such as intelligence and likability (Parker & Borrie, 2018), suggesting that listeners do not simply ignore variability, but use it to make social decisions.

Within speech perception, as well, a great deal of previous research has demonstrated that much of the variability observed in the speech signal is not random, but predictable within a given context. Lahiri et al. (1984) showed a type of invariance they called "dynamic relative invariance", through which place of articulation of stop consonants could be determined phonetically by comparing *relative* spectral changes across time (Blumstein, 1986). Arnon and Cohen Priva (2013) demonstrated that phonetic durations are reduced as a function of sentence frequency. Other previous research suggests that systematic variation such as this can be remembered and used by listeners during speech processing—listeners have demonstrated sensitivity to variation in talker's voice (Magnuson & Nusbaum, 2007; Mullennix & Pisoni, 1990). They also appear to remember talker-specific characteristics (Bradlow et al., 1999; McLennan & Luce, 2005; Nygaard et al., 2000). This is all to say that many types of variability are systematic and can therefore be used by listeners to overcome the problem of the lack of invariance.

Given that variability is an inescapable fact of speech and that listeners seem to use variability during speech processing, it follows that variability of the speech signal may affect how units of language, such as phonemes, are stored within the mind.

It appears that native listeners possess highly detailed knowledge of the phonological categories of their language, and they can harness this detailed knowledge during word recognition (Goldinger, 1998; Nygaard & Pisoni, 1998; Luce & Lyons, 1998 and works cited therein). Adults are sensitive to subphonemic details within words, showing sensitivity to differences in phonemes along phonological dimensions such that /p/ and /b/ are more "related" than /p/ and /v/, for example (Miller & Nicely, 1955; White, Yee, Blumstein & Morgan, 2013; Franklin, Tin, Gasdaska, & Morgan, in preparation). Adults also show better performance in priming and word recognition tasks when phoneme tokens are prototypical (Sumner, 2013; Sumner & Samuel, 2004; Grieser & Kuhl, 1989). Additionally, native speakers show continuous perception of vowels across between-category continua, suggesting detail extends to differences even within a single phonemic category, at least for vocalic phonemes (Fry, Abramson, Eimas, & Liberman, 1962).

Toddlers, early in their language-learning journey, also exhibit evidence of detailed phonological representations. For example, Swingley and Aslin (2000) observed that when phonemes in recognizable words were replaced with similar, highly confusable phonemes (e.g. *baby* > *vaby*), 23-month-olds could still recognize the words, but recognition was significantly poorer than for correctly pronounced words. Bailey and Plunkett (2002) also showed that 18- and 24-month-olds were able to detect slight (1- or

2-feature) mispronunciations in familiar words, again suggesting detailed phonological representations for words they know. Indeed, toddlers have demonstrated adult-like sensitivity to phoneme features for consonants and vowels in various positions within words (White & Morgan, 2008; Ren & Morgan, 2011; Franklin & Morgan, 2020). Work such as this demonstrates that from an early age, humans have detailed but flexible representations of the sounds and words of their native language, which allow them to recognize familiar forms, generalize to new similar forms, and adapt to novel forms (Kleinschmidt & Jaeger, 2015).

While representations are detailed, listeners also demonstrate an ability to use top-down information to "fill in" informational gaps in the speech signal. Ganong (1980), for example, showed that when participants heard an ambiguous sound, such as one with a VOT between /g/ and /k/, they were likely to categorize it as /g/ when it preceded /ɪft/, but /k/ when it preceded /ɪs/; that is, ambiguous forms were heard as words rather than nonwords. Fox and Blumstein (2016) extended these effects to semantic context, embedding phonetically ambiguous target words into sentences that biased listeners toward one of the two targets (e.g. "Valerie hated the…" or "Brett hated to…" followed by "buy"/"pie" with an ambiguous VOT between /b/ and /p/. They found that perception of these ambiguous onsets was in fact biased by semantic context, demonstrating top-down effects of lexical feedback on prelexical representations (Fox & Blumstein, 2016). Similar effects have been shown with grammatical number cues (Brown, Fox, & Strand, 2020). Predictable contexts have also shown to improve intelligibility of speech in adverse listening conditions (e.g. masked with speech-shaped noise or multi-talker

babble, or gated) (van der Feest, Blanco, & Smiljanic, 2019; Van Engen, Phelps, Smiljanic, & Chandrasekaran, 2014; Gilbert, Chandrasekan & Smiljanic, 2014). It appears that listeners flexibly modulate their use of bottom-up and top-down information in the face of variability, but it is not clear how this process alters their short- and long-term phonological representations, as listeners are able to remember previous experiences and use them in similar environments to enhance processing efficiency (Baese-Berk, Bradlow, & Wright, 2013; Bradlow & Bent, 2008; Sidaras, Alexander, & Nygaard, 2009, White & Aslin, 2011).

A source of variability in language that has garnered attention in recent years is accent. Accent is uniquely posed to help answer questions regarding phonological processing, as it is a common source of variability that includes categorical mismatches in phonemes across dialects and nonnative accents. Furthermore, accents may be completely novel (i.e., an accent that the listener has never heard before) or they may be familiar to varying degrees (e.g., a British accent to American listeners). Finally, regional accents often differ *systematically* from each other, allowing us to ask questions about generalization during accent adaptation.

In English, there are often different realizations of vowels in different regional varieties, making processes of vowel perception important components in understanding how listeners "hear through" divergent accents. Previous work has shown that listeners perceive vowels and consonants differently—while consonants are perceived discretely, with clear boundaries between consonant categories (especially between obstruents, somewhat less for sonorants), vowels are generally perceived more continuously, with

less distinct boundaries between vowel categories (Liberman et al., 1954; Fry et al., 1962; Pisoni, 1973; etc.). During lexical access in particular, research has demonstrated a consonant bias, in which listeners treat vowels as more mutable than consonants, and show that consonants constrain lexical access more strongly than vowels (Cutler et al 2000; Nespor et al 2003; Hochmann et al 2011; Nazzi 2005; Havy et al 2014; Floccia et al 2014; Nazzi et al 2009). For example, Cutler and colleagues' (2000) word reconstruction task showed that when participants are presented with a non-word, such as /kibrə/, in which either one consonant or one vowel can be changed to create a real word (e.g., zebra or cobra), participants are more likely to change vowels than consonants. Additionally, semantically related lexical items from the same root often share more consonants than vowels, e.g., *serene* and *serenity* (Cutler, 2012) and in Semitic languages, lexical roots are entirely consonantal, and different vowel patterns denote different but semantically related meanings (e.g., in Hebrew, *lamad* means "he studied" while *limed* means "he taught"). Additionally, in corpus studies of functional load, vowel contrasts were shown to be less informative overall than consonant contrasts, demonstrating differences in representation and processing of consonants and vowels in a number of unrelated languages (Oh et al., 2015).

Finally, Franklin, Tin, Gasdaska, and Morgan (in preparation) similarly show that listeners are more likely to accept changes of vowels as mispronunciations of known words (rather than novel words) than changes of consonants. However, looking patterns in these experiments showed that while vowel changes did not constrain lexical access as strongly as consonant changes, even single-feature mispronunciations of vowels resulted

in delayed looking, indicating that listeners are *sensitive* to even single-feature mispronunciations of vowels. Therefore, it is not that listeners' knowledge of vowel categories is less detailed, but rather that listeners accept, or possibly even *expect*, variation within and across vowel categories, as is often realized in accent. However, the way people process and represent vowel categories has not yet been well studied with respect to accent perception. Thus, this dissertation will focus on vowel manipulations in order to elucidate the effect of accent on vowel representations.

Previous work suggests that although accented speech can initially result in slower or less efficient processing (Adank, Evans, Stuart-Smith, & Scott, 2009; Floccia, Goslin, Girard, & Konopczynski, 2006; Adank & Janse, 2010), listeners can adapt to novel accents rapidly (Bent, et al., 2016; Bradlow & Bent, 2003; Clarke & Garrett, 2004), and that with continued experience, listeners encode subtle phonetic details and use this information to efficiently access lexical items and generalize to other speakers from the same accent community (Clopper and Pisoni, 2004; Evans and Iverson, 2004; Bradlow & Bent, 2003). Maye, Aslin, and Tanenhaus (2008) investigated how listeners adjust to the specificity of a novel accent and the subsequent effects of this exposure on lexical access. In this study, participants listened to a 20-minute passage of speech produced with a novel accent that was created by lowering all of the front vowels by one phonological position in American English vowel space such that the phrase "the wicked witch of the West" was heard as "the weckud wetch of the Wast" (/ðʌ wɛkəd wɛt͡ʃ ʌv ðʌ wæst/). Participants then completed a lexical decision task in which they judged whether the items they heard were words or non-words. Listeners were able to generalize their

knowledge of the shifted vowels to words they had not heard in the passage: when the word in the lexical decision task included a lowered front vowel, but not when the word included a shifted back vowel, participants were more likely to identify the word after listening to the passage.

When listeners hear a new speaker with an unfamiliar accent, their phonological representations should be flexible enough to accommodate this new speaker. Because listeners often display detailed knowledge of native phonological categories, modulated by speaker characteristics, lexical context, and a variety of other factors, it appears to be the case that listeners adjust their category representations based on the input they receive. It is possible that these adjustments influence future speech perception, pointing to a possible change in underlying phonemic representations.

Exemplar models of lexical access approach the issue of variability by positing that listeners store acoustic traces within phonological and lexical representations such that incoming speech can be matched to existing representations. Using exemplar models as a basis for our exploration of accent's effects on phonological representations, I ask the following questions:

When listeners encounter category differences, how do their phonological representations change to incorporate this new input? Do the category boundaries "shift" toward the new input? Do they "expand" to incorporate *more types* of input? Or does a "pathway" form between the new input and older representations?

These three questions suggest possible mechanisms for how people adapt their sound categories to accented speech. For example, the *expand* hypothesis suggests that

the distribution of sounds considered to fall within a given category is broadened, perhaps via an overall lowering of the decision criterion used, allowing for overall flexibility of the category's boundaries. With this sort of hypothesis/mechanism, one might expect symmetric adaptation effects in all directions around the category prototype. In particular, if a category is symmetrically expanded to include tokens similar to exposure exemplars, it should also be expanded to include tokens on the "other side" of the category. This hypothesis is supported by studies in which listeners heard phonological changes to lexical items in exposure and then in test adapted to changes that they were not familiarized to (Schmale et al., 2012, Witteman et al., 2010, Eisner et al., 2010, Weatherholtz, 2015).

The *shift* hypothesis, on the other hand, suggests that the distribution of sounds within a given category is moved along some relevant dimension, perhaps due to some modification of what is considered prototypical for that category. Like the *expand* hypothesis, the *shift* hypothesis predicts changes in category boundaries, but these differ from the predictions from the *expand* hypothesis in one key aspect: now, tokens on "other side" of the category prototype might not be accepted as possible exemplars of the target category. In this case, even exemplars on the "other side" that were previously considered to be part of the category might no longer be accepted. Evidence for the shift hypothesis comes mainly from work by Maye et al. (2008). These authors found that after exposure to a passage with lowered front vowels, listeners were more likely to accept words with lowered front vowels as part of their intended lexical category, but were not more likely to accept words with raised front vowels.

Finally, the *pathway* hypothesis might also be considered an asymmetric expansion hypothesis, similar to the *shift and expand* hypothesis suggested by Hitczenko and Feldman (2016). According to the *pathway* view, the category is broadened to include a new set of exemplars (and presumably tokens falling between the old category and this new set), but tokens falling on the "other side" of the category are unaffected. A landscape model like DRIBBLER (Morgan, ms.), in which exemplars can be visualized as water droplets eroding the representational landscape and creating slopes, valleys, and deep reservoirs, could provide the mechanism for such category adjustments. Put another way, acoustic forms are encoded as weight modifications on a category, creating a quantized perceptual space (Johnson, 1997).

In the following thesis, the *expand, shift* and *pathway* hypotheses are tested empirically in order to better understand adult perception of accented speech. In

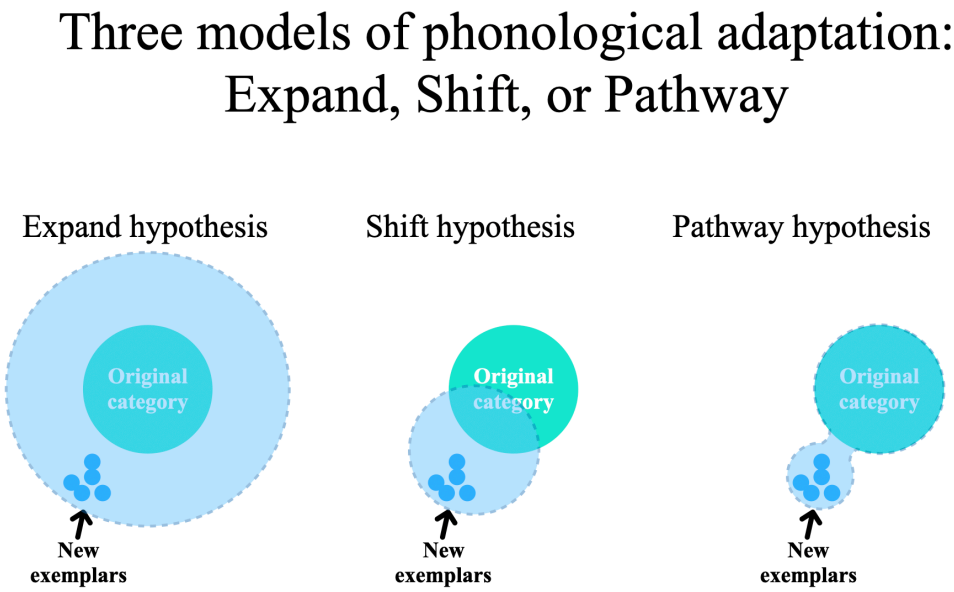

Figure 1: Three models of phonological adaptation: expand, shift, or pathway

Experiment 1, systematic features of Spanish-accented English (Experiment 1A) and British-accented English (Experiment 1B) are used to test how the phonological category representations change as listeners are tasked with understanding an accent different from their own. I explore the level of detail available to listeners who have some existing representation of an accent that is different from their own. In Experiment 2, I ask how exposure to a novel accent affects word endorsement rates in a lexical decision task using a paradigm similar to Maye, Aslin, and Tanenhaus (2008). Finally, in Experiment 3, an eye tracking paradigm is used to explore how minimal exposure to a novel accent affects perception of mispronounced words. With these experiments, I examine how listeners' phonological perceptions change across different experimental tasks, and how participants' performance aligns with the varying predictions of these competing *expand, shift,* and *pathway* hypotheses.

Chapter 2: Representations of Familiar Accents

Introduction

Listeners can easily overcome mismatches that occur due to accent when processing speech. Prior research has demonstrated that, often, listeners use non-acoustic information, such as lexical, semantic, syntactic, and probabilistic knowledge, to disambiguate ambiguous phonetic signals (Ganong, 1980; Davis et al., 2008; Maye et al., 2008; Fox & Blumstein, 2016; Kleinschmidt, Weatherholtz, & Jaeger, 2018; Luthra et al., preprint). Furthermore, some work has shown that while using top-down information in the face of variability, phoneme category boundaries can (at least, temporarily) shift. For example, Norris et al. (2003) presented Dutch listeners an ambiguous sound between /f/ and /s/ in various lexical contexts. Some participants heard lexical contexts that biased them toward hearing /f/ (e.g. *witlof*, "chicory") while others heard contexts that biased them toward /s/ (e.g. *naaldbos,* "pine forest"). In a subsequent sound categorization task, Norris et al. found that listeners' phonemic boundary between /f/ and /s/ was shifted as a function of the biasing condition that they heard. For vowels, too, category boundaries can be influenced by formant manipulations of neighboring vowels (Mitterer, 2006; Sjerps, McQueen, & Mitterer, 2013). Results such as these suggest that exposure to variability can affect the *shape* of a phonemic representation, but for vowels in particular, which are relatively mutable, to what extent or for how long is still not fully understood.

There is evidence that vowels and consonants are treated differently in speech processes, as vowels are often perceived less categorically than consonants (Liberman et al., 1954; Fry et al., 1962; Pisoni 1973; Kronrod, Coppess, & Feldman, 2016) and vowels

constrain lexical processing less strongly than consonants (Cutler et al., 2000; van Ooijen, 1996; New, Araùjo, & Nazzi, 2008; New & Nazzi, 2015; Mehler, Peña, Nespor & Bonatti, 2006; Toro, Nespor, Mehler & Bonatti, 2008; Franklin et al., in preparation). However, at least in English, native accents tend to differ most in how vowels are realized. This observation, in conjunction with the effects of variability on phonemic representations and the more continuous representation of vowels raises the question: how does exposure to accents affect the shape of vowel representations?

Evans and Iverson (2004) tested listeners from Northern England, half of whom had lived only in Northern England, and half who had been born and raised in the North but had been living in London for at least a year. These participants were asked for goodness ratings of vowels presented in Northern- and Southern-accented sentences. Evans and Iverson found that northerners living in London adjusted their goodness ratings as a function of which accent they heard, but northerners still living in the North did not—all of their goodness ratings reflected their native vowel categories. These findings were surprising, as northern speakers tend to have extensive experience with Southern English dialects through media, despite not living in a place where those dialects are spoken. Using the same methods, Evans and Iverson (2007) were not able to replicate these findings with Northern participants who had been attending university with speakers of Southern dialects, once again calling into question how much experience and what quality of experience with a dialect is needed in order to internalize its phonological system. In a similar vein of research, Shaw et al. (2018) tested if Australian listeners would recalibrate goodness ratings of vowels when exposed to the accent in

which those vowels occurred. Results showed no adaptation effects for any of the four regional accents they were exposed to, suggesting tolerance of accented vowels, but no perceptual shift.

Notably, Evans and Iverson (2004, 2007) and Shaw et al. (2018) relied on goodness ratings, but this measure may not be sensitive enough to demonstrate if and how listeners incorporate accented exemplars into their phonological representations. In other words, while listeners may not rate a vowel as a good example of a given category, they may nevertheless remember features of a given accent, which can be used to generalize to new speakers. Studies on dialect and accent classification show that some children as young as four are able to differentiate between talkers with different regional accents, and that these skills continue to improve into adulthood, at which point listeners can use accent information to accurately classify talkers based on regional background (Jones et al., 2017; Clopper & Pisoni, 2004; Van Bezooijen & Gooskens, 1999; Yan, 2015; Clopper & Pisoni, 2007; Clopper & Bradlow, 2009). The ability to accurately group talkers in this way suggests either long term or short term adaptations. That is, participants may implicitly remember previous experiences with similar speakers and use that information to classify study speakers, or participants may be able to track similar phonetic patterns across speakers from a given region, suggesting short-term calibration to dialectal features.

In the literature, there is evidence for both of these possibilities. Evidence for long-term effects come from work like McGarr (1983), who investigated how listeners adjust to deaf speech, which is similar to accented speech. Comparing individuals who

had years of experience listening to deaf talkers and those who had little-to-no experience, McGarr found that listeners with experience performed better than those without on every task, suggesting that those with experience encoded information about deaf speech and were able to access that information during experimental tasks. Long-term effects of adaptation have also been shown for synthetic speech (Schwab, Nusbaum, & Pisoni, 1985; Greenspan, Nusbaum, & Pisoni, 1988), time-compressed speech (Dupoux & Green, 1997; Pallier et al., 1998), and speech with shifted phonetic category boundaries (Kraljic & Samuel, 2005; Eisner & McQueen, 2006). Work like that of Clarke and Garrett (2004) show that with less extensive training, participants could still adapt to accented speech. When exposed to short blocks of either Spanish- or Chinese-accented sentences, Clarke and Garrett's participants showed gradual but substantial improvement on an intelligibility task throughout the experiment, showing that in the short term, listeners track accented features to improve speech processing.

Additionally, the contrast tested in Evans and Iverson (2004, 2007) relied on participants to make distinctions within a single category (/ʌ/, which is raised in Northern English accents compared to Southern accents). However, many accents differ from each other more drastically, as words in one accent may be produced with categorically different phonemes than the same words in another accent. Studies that include cross-category accent mismatches have demonstrated more rapid adaptation to accented speech (Clarke & Garrett, 2004; Bradlow & Bent, 2008, Maye, Aslin, & Tanenhaus, 2008). Perhaps when accented speech is close enough to a target, i.e. within a given category, participants' phonemic representations do not need to change greatly to accommodate the

differences that arise due to accent, as the new exemplar is already within the category distribution. There may only be evidence of a change in phonemic representations when the distance between the heard phoneme and the target phoneme spans category boundaries.

Therefore, in the current experiments, which focus on long-term effects of accent representation, participants are tested on cross-category distinctions between somewhat familiar accents in order to better understand the level of phonological detail available to listeners when they are asked about their representation of a given accent. This experiment investigates if participants are sensitive to cross-category accent mismatches (rather than within-category shifts, as in Evans and Iverson (2004, 2007), as well as if participants retain detailed information about the nature and direction of said accent mismatches over the long-term, or if they use a strategy of global expansion to widen the pool of possible phonemic/lexical candidates during accented speech processing.

In Experiment 1A, participants are tested on their representations of a systematic difference between Spanish-accented English (SpE) and Standard American-accented English (SAE)—where SAE speakers say /ɪ/, SpE speakers say /i/ (cf. Sidaras, Alexander, & Nygaard, 2009). Because SAE speakers generally have at least some exposure to SpE (through television, film, and other media, as well as, often, in person), I sought to test if SAE speakers are aware of this specific alternation between SAE and SpE or if they, rather, generally expand their criteria in their phonological representations when they encounter "accentedness". In this experiment, listeners heard versions of the word "fish" that ranged either from [ɪ] to [i] or [ɪ] to [ɛ] and were asked to indicate which sounded

more “Spanish-accented”. I hypothesized that if listeners have a detailed representation of SpE, they would only accept [i] as Spanish-accented. However, if they do not store details about other accents in their phonological representations, they would be just as likely to accept [ɛ] as Spanish-accented, since it is also diverges from SAE [ɪ/.

In Experiment 1B, participants were tested on their representations of a slightly more complicated systematic difference, this time between Received Pronunciation British-accented English (BE) and SAE. Among some documented differences in American and British English, one in particular involves a shift from /æ/ to /ɑ/ in environments before voiceless fricatives /_θ/, / _ʃ/, and /_f/ (Wells, 1982; Hosseinzadeh, 2015). For example, the word “bath” is pronounced as /bæθ/ in SAE, but /bɑθ/ in BE. However, before stop consonants, both varieties use /æ/, so “bat” sounds like /bæt/ in both SAE and BE. Similar to SpE, Americans also likely have at least some exposure to BE, again through television, film, and other media. Therefore, they are once again likely to have some representation of BE—our question is, how detailed is this representation? Have listeners internalized this phonological rule in their representation of BE, or do they over-generalize this varietal difference? To test this, participants were played versions of “bath” and “bat” with word-medial vowels ranging between /æ/ and /ɑ/ and were asked which ones sounded more “British-accented.” I hypothesized that if participants correctly chose /bɑθ/ as more British-accented but also incorrectly chose /bɑt/ as more British-accented, this would indicate that participants have *some* representation of BE’s phonology, but it is not robust or highly detailed. On the other hand, if they correctly chose both /bɑθ/ and /bæt/ as more British-accented, this would indicate that listeners

store detailed representations of other native accents that they can access during speech processing.

Experiment 1A: Spanish-accented English

Methods

Subjects

24 monolingual speakers of Standard American English age 18-33 (17 female, mean age 20.5, 21 right-handed) participated in this experiment. All subjects were recruited from the Brown University community, and although many had experience with languages other than English, none were fluent in a language other than English during childhood. Participants were asked about their experience with Spanish, Spanish-accented English, experience with other accents, and difficulty understanding accents in general.

Stimuli

The auditory stimuli were various versions of the word "fish". These versions had the same onset, coda, and duration, but differed in the F1 and F2 of the medial vowel. The vowels were situated along a 10-point continuum between /ɪ/ and /i/ or along a ten-point continuum between /ɪ/ and /ɛ/, so that stimuli sounded either like /fiʃ/, /fɪʃ/, or /fɛʃ/. The endpoints of the continuum were produced by a phonologically-trained female speaker in her twenties, and then the stimulus continua were synthesized in Praat

(Boersma & Weenick, 1992, ver. 6.1). The stimuli were verified by members of the Metcalf Infant Research lab to sound equally natural across both continua.

In the experimental trials, each version of fish was paired with all 10 versions of fish along the same continuum (including itself), resulting in 100 pairings per continuum, and 200 pairings overall. Each pair was presented with a 1000 ms ISI between stimuli.

Procedure

Participants completed the study at the Metcalf Infant Research Lab on Brown University's campus. Before starting the experiment, participants completed a language background questionnaire adapted from Northwestern University's Q-LEAP questionnaire. Then, participants were seated in a sound attenuated booth about 2 feet away from a computer monitor. In their lap they held a SuperLab response box with the far left button labeled "First" and the far right button labeled "Second". Participants were told that they would hear two versions of the word "fish" at a time, and they were instructed to listen to both versions and then indicate on the response box whether they thought the first version sounded more like it could have been spoken by someone with a Spanish accent or if the second version sounded more like it could have been spoken by someone with a Spanish accent. The 200 trials were then presented in random order to the participant. After they completed the experiment, participants were asked follow-up questions about their familiarity with other accents of English, Spanish-accented English, the Spanish language (all on a three-point scale, with "none", "some", "a lot"), and their hometown.

Results

Number of times stimuli were chosen as "Spanish-accented"

In order to assess the effect of trial type (authentic vs. inauthentic) on the number of times each stimulus was chosen as "more Spanish-accented", a logistic regression was fit using the glmer function in the lme4 package ver 1.1-21 (Bates et al., 2015) using R statistical software ver 1.1.453 (R Core Team, 2019). The dependent variable was whether a stimulus was chosen as "more likely to be Spanish-accented" (coded as 0 or 1 for each stimulus in each trial), with fixed effects of trial type, stimulus, and an interaction of these two measures, and with subject as a random effect on intercept and on the slope of trial type:

```
choice (yes or no) ~ stimulus step * trial type + (1 + trial type |
subject)
```

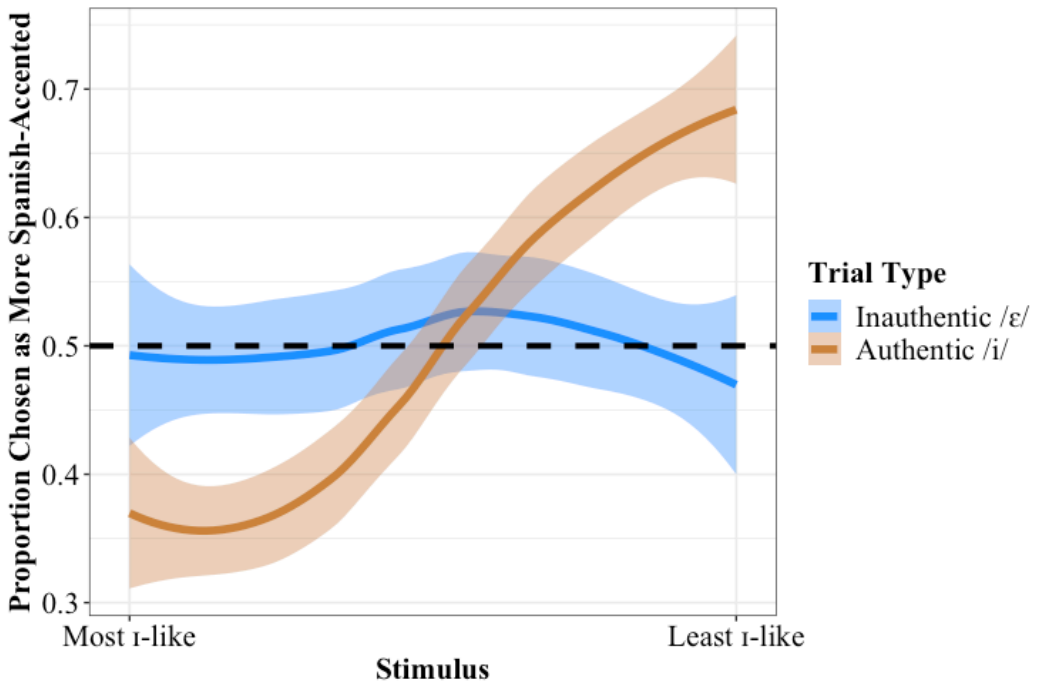

Figure 1: Sensitivity to extant Spanish-English vowel difference. Proportion (by subject) of times stimuli along continua from most /ɪ/-like to least /ɪ/-like were chosen as likely to be Spanish-accented with 95% confidence intervals. Dotted line is chance.

This regression showed a significant interaction effect of trial type and stimulus choice on stimulus choice likelihood ($\chi^2(2) = 275.04$, $p < 0.0001$), suggesting that patterns of stimulus choice were different within authentic and inauthentic Spanish-accented trial type (See Table 1 for estimated fixed effects). Fixed effects also showed that trial type had a significant main effect on the likelihood of stimulus choice ($z = -10.08$, $p < 0.0001$).

Reported experience with Spanish (scale from 0-2), reported experience with Spanish-accented speakers (scale from 0-2) and reported difficulty understanding foreign accents (scale from 0-2) did not have significant effects on the model ($\chi^2(3) = 0.023$, $p = 0.999$). The majority of people reported at least some experience with Spanish and Spanish-accented English (21 and 23 participants, respectively), and most participants reported little difficulty understanding accented speech. This suggests that our American participants have had enough exposure to know that the /ɪ/ > /i/ distinction is authentic to that accent while the /ɪ/ > /ɛ/ distinction is not.

Indeed, participants appear to choose /i/-like stimuli in greater proportions than their /ɪ/-like counterparts, but they treated all stimuli along the /ɪ/-/ɛ/ continuum as equally likely (or unlikely) to be Spanish-accented (Figure 1).

Table 1: Estimated model fixed effects for logistic regression on stimulus step choice by trial type and stimulus for Experiment 1A.

	stimulus step choice			
Effect	**Estimate**	**Std Err**	**z value**	**p**
Intercept	0.177	0.067	2.637	**<0.01 ****
stimulus step	0.004	0.011	0.351	0.726

Table 1: Estimated model fixed effects for logistic regression on stimulus step choice by trial type and stimulus for Experiment 1A.

	stimulus step choice			
Effect	**Estimate**	**Std Err**	**z value**	**p**
trial type	-0.956	0.095	-10.076	**<0.0001** ***
stim step : trial type	0.175	0.016	11.309	**<0.0001** ***

Psychoacoustic measures

Additionally, in order to assess the impact of psychoacoustic measures (F1 and F2 in mel) on participants' performance, the predictive power of F1 and F2 was analyzed (measured at the midpoint of the vowel, converted to mel and scaled) and their interaction on the likelihood of stimulus choice using a logistic regression with random effects of subject on intercept and the slope of trial type. Previous research has shown that F1 and F2 are the most important formants for vowel discrimination (Flege et al., 1997). Indeed, the model including scaled mel F1 and F2 was significantly more predictive of the likelihood of stimulus choice than the null model ($\chi^2(2) = 192.36$, $p < 0.0001$), but compared to this model, the significantly best-fitting model included these fixed effect terms as well as a fixed effect of trial type and an interaction of the three terms ($\chi^2(1) = 11.61$, $p < 0.001$):

```
choice (yes or no) ~ trial type * scaled mel F1 * scaled mel F2 + (1
+ trial type | subject)
```

Table 2: Estimated fixed effects for logistic regression model including trial type, F1, F2, and their interactions. P-values were obtained using sjPlot's tab_model function in R (Lüdecke, 2020).

	stimulus choice			
Effect	**Estimate**	**Std Err**	**z value**	**p**
Intercept	-0.352	0.203	-1.732	0.083 .
trial type	-1.519	0.327	-4.649	**< 0.0001** ***
F1	-7.709	3.628	-2.125	**< 0.05** *
F2	-8.439	3.741	-2.256	**< 0.05** *
trial type : F1	-21.484	6.282	-3.420	**< 0.001**
trial type : F2	-19.292	6.161	-3.131	**< 0.005** **
F1 : F2	0.074	0.152	0.438	0.629
trial type : F1 : F2	-0.961	0.222	-4.321	**< 0.0001** ***

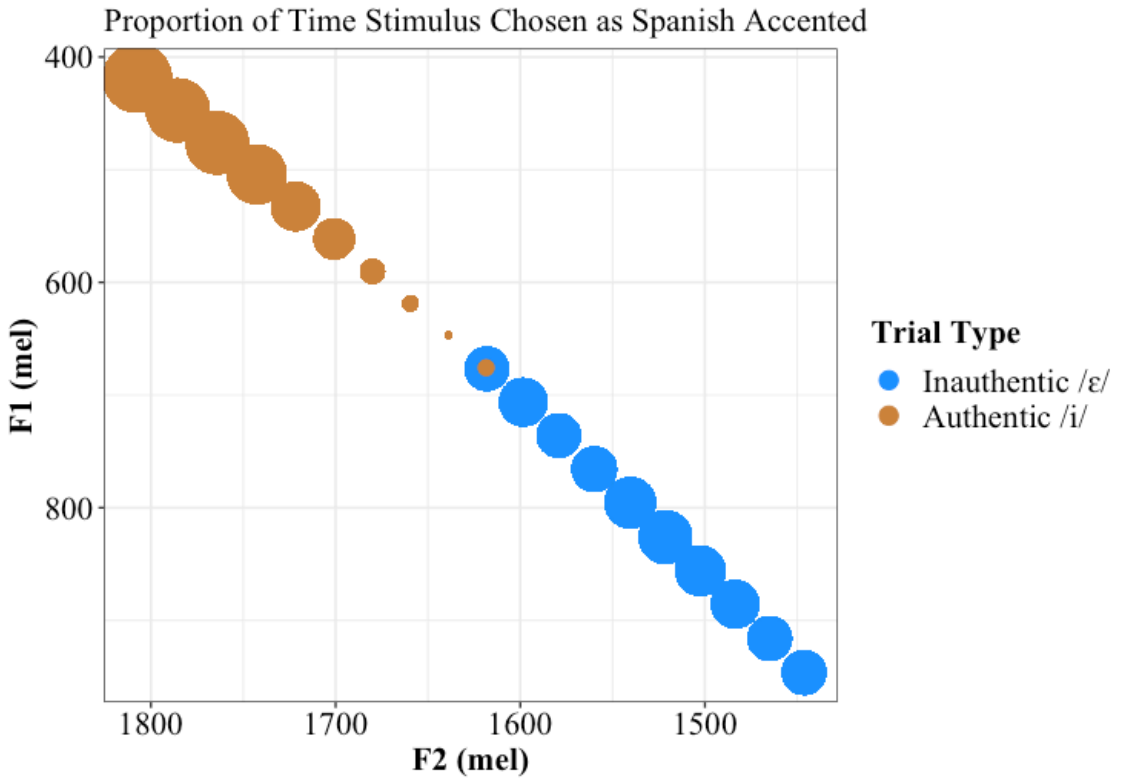

Figure 2: The proportion of time each stimulus was chosen as Spanish-accented, denoted by the size of each dot, as a function of F1 and F2 (mel) of the vowel midpoint with axes reversed to simulate acoustic vowel space.

As F1 and F2 were manipulated in equidistant steps to create 10 stimulus steps along each continuum, F1 and F2 manipulations are fairly analogous to stimulus steps. Interestingly, there was not a main effect of stimulus steps nor an interaction effect of F1 and F2 on either of the models that included these terms. Both stimulus steps and the interaction of F1 and F2 showed strong interaction effects with trial type. Participants were more likely to accept stimuli as they approached /i/, but showed no difference in endorsement along the /ɪ/ > /ɛ/ continuum, providing evidence that participants know that /i/-like stimuli are most Spanish-accented, while /ɪ/ and /ɛ/ stimuli are not (Figure 2).

Reaction time

The effects of stimulus choice and trial type on reaction time were also assessed by fitting a linear regression with reaction time as the dependent variable, with chosen stimulus, trial type, and an interaction of chosen stimulus and trial type as fixed effects, and a random effect of subject on intercept and slope of trial type:

```
reaction time ~ chosen stimulus * trial type + (1 + trial type |
subject)
```

Results of this model showed no main effect of trial type or stimulus choice, but an effect of trial type/stimulus choice interaction ($\chi^2(1) = 10.59$, $p = 0.001$, see Table 2). These results suggest that only within the authentic SAE condition, as stimuli became more /i/-like, reaction times decreased. Within the inauthentic condition, however, reaction time was the same for all stimulus steps and was the same as the reaction times for the most /ɪ/-like stimuli within the authentic condition (Figure 3). Reaction times were overall

higher for stimuli along the /ɪ/ > /ɛ/ continuum than the /ɪ/ > /i/ continuum, suggesting perhaps that participants were having a harder time deciding which stimulus was more Spanish-accented, as neither were likely. Additional analyses showed that Spanish experience, Spanish accent experience, and reported difficulty understanding foreign accents also had no effect on reaction time ($\chi^2(3) = 2.40$, $p = 0.49$), once again providing evidence that even with relatively little exposure to SpE, listeners are able to form detailed representations of the accent's features and its mismatches with SAE.

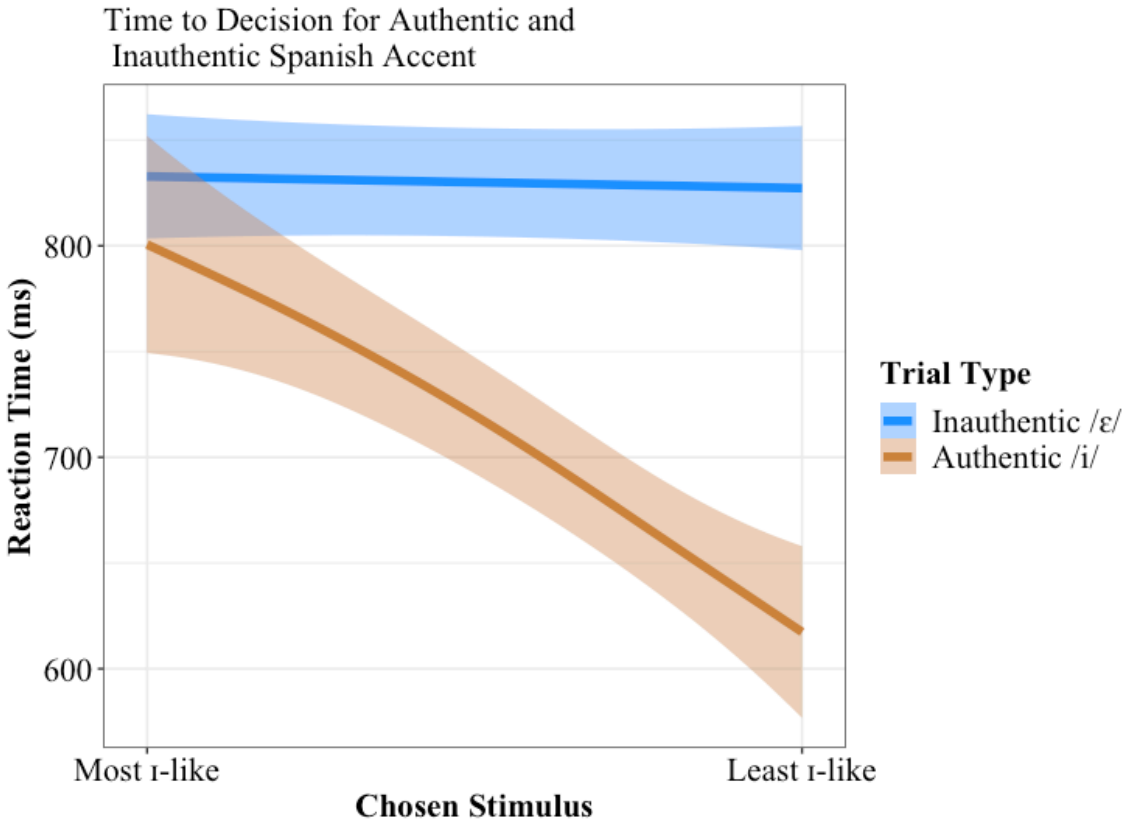

Figure 3: Time from presentation to stimulus choice plotted along the continuum of most /ɪ/-like to least /ɪ/-like chosen stimulus with 95% confidence intervals for trial type. In authentic /ɪ/-/i/ continuum, participant reaction times decreased as a function of SAE authenticity, while in inauthentic /ɪ/-/ɛ/ continuum, there was no change in reaction time across the continuum.

Table 3: Estimated model fixed effects for linear regression on reaction time for Experiment 1A with 95% confidence intervals. P-values were obtained using sjPlot's tab_model function in R (Lüdecke, 2020).

	reaction time				
Effect	**Estimate**	**Std Err**	**t**	**CI**	**p**
Intercept	3151.53	50.93	61.88	3051.70 - 3251.36	**<0.001** ***
stimulus choice	-2.23	4.23	-0.53	-10.52 - 6.07	0.599
trial type	-16.68	44.85	-0.37	-104.58 - 71.22	0.710
stim choice: trial type	-19.17	5.89	-3.26	-30.70 - -7.63	**<0.001** ***

Discussion

Experiment 1A tested the specificity of participants' representations of the /ɪ/ > /i/ feature of Spanish-accented English. Participants heard the word "fish" with vowels along two continua, /ɪ/ > /i/ and /ɪ/ > /ɛ/. Results demonstrated that participants were highly likely to select /i/-like stimuli as more Spanish-accented than /ɪ/ stimuli, but showed no preference between /ɪ/ and /ɛ/ stimuli. This pattern of results suggests that our American participants have somewhat detailed representations of SpE regardless of the amount of exposure to SpE that they reported. At least for this cross-category distinction between SpE and SAE, participants appear to store aspects of SpE in long-term memory.

The results of Experiment 1A provide modest evidence against the *expand* hypothesis, which posits that phonemic category boundaries expand in all directions to accommodate variability, such as that which arises from accented speech. Because participants did not choose /ɛ/ stimuli more than /ɪ/ stimuli, it is likely that with enough experience, listeners do not perceive accented speech as universal variability, but rather

as systematic, directional variability. This interpretation is supported by work showing that after being trained on a single foreign accent, listeners can adapt to new speakers from the same language background, but cannot use that information to adapt to speakers from a different language background (Witteman et al., 2013; Cooper & Bradlow, 2016).

Rather, results of the current experiment provide evidence for the *shift* hypothesis, which states that the boundaries of a phonemic category shift toward new exemplars during speech processing. The results are interpreted in this way because participants did not accept /ɪ/-like stimuli as a good examples of Spanish-accented "fish," even when they were paired with /ɛ/-like stimuli, showing that /fiʃ/ is not considered a good category exemplar in Spanish-accented speech. However, this interpretation does not necessarily rule out the *pathway* hypothesis, which suggests that the category asymmetrically expands to incorporate new exemplars, but lexical access is ultimately guided by a type of gravitational pull that attracts exemplars to a nearby category. This is because the task did not require participants to access "fish" as a lexical item, as participants were specifically asked about their representations of Spanish-accented English, not their representation of "fish." Therefore, while participants retain knowledge about the likelihood of one phonetic distribution over another in a familiar accent, they may not actually shift their phonemic representations during online speech processing.

In order to better understand the level of detail with which participants store familiar accent information, a second experiment (1B) was conducted, in which participants were tested on their knowledge of a phonological rule in RP British English (BE). Native BE speakers produce /ɑ/ where native SAE speakers produce /æ/ before

voiceless fricatives /_θ/, / _ʃ/, and /_f/, but not before voiceless stops at the same place of articulation (Hosseinzadeh, 2015). In other environments, BE speakers also produce /æ/. While it is likely that Americans have some knowledge of the /æ/ > /ɑ/ difference in BE, it is less clear that they would know the rules constraining its production.

Experiment 1B: British English

Methods

Subjects

Participants were recruited using Psiturk (Gureckis & McDonnell, 2016) and Amazon Mechanical Turk. A total of 50 participants completed the experiment, but 25 were excluded from the results because they did not reach the "attentive worker" threshold, which was calculated as the number of times the participants chose the two most /ɑ/-like "bath" stimuli as more British-accented as the two most /æ/-like "bath" stimuli at least 7/8 times (chance = 4, $p = 0.03$). This threshold was chosen as American English speakers are highly likely to recognize this distinction between British and American English. This left a total of 25 participants (8 female, 14 male, 3 no response; average age approx. 42 years). Three of these participants were early bilinguals, but analyses within a maximal generalized linear model showed no effect of bilingualism on the results, so these participants were included in analyses.

Stimuli

Authentic British English utterances of the words “bath” (/bɑθ/) and “bat” (/bæt/) were recorded by a female native speaker of RP British English. These words were then used to synthesize two 10-point continua in Praat (Boersma & Weenick, 1992, ver 6.1), one of the word “bath” and one of the word “bat”. The vowels of both continua ranged from /ɑ/ to /æ/, so that stimuli sounded like /bɑθ/ or /bæθ/ and /bɑt/ or /bæt/. The onset, coda, and duration were equal for items along each continuum (duration was the duration of the original recording). The stimuli were verified by members of the Metcalf Infant Research Lab to sound equally natural across both continua. In each continuum, all ten stimulus steps were paired with each of the other nine steps within the continuum, resulting in 90 stimulus pairs per continuum and a total of 180 trials.

Procedure

In order to ensure that Mechanical Turk participants were using headphones and were in relatively quiet listening conditions, participants first had to pass a headphone screening test in which listeners were asked to judge which of three pure tones was quietest, with one tone presented 180 degrees out of phase across the left and right headphones (Woods et al., 2017). Participants had two chances to pass the headphone screening, and after the second failure they were excluded from participation. This headphone screening test also provided a way to deter bots from taking participant slots on MTurk. After passing the headphone screening, participants were told that they would hear two versions of a word and then would indicate which of the two sounded more

"British Accented" using the 1 and 2 keys on their keyboard. They were asked to respond as quickly and accurately as possible. They heard the stimulus pairs presented in random order with a 1000 ms ISI between first and second stimulus and with trials of "bath" and "bat" intermixed. For each trial (including the practice trials), a clipart image of the object referred to (e.g., a bathtub or a baseball bat) was shown in order to ensure some degree of lexical access (see appendix for images). After hearing all 180 trials, participants completed a follow-up questionnaire asking about their experience with British English and their ease of understanding for British accents, other native accents of English, and foreign accents of English, all on a scale from 0-10. They were also asked additional questions about their language background and demographic background.

Results

Number of times stimuli were chosen as "British-accented"

To assess whether participants knew that the /æ/-/ɑ/ distinction between American and British English occurs before voiceless fricatives (but not before voiceless stops), a logistic regression was fit using the glmer function in the lme4 package ver 1.1-21 (Bates et al., 2015) using R statistical software ver 1.1.453 (R Core Team, 2019). As in the analysis on Spanish-accented English, the dependent variable was whether a stimulus was chosen as "more likely to be British-accented" (coded as 0 or 1 for each stimulus in each trial), with fixed effects of word, stimulus step from most authentic to least authentic, and an interaction of these two measures, and with subject as a random effect on intercept and on the slope of word:

```
choice (yes or no) ~ stimulus step * word + (1 + word | subject)
```

This model was a significantly better fit than the null model ($\chi^2(3) = 1537.1$, $p < 0.0001$), and fixed effects estimates revealed a main effect of stimulus, a main effect of word, and an interaction of stimulus and word (Table 4). These results suggest that participants knew of a general /æ/ > /ɑ/ difference between American and British English, and therefore correctly identified /bɑθ/ as British-accented and /bæθ/ as *not* British-accented. However, the pattern of results for "bat" reveals that American listeners' representations of this vowel difference in British English is not fully specified, as they tended to choose /bɑt/ as more British-accented than /bæt/. It is of note, however, that confidence

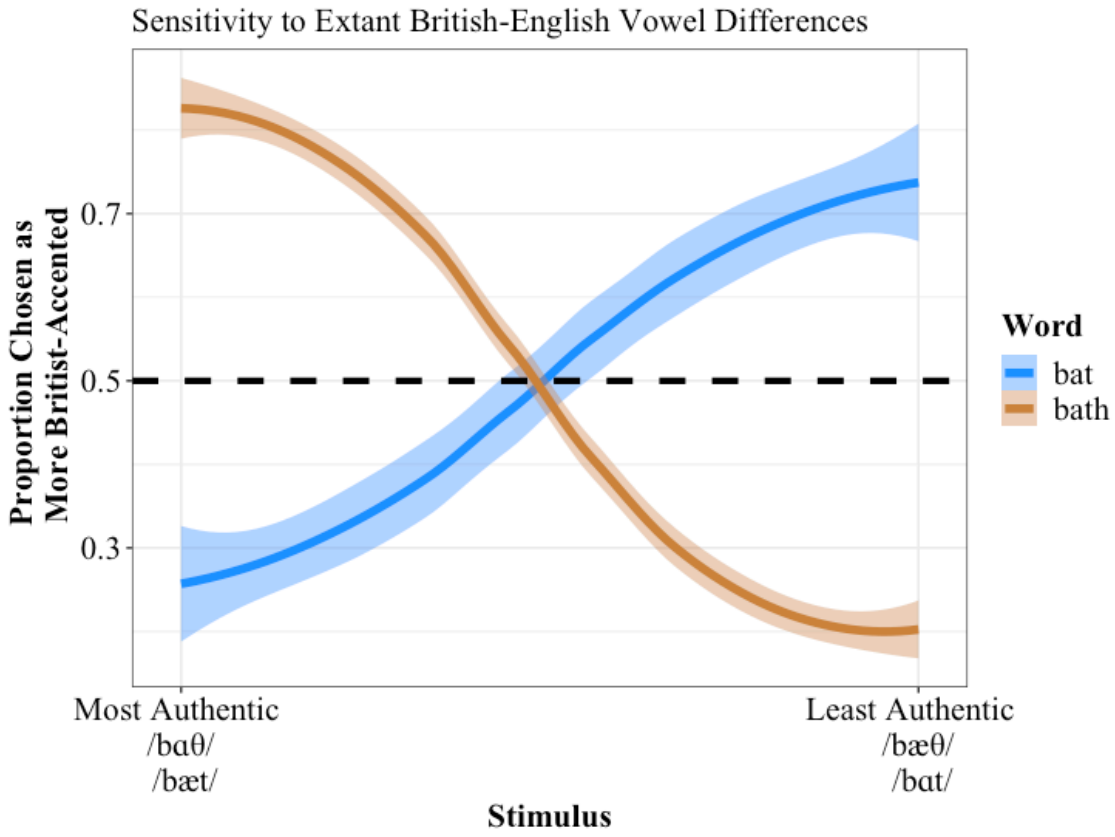

Figure 4: Proportion of times stimuli were chosen as "more British-accented" from most to least authentic (for bath: /bɑθ/ > /bæθ/, for bat: /bæt/ > /bɑt/), with 95% confidence intervals. Participants correctly identified British "bath" but not British "bat". Dotted line is chance.

intervals for proportions along the "bat" continuum are larger than those of the "bath" continuum, suggesting that participants may have been less sure about /bɑt/ than they were about /bɑθ/ (Figure 4).

Table 4: Estimated fixed effects for British English logistic regression model including stimulus step, word, and their interaction.

	BE stimulus choice			
Effect	**Estimate**	**Std Err**	**z value**	**p**
Intercept	-1.380	0.075	-18.43	**< 0.0001** ***
stimulus step	0.249	0.012	20.57	**< 0.0001** ***
word	3.581	0.113	31.74	**< 0.0001** ***
stimulus step : word	-0.650	0.018	-35.20	**< 0.0001** ***

As in the Spanish-accented English analysis, participants' reports of familiarity with British English (on a scale from 0-10, with 0 corresponding to "not familiar at all" and 10 corresponding to "extremely familiar", mean = 4-55, std err = 0.48) and ease of understanding other native accents of English (on a scale from 0-10, with 0 corresponding to "very easy" and 10 corresponding to "very difficult", mean = 5.05, std err = 0.44) were also analyzed. When these fixed effect terms were added to the model reported above, they did not significantly improve model fit ($\chi^2(2) = 0.018$, $p = 0.991$). Therefore, it appears that even American participants who reported knowing little about British accents had enough exposure to recognize that in certain instances where Americans say /æ/, British speakers say /ɑ/, while those who reported knowing a lot about British accents did not reliably know that this distinction occurs before voiceless fricatives but not before voiceless stops.

Represented in psychoacoustic space

Instead of stimulus steps, the effects of psychoacoustic measurements on stimulus choice were also analyzed by taking the average of F1 and F2 in Hz for each stimulus in both continua, converted them to Mel (a psychoacoustic scale), and then scaled for regression. A model was then fit with stimulus choice (yes or no) for each trial as the dependent variable, scaled Mel F1 and F2 and word and interactions of all three terms as fixed effects, and random effects of subject on the intercept and the slope by word. This model was a significantly better fit than the null model ($\chi^2(7) = 1549.9$, $p < 0.0001$), and fixed

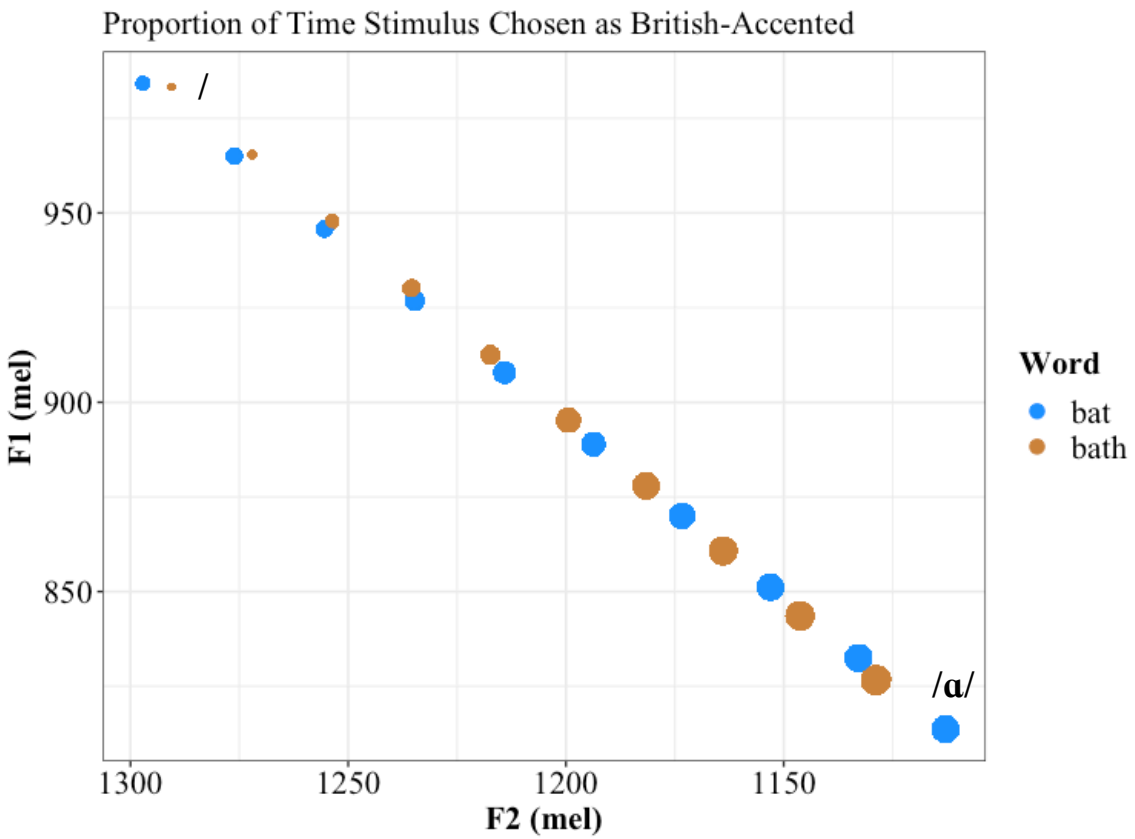

Figure 5: Proportion of time each stimulus was chosen as "more British-accented" (denoted by size of dot) plotted along psychoacoustic dimensions (reversed to mirror vowel space).

effects revealed all terms were significant except the interaction of F1 and F2, which was marginal (Table 5). The main effects and interaction effects of word show that although participants did choose /ɑ/-like stimuli as more British-accented overall, they also treated "bath" and "bat" stimuli differently. As such, participants may have more detailed representations of these lexical items or of the in-question phonological rule than would seem at first glance.

Table 5: Estimated fixed effects for British English logistic regression model including stimulus step, word, and their interaction. P-values were obtained using sjPlot's tab_model function in R (Lüdecke, 2020).

	BE stimulus choice			
Effect	**Estimate**	**Std Err**	**z value**	**p**
Intercept	0.418	0.102	4.112	**< 0.0001** ***
word	-2.500	0.161	-15.560	**< 0.0001** ***
scaled F1	38.037	8.162	4.660	**< 0.0001** ***
scaled F2	-37.906	7.989	-4.745	**< 0.0001** ***
word : F1	109.228	13.155	8.303	**< 0.0001** ***
word : F2	-114.289	13.21	-8.649	**< 0.0001** ***
F1 : F2	0.075	0.042	1.778	0.075 .
word : F1 : F2	0.396	0.069	5.703	**< 0.0001** ***

Reaction time

To further understand participants' representations of the /æ/-/ɑ/ distinction between American and British English, reaction time as a function of word and stimulus step was analyzed using a linear regression, with reaction time as the dependent variable, with

chosen stimulus, word, and an interaction of chosen stimulus and word as fixed effects, and a random effect of subject on intercept and slope of word:

```
reaction time ~ chosen stimulus * word + (1 + word | subject)
```

This model fit the data significantly better than the null model, models that only one of the two terms, and a model that included both terms but no interaction ($\chi^2(1) = 169.209$, $p < 0.0001$). Furthermore, this model revealed main effects of chosen stimulus and reaction time in addition to an effect of their interaction (Table 6).

Reaction time patterns are overall similar to stimulus choice patterns reported above. However, while accurate-to-BE chosen stimuli showed a large reaction time difference for "bath" and "bat", inaccurate-to-BE chosen stimuli did not.

Table 6: Estimated model fixed effects for linear regression on reaction time for British English experiment with 95% confidence intervals. P-values were obtained using sjPlot's tab_model function in R (Lüdecke, 2020).

	reaction time				
Effect	**Estimate**	**Std Err**	**t**	**CI**	**p**
Intercept	0.552	0.027	20.27	0.50 - 0.61	**<0.001** ***
stimulus choice	-0.012	0.002	-7.849	-0.01 - -0.01	**<0.001** ***
word	-0.219	0.016	-13.343	-0.25 - -0.19	**<0.001** ***
stimulus choice: word	0.028	0.002	13.101	0.02 - 0.03	**<0.001** ***

For example, for stimuli in step 1 (most authentic), a linear regression revealed a highly significant difference in reaction times between "bath" and "bat" stimuli ($t = -6.813$, $p < 0.001$), while for stimuli in step 10 (least authentic), there was no significant difference in reaction time between "bath" and "bat" stimuli ($t = 0.723$, $p = 0.470$). Therefore, while stimulus choice data shows that overall participants thought that both /bɑθ/ (authentic)

and /bɑt/ (inauthentic) were both better examples of British-accented speech than /bæθ/ (inauthentic) and /bæt/ (authentic), reaction time data suggests that participants were less sure about "bat" stimuli, resulting in overall longer reaction times.

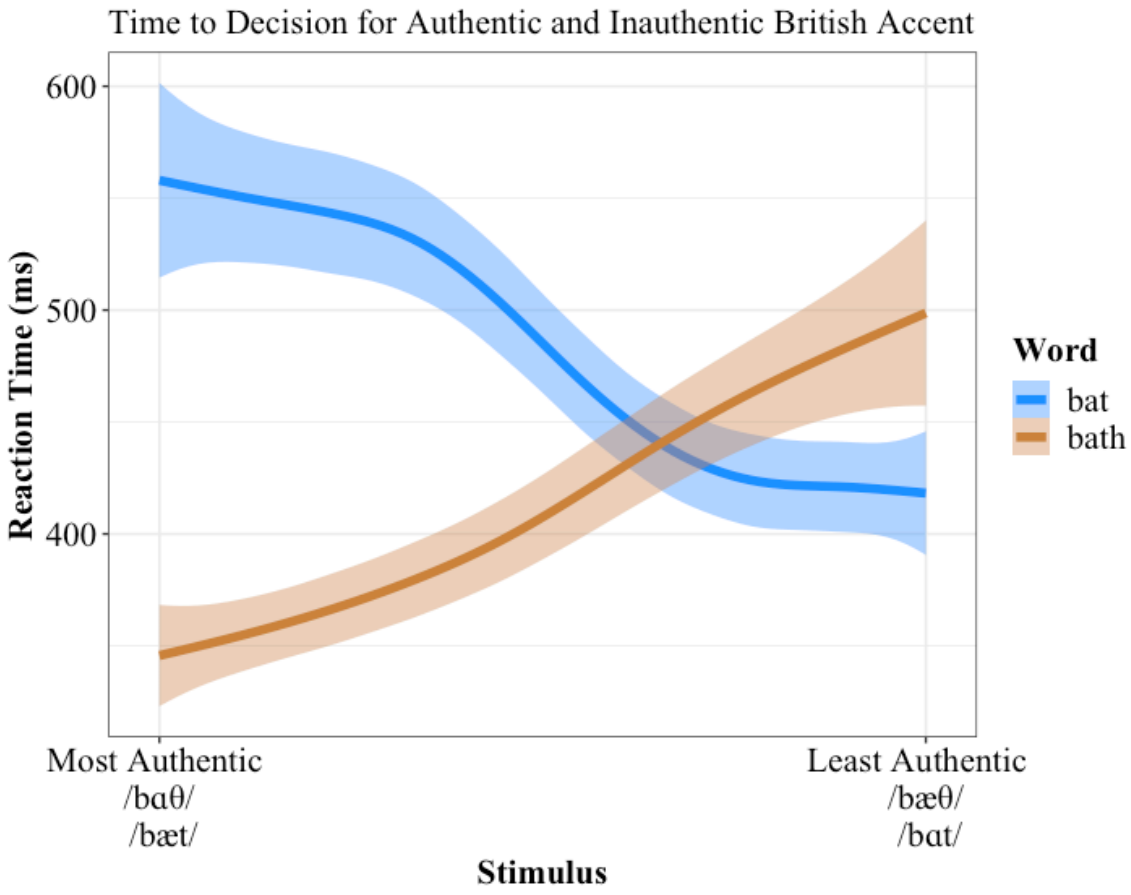

Figure 6: Reaction time in ms for stimuli from most authentic to least authentic in RP British English with 95% confidence intervals.

Discussion

The goal of experiment 1B was to test the level of detail with which American listeners represent RP British-accented English. Specifically, participants were tested on

a phonological rule difference between SAE and BE. As in experiment 1A, results of experiment 1B demonstrated that aspects of familiar accents are stored in long-term memory, again regardless of reported experience with that accent. However, although participants correctly chose /bɑθ/ as more British-accented than /bæθ/, they also incorrectly chose /bɑt/ as more British-accented than /bæt/. This pattern provides evidence for storage of a general phonological mismatch, but not of an environmentally constrained phonological rule. Previous work on perceptual adaptation has demonstrated that, with the same amount of exposure, participants could reliably learn new phoneme boundaries for an ambiguous /s/ when it was contextually independent (i.e., intervocalically), but not when it was contextually constrained (i.e., only preceding [_tr]) (Kraljic, Brennan, & Samuel, 2008). It may be the case that listeners are sensitive to shifts such as /æ/ > /ɑ/ in BE, but without extensive experience, will generalize them to inauthentic environments.

Experiment 1B therefore provides support for accounts of pre-lexical accent adaptation, as participants showed less accurate responses in the "bat" condition than would be expected if they were processing accents mainly at the lexical level. Conversely, greater variability and longer reaction times in the "bat" condition also suggest that participants were not confident in their generalizations of /æ/ > /ɑ/. This is especially clear in reaction times for "bat" stimuli, which are significantly longer than "bath" stimuli overall, but it is also suggested in the results of the psychoacoustic analysis that demonstrated an effect of word above and beyond the acoustic properties of the stimuli.

It is possible that this behavior stems either from a knowledge of BE's phonological rule or a lack of acoustic trace for /bɑt/ at the lexical level of processing, but unfortunately it is not possible to differentiate between these two possibilities in the current study. If participants have knowledge about the phonological rule, this knowledge was not weighted heavily enough for participants to correctly choose /bæt/-like stimuli, but it did disrupt processing to some extent. Findings of Sumner and Samuel (2005), which demonstrated equivalent priming effects for three variants of final /t/ in American English, support an interpretation involving pre-lexical processing of regular (rule-based) variant forms. However, Sumner and Samuel's (2005) work critically differs from our own in a few ways. Sumner and Samuel examined consonant variation that exists within American participants' native variety of English, while the current experiments examine vowel variation that stems from accent varieties that are different from participants' native variety. Because vowels and consonants appear to play different roles in speech processes, it is not possible to assume that listeners use the same mechanisms to process variant consonants and vowels (although Kronrod et al. (2016) have developed a parsimonious account of vowel and consonant perception through Bayesian modeling). As a result, the question of how listeners store and access information about accented vowels remains open.

If the effects of *word* that were observed in the current study actually come from acoustic traces of BE-accented items in the lexicon rather than from pre-lexical phonological processes, participants' ultimate decisions about accentedness suggest that pre-lexical processing of phonological forms pits these traces and generalizations of

phonological mismatches against each other, resulting in slower processing but eventual generalization of the /æ/ > /ɑ/ contrast. One caveat to this interpretation is that although pictures of the lexical item were included in order to guide lexical access, it's possible, as it was in experiment 1A, that participants did not fully access lexical items, as the task demands do not distinctly require participants to do so. Therefore, even if participants store acoustic traces of BE instances of "bat," it may be that those traces weren't sufficiently activated in this experimental task, as participants ultimately chose /bɑt/ as more British accented.

Like Experiment 1A, Results of experiment 1B also appear to support the *shift* hypothesis over the *pathway* hypothesis at the level of phonological representation, as participants appeared to shift their /æ/ category toward /ɑ/, irrespective of context (note that the design of this experiment did not allow for the evaluation of the *expand* hypothesis). These findings call into question the effect of saliency on storage of accented forms—when listeners hear /ɑ/ when expecting /æ/, they are more likely to notice that there is a difference than when their expectations are met. Extending this notion, it may be that the changing shape of phonological representations may be modulated by the degree of saliency—noticeable changes are assigned greater weight, and more salient differences are more reliably stored. To our knowledge, the effects of saliency of mismatch on long-term storage has not yet been explicitly tested.

Through experiments 1A and 1B, I demonstrated that listeners accurately remember features of familiar accents and are able to access that knowledge later. However, accent representations are not fully specified or detailed, but rather involve

generalizations mainly at a pre-lexical phonological level, as shown in experiment 1B. Although learning may initially be lexically driven, variation arising from a familiar accent appears to be used to adjust how vowels are mapped onto the existing lexicon (Witteman et al., 2015; Cutler, Smits, & Cooper, 2005; McQueen, Cutler, et al., 2006; Sjerps & McQueen, 2010). Additionally, the current results suggest that how listeners encode variability may be modulated by the saliency of a mismatch, such that the details of an accent that are acquired most strongly are also the most salient. In terms of processing accented speech, this mechanism would likely allow listeners to overcome the most challenging mismatches first, allowing for rapid adaptation. In this case, phonological features of an accent that overlap nearly or completely with one's native variety (e.g., /æ/ in "bat"), are likely not encoded as accented features and therefore not recalled by listeners in tasks like ours that specifically ask about accent authenticity.

Results from these two experiments align with accounts of the Natural Referent Vowel (NRV) framework, which suggest that directional asymmetries occur in vowel discrimination that favor vowels closer to the periphery of F1/F2 vowel space (Polka & Bohn, 2011; Masapollo et al., 2018). In both experiments, the most endorsed vowels were also the more peripheral(/i/, /ɑ/). It is therefore impossible rule out that NRV asymmetries could have had an effect on the current results. However, there is evidence that listeners are also sensitive to accent shifts from more-to-less peripheral vowels. For example, Witteman et al. (2015) tested Dutch speakers on priming effects of words with /i:/ (more peripheral) and /ɪ/ (less peripheral), which are both realized as /ɪ/ in Hebrew-accented Dutch. They found that listeners were able to maintain representations for /ɪ/

words while simultaneously extending /i:/ words to include /ɪ/ exemplars (Witteman et al., 2015). Therefore, it is unlikely that the current results stem solely from NRV effects.

The current study is also unable to differentiate between lexical and pre-lexical processing. Although steps were taken to increase the likelihood that participants would access lexical representations during the current task, the question "Which is more X-accented?" does not explicitly require lexical information. Therefore, while listeners *can* use lexical information during online accent adaptation (e.g., Cooper & Bradlow, 2016), it is possible that they did not do so in the current experiment, thereby demonstrating only the effects of pre-lexical phonological representations.

Conclusion

The current study investigated whether Americans store detailed information about Spanish-accented and British-accented English. Experiment 1A showed that American listeners have a detailed representation of the /ɪ/ > /i/ feature of Spanish-accented English—when faced with Spanish-accented speech, listeners do not appear to rely expansion of categories to increase flexibility, but rather use details to build more precise categories. Listeners appear to shift their category representations toward the /i/ exemplars, excluding standard American English /ɪ/ exemplars: the current experiment provides some support for the shift model of phonological adaptation, but more research needs to be done on this front. This finding was not modulated by reported experience with Spanish-accented English, suggesting that listeners don't need extensive experience with an accent to form some representation of it. Experiment 1B provided further

evidence for the *shift* model, showing that American listeners are also aware of the /æ/ > /ɑ/ mismatch in RP British-accented English, which occurs before voiceless fricatives /_θ/, /_ʃ/, and /_f/. Critically, however, participants extended this mismatch to environments where it does not occur, specifically before a voiceless stop. These results show that representations of accents are not fully detailed, but rather may only be encoded when there is a salient mismatch. Finally, these experiments suggest that at least some of accent adaptation takes place at the pre-lexical level, as phonological mismatches were generalized across environments.

Chapter 3: Rapid Adaptation to a Novel Accent: Lexical Decision

Introduction

Experiment 1 showed that listeners retain knowledge about familiar accents, but that this knowledge is not always detailed, suggesting that listeners may also use rapid adaptation strategies when interacting with speakers whose accents differ from their own. Indeed, there is ample evidence showing that listeners do not need much exposure to an unfamiliar accent in order to adapt to it: for example, Evans and Iverson (2004) presented listeners from northern and southern England with a 2-minute story spoken in either a familiar or unfamiliar accent and then asked them to categorize synthesized vowels in carrier words. They found that listeners adjusted their vowel categorizations according to the accent used in the 2-minute story, providing evidence that listeners need only a little bit of experience with an accent to adjust their phonological system, at least temporarily. Similarly, Clarke & Garrett (2004) showed that with less than one minute of exposure to Spanish- or Chinese-accented English, English-speaking listeners were able to recognize words spoken in these accents as quickly and accurately as non-accented English words, and Cooper and Bradlow (2016) demonstrated that participants were able to adapt to Mandarin-accented English presented in noise when using semantic context to bootstrap word recognition.

Although many studies of accent adaptation have been done with real-world accents, rapid adaptation has also been shown with lab-created synthetic accents. Maye, Aslin, and Tanenhaus (2008) investigated how listeners adjust to the specificity of a novel accent and the subsequent effects of this exposure on lexical access. In Maye et al.,

participants first listened to a 20-minute passage of speech produced in Standard American English and then completed a lexical decision task in which they had to indicate which words were real words of English. On a subsequent day, participants head the same story with a novel accent that was created by lowering all of the front vowels by one category in American English vowel space such that the phrase "the wicked witch of the West" was heard as "the weckud wetch of the Wast" (/ðʌ wɛkəd wɛt͡ʃ ʌv ðʌ wæst/). Participants then completed a second, identical lexical decision task. Listeners were able to generalize their knowledge of the shifted vowels to words they had not heard in the passage: when the word in the lexical decision task included a lowered front vowel, but not when the word included a shifted back vowel, participants were able to identify the word.

Maye et al. replicated these findings with raised front vowels, showing that participants were paying attention to the direction of shifted vowels in the accent they heard. Endorsement of unaccented words did not change across lexical decision tasks. These findings provide evidence for a model of accent adaptation in which listeners create a "pathway" between their existing lexical representations and an accented version of the same lexical item. In this way, they are restricting the representational space to include only category shifts that match the accent they have been hearing, but not shifting their representational space such that it might cause them to reject tokens in a highly familiar (and likely their native) dialect. Maye et al. also found that increased endorsement rates extended to words that participants did not hear in the story, suggesting

that participants can shift sound-word mappings across the entire lexicon in the face of accented speech, generalizing beyond heard lexical exemplars.

The current study seeks to replicate and extend the work of Maye et al. Experiment 2A considers a between-subject approach, testing the difference in endorsement rates across groups that heard the a story in either a Standard American English (SAE) accent, an accent in which all front vowels were lowered by one SAE vowel category, or an accent in which all front vowels were raised by one SAE vowel category. Experiment 2B employs a within-subject approach in which participants complete a lexical decision task, listen to the story in one of the three accents, and then complete another lexical decision task with different words. These experiments test how phonological representations of lexical items change with exposure to a novel accent, and whether those changes are comparable to those found in Maye et al.

Experiment 2A - Between subjects measures

Methods

Subjects

Participants were recruited using Psiturk (Gureckis & McDonnell, 2016) on Amazon Mechanical Turk. A total of 250 participants completed the experimental tasks, but 149 of these participants were excluded from the analysis because they did not reach an "attentive worker" threshold, which measured if participants' d-prime scores were at

least 1.5. D-prime was calculated by subject using the formula described in Huang and Ferreira (2020):

$$d' = z(H) - z(FA)$$

where H is a "hit" response (participants endorsed standard words), FA is a "false alarm" response (participants endorsed filler nonce words), and z(x) is the z-score of x. D-prime uses false alarm rates to obtain a measure of participants' bias to identify items as words, and larger d-prime scores indicate a greater difference in treatment of SAE words and nonce words. An additional 10 participants were excluded for failing to correctly answer at least 3 out of 4 follow up questions about the Alice in Wonderland excerpt that they were asked to listen to. This left 91 participants: 36 in the Standard American English group, 23 participants in the raised group, and 32 participants in the lowered group (mean age 41.3, 31 female, 56 male, 1 nonbinary, 3 preferred not to answer).

Stimuli

All auditory stimuli were synthesized using Amazon Polly's Neural Joanna voice. A summary of 5 chapters of Alice in Wonderland was transcribed into IPA (International Phonetic Alphabet) for the Standard American English Accent and then the front vowels were subsequently raised or lowered (see Table 1 for patterns of raising and lowering) in order to create a Raised Accent condition and a Lowered accent condition. These transcriptions were synthesized in Amazon Polly to create an auditory story about 10 minutes long.

In order to create test stimuli for the lexical decision portion of the task, 6 words containing each vowel were selected. Half of the words were selected from the Alice in Wonderland passage (i.e., trained words) and the other half were not from the passage (i.e., untrained words). Within these groups, the vowels all had three representative items for each of the three conditions (Standard, Raised, and Lowered), with the exception of /i/ and /a/, which had only standard and the condition in which they were changed. Words were assigned to conditions such that they would be considered a nonword in SAE (e.g. "race" was raised to become /rɪs/). Raised and lowered versions of words containing back vowels were also created using the same methods in order to test if and how much participants generalize front vowel alternations to back vowels. 33 nonword fillers were also included (3 for each vowel). This resulted a total of 207 items (see Table 2 for breakdown of items, or see Appendix for full list of test items).

Table 1: Pattern of Raising and Lowering for Accent Synthesis

Standard American English Accent	Raised Accent	Lowered Accent
i	i	ɪ
ɪ	i	eɪ
eɪ	ɪ	ɛ
ɛ	eɪ	æ
æ	ɛ	a
a	æ	a

Table 2: Breakdown of test items per test condition

Trained/ Untrained	Accent Condition	Front/Back	Total number of test items	Vowels included
Trained	Standard	Front	18	i,ɪ,eɪ,ɛ,æ,a
Trained	Standard	Back	15	u,ʊ,o,ʌ,ɔ
Trained	Raised	Front	15	i,ɪ,eɪ,ɛ,æ
Trained	Raised	Back	12	u,ʊ,o,ʌ
Trained	Lowered	Front	15	ɪ,eɪ,ɛ,æ,a
Trained	Lowered	Back	12	ʊ,o,ʌ,ɔ
Untrained	Standard	Front	18	i,ɪ,eɪ,ɛ,æ,a
Untrained	Standard	Back	15	u,ʊ,o,ʌ,ɔ
Untrained	Raised	Front	15	i,ɪ,eɪ,ɛ,æ
Untrained	Raised	Back	12	u,ʊ,o,ʌ
Untrained	Lowered	Front	15	ɪ,eɪ,ɛ,æ,a
Untrained	Lowered	Back	12	ʊ,o,ʌ,ɔ
Non-word Filler	Non-word Filler	Front	18	i,ɪ,eɪ,ɛ,æ,a
Non-word Filler	Non-word Filler	Back	15	u,ʊ,o,ʌ,ɔ

Procedure

To ensure that Mechanical Turk participants were using headphones and were in relatively quiet listening conditions, participants first had to pass a headphone screening test in which listeners were asked to judge which of three pure tones was quietest, with one tone presented 180 degrees out of phase across the left and right headphones (Woods et al., 2017). The headphone task had a total of six trials. Participants had two chances to pass the headphone screening, and after the second failure they were excluded from participation. This headphone screening test also provided a way to deter bots from

taking participant slots on MTurk. After passing the headphone screening, participants were instructed that they would first listen to a story about Alice in Wonderland. They were asked to pay close attention, as they would be quizzed on the content of the story later. Participants heard either the SAE version, the Raised Accent version, or the Lowered Accent version of the story. While the story was playing, images of Alice in Wonderland illustrations were shown in order to help hold participants attention throughout the story. After the story ended, participants were asked to complete a lexical decision task, in which they would hear words and had to indicate if words were real words of English or not using the 'z' and '/' keys on the keyboard. Choices were counterbalanced so that participants either used 'z' to denote "Real word" and '/' to denote "Non-word" or 'z' to denote "Non-word" and '/' to denote "Real word". Participants heard all of the test words during this section of the experiment. Once participants completed the lexical decision task, they completed a follow-up questionnaire, which included a quiz on the Alice in Wonderland passage, questions about experience with accents, questions about their language background, and general demographic questions.

Results

Effects on Endorsement Rates

To investigate participants' word endorsement behavior, a logistic regression was fit using the glmer function in lme4 package ver 1.1-21 (Bates et al., 2015) using R

statistical software ver 1.1.453 (R Core Team, 2019). Comparison of nested models showed that the best fitting model for word endorsement as the binary dependent variable included fixed effects of LDT accent type, training (whether the word was presented in the story or not), and word frequency, with a random intercept of subject, a random slope of LDT accent by subject, and a random intercept of lexical item ($\chi^2(5) = 170.84$, $p < 0.0001$, log-likelihood = - 5647.2, see Table 3 for fixed effects). Word frequency was included in the model because the trained words in this experiment tended to be more frequent than untrained words ($R^2 = -0.31$). *Story condition did not improve model fit ($\chi^2(2) = 0.250$, $p = .883$).*

Fixed effects of this model revealed a significant effects of LDT accent and training and a marginal effect of word frequency (see Table 3).

Table 3: Fixed effects for logistic regression of endorsement rate.

	Estimate	Std. Error	z value	p-value
(Intercept)	**4.3423**	**0.3292**	**13.192**	**<0.001 *****
raised LDT	**-4.0455**	**0.3969**	**-10.192**	**<0.001 *****
lowered LDT	**-4.8705**	**0.3972**	**-12.261**	**<0.001 *****
filler LDT	**-6.4932**	**0.4972**	**-13.060**	**<0.001 *****
word frequency	0.3027	0.1572	1.926	0.086 .
training	**-0.8736**	**0.3296**	**-2.651**	**<0.01 ****

Overall, accented LDT words were endorsed less than standard words and trained words were endorsed more than untrained words (see Fig. 1).

Tukey HSD tests for pairwise comparison within LDT accent type and front/back vowels showed that the majority of comparisons between trained and untrained words

within story condition were significant (see Appendix for full results). These, along with the fixed effects of the logistic regression reported in Table 3 suggest that hearing the word at least once during the story resulted in a priming effect, which resulted in increased endorsement. However, none of the story accent conditions had significantly different endorsement rates when grouped by trained vs. untrained words and LDT accent type. This suggests that accented stories primed words to the same degree as the SAE story.

Reaction Time

Reaction time (RT) was also analyzed. RT outliers more than two standard deviations from the mean were removed from the analysis, so that the minimum RT was 0.001 seconds after the offset of the test word and the maximum was 1.133 seconds after the offset of the test word. Comparison of nested linear models revealed that the best fitting model for RT included LDT accent type and trained/untrained words as fixed effects, with random intercepts of subject and word and random slopes of LDT accent type on subject and story accent condition on word ($\chi^2(4) = 91.92$, $p < 0.0001$), see Fig. 2 and Table 4.

```
RT ~ LDT accent + trained + (1 + LDT accent | subject) + (1 +
story accent | word)
```

Table 4: Fixed effects for RT model that includes LDT accent type and trained/untrained as predictors.

	Estimates	CI	p-values
(Intercept)	**0.34**	**0.31 – 0.38**	**<0.001 *****

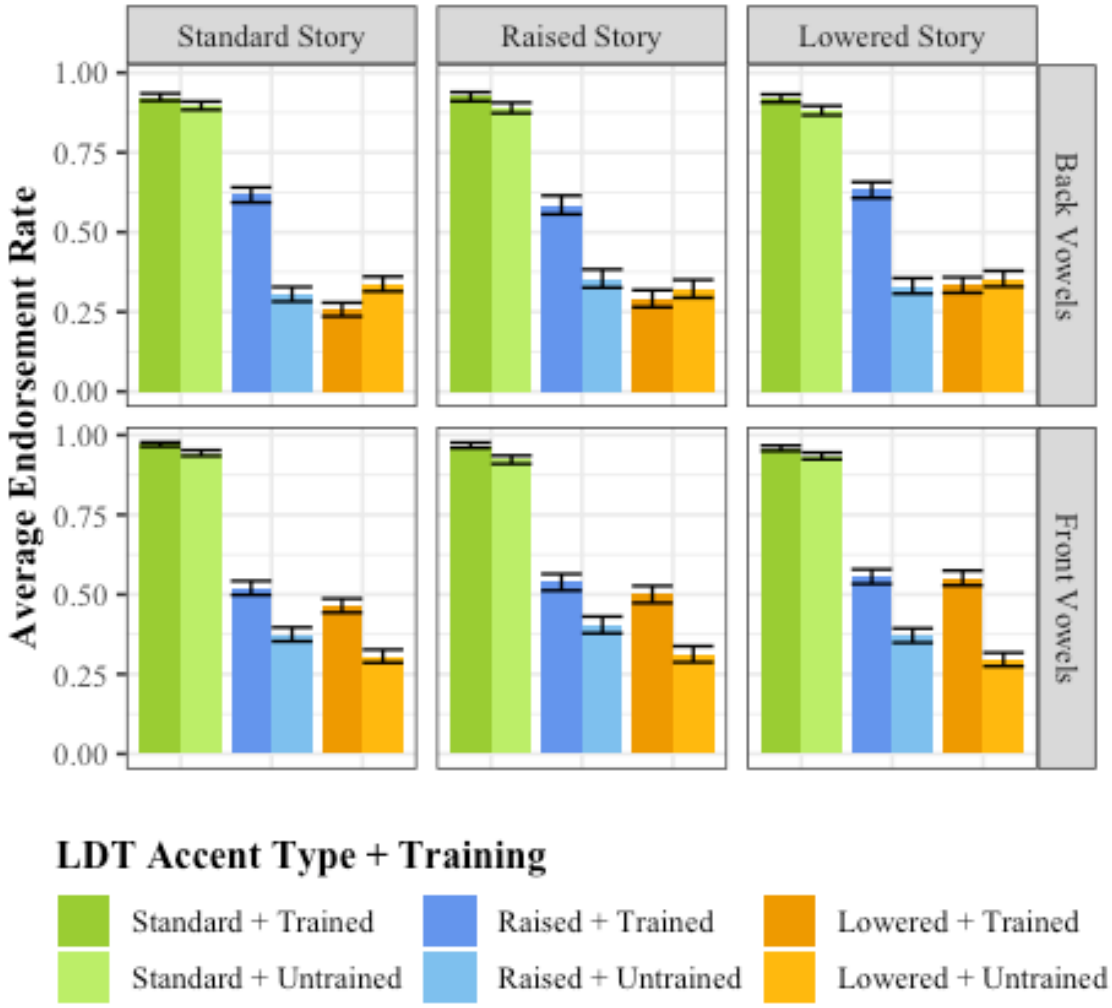

Figure 1: Average endorsement rate by story condition, LDT accent type, training, and front/back vowel classification, with standard errors.

Table 4: Fixed effects for RT model that includes LDT accent type and trained/untrained as predictors.

	Estimates	CI	p-values
Raised LDT	**0.09**	**0.06 – 0.12**	**<0.001 *****
Lowered LDT	**0.13**	**0.10 – 0.16**	**<0.001 *****
Nonce LDT	**0.18**	**0.14 – 0.22**	**<0.001 *****
Trained/Untrained	**0.06**	**0.03 – 0.08**	**<0.001 *****

RTs were significantly faster for standard LDT accent type than for all other LDT types ($p < 0.0001$ for all comparisons). RT was also faster for raised LDT words than lowered and filler words ($p < 0.0001$ for both), but lowered and filler words were only marginally significantly different from each other. ($p=0.054$), suggesting that participants processed raised and lowered words differently, and that lowered words were more difficult to recognize than raised words.

Psychoacoustic Difference

While the paradigm employed in this experiment explicitly uses vowel phoneme categories, there is considerable evidence to suggest that vowels are perceived continuously by listeners (Fry et al. 1962; Stevens, 1989; Chistovich, 1960; Franklin et al, in preparation; and others). Therefore, differences in endorsement rates across LDT conditions may be better explained by a measure of acoustic distance between the SAE pronunciation and the accented pronunciation than by the measures we've explored so far. In order to get this measure, first and second formant values (F1 and F2) in Hertz were obtained from the midpoint of vowels in accented test words using Praat (Boersma & Weenick, 1992, ver 6.1). SAE versions of each of these words were also synthesized using Amazon Polly in the same manner as the test words, and the F1 and F2 values at the midpoints of these words were also obtained. These values were then converted to Mel in order to get a measure of psychoacoustic distance and account for differences in distances across higher or lower frequencies. Finally, Euclidean distance in F1-F2 acoustic space was calculated using the following equation:

$$d(p,q) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

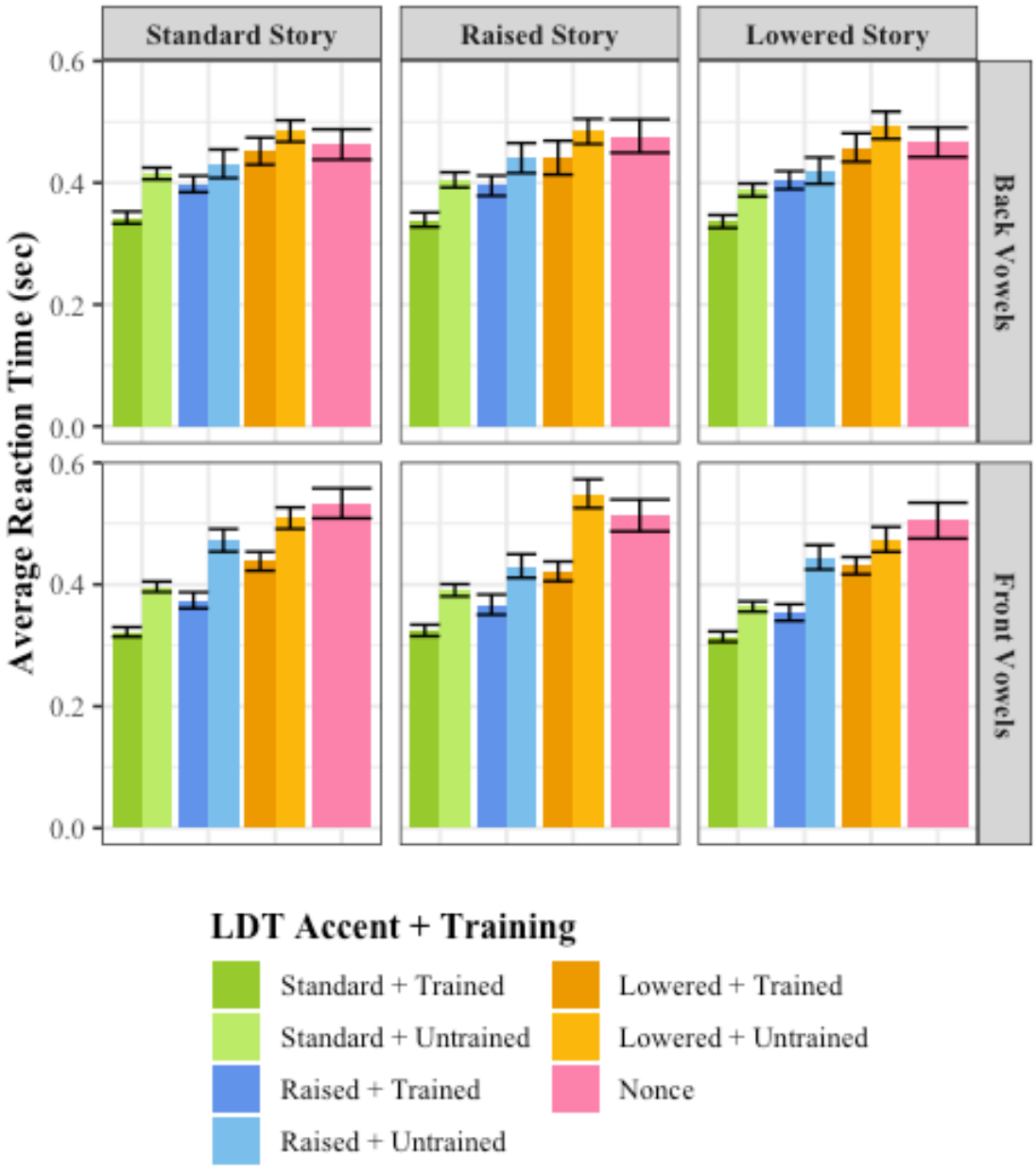

Figure 2: Reaction times for story conditions and front/back vowel words. Nonce items did not occur in the story and were therefore untrained.

where p and q are coordinates in F1-F2 acoustic space for the SAE and accented vowels. Distance of SAE words was coded as 0. Then, the relative fit of nested models containing this measure of psychoacoustic distance was tested. The best fitting model is shown below:

```
endorsement ~ psychoacoustic distance * front/back + LDT accent +
trained + (1 + LDT accent | subject) + (1 + story accent | word)
```

Compared to the best fitting model reported above that included LDT accent, front/backness, trained/untrained, and word frequency, this model fits the data significantly better ($\chi^2(1) = 15.93$, $p < 0.0001$), see Table 5 for fixed effects.

Table 5: Fixed effects for model including psychoacoustic distance.

	Estimate	Std. Error	z value	p value
(Intercept)	3.765	0.349	10.787	**< 0.0001** ***
Raised LDT	-4.013	0.378	-10.621	**< 0.0001** ***
Lowered LDT	-4.860	0.379	-12.811	**< 0.0001** ***
Psychoacoustic Distance	-1.875	0.407	-4.608	**< 0.0001** ***
Front/Back	0.900	0.306	2.939	**< 0.005** **
Trained/Untrained	-0.882	0.307	-2.875	**< 0.005** **
Distance:Front	1.879	0.464	4.053	**< 0.0001** ***

A model that included story accent condition was not significantly different from this model ($\chi^2(2) = 0.163$, $p = 0.92$), nor was a model that included word frequency ($\chi^2(1) = 2.62$, $p = 0.11$). The same model was tested on a subset of the data that contained only raised and lowered LDT accent types, and it, too, fit the data significantly better than the previous best model ($\chi^2(1) = 14.01$, $p < 0.0005$), see Table 7 for fixed effects.

Table 6: Fixed effects for Lowered/Raised LDT model.

	Estimate	Std Error	z value	p value
(Intercept)	-0.328	0.457	-0.717	0.474
Lowered LDT	-0.883	0.405	-2.184	**< 0.05 ***
Front/Back	0.970	0.426	2.277	**< 0.05 ***
Psychoacoustic Dist	-1.905	0.452	-4.220	**< 0.0001 *****
Trained/Untrained	-0.773	0.439	-1.759	**0.079 .**
Front/Back:Distance	1.913	0.511	3.743	**< 0.0005 *****

Discussion

Overall, the results of Experiment 2A show that although psychoacoustic distance improved model fit, LDT accent type was still a significant predictor of endorsement, with participants endorsing lowered accent LDT words significantly less than raised accent LDT words. Additionally, front and back vowels were treated differently by participants, as were words they had heard in the story vs. words they hadn't. Finally, there was a significant interaction of vowel backness and psychoacoustic distance on endorsement rates, with endorsement of back vowel words appearing to be more contingent on distance than front vowels. Critically, story accent condition did not significantly improve the fit of any of the models tested, suggesting that the accent of the story did not greatly affect participants' subsequent ratings of similarly accented words (although those in the lowered condition did demonstrate slightly better performance overall in the lexical decision task than those in the raised condition).

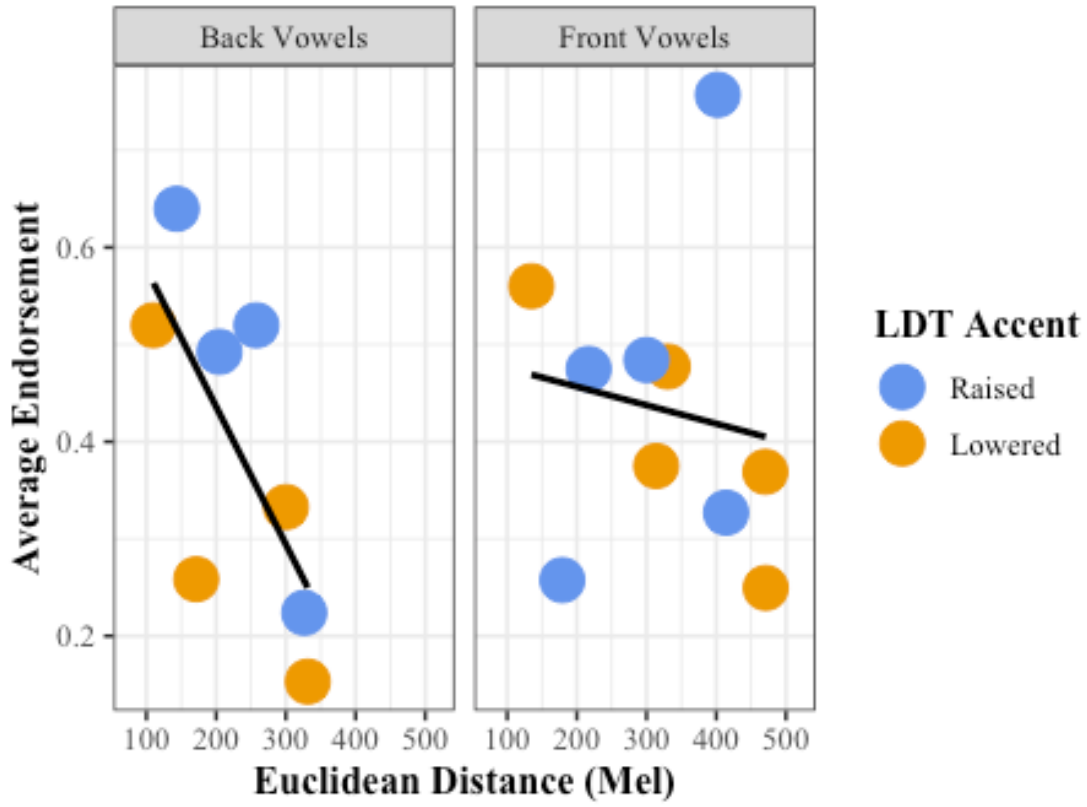

Figure 3: Average endorsement rate as a function of Euclidean distance between LDT accented words and their SAE pronunciations.

Participants' significantly higher endorsement on trained words than untrained words along with participants' performance on the post-experiment quiz suggests that listeners quickly adapted to the synthetic accents while listening to the story, resulting in story comprehension and increased priming in the LDT task. Participants in the raised story condition did not show greater endorsement rates for words that matched the accent condition they heard, suggesting that the accent heard in the 10 minute story did not reliably shift their phonological representations, either for words they had heard in the story or words that they had not. Participants in the lowered story condition, however, endorsed raised and lowered front vowel words more than those in the raised story

condition, providing some evidence that those who heard lowered front vowels expanded their front vowel categories in a way that those in the other story condition did not.

It is possible that this experiment failed to provide strong evidence of adaptation (as was found in Maye et al.) because of the between-subject nature of this study. In English, even across dialects of American English, vowels are highly variable. Especially given that the current study was conducted on MTurk and therefore potentially includes speakers of many different dialects of American English, it may be that comparisons across speakers are not sensitive enough to detect short-term changes to phonological representations as a result of accented input because variability across speakers may have simply been too great. Therefore, Experiment 2B, uses a within-subjects design to test whether participants change their endorsement rates of words with raised or lowered vowels before and after listening to the same story. This paradigm is even more similar to the original Maye, Aslin, and Tanenhaus (2008) study, but, as in Experiment 2A, MTurk and the Alice in Wonderland story are used in order to replicate and extend Maye et al.'s original findings.

Experiment 2B - Within-subject measures

Methods

Subjects

Participants were recruited using Psiturk (Gureckis & McDonnell, 2016) and Amazon Mechanical Turk. A total of 143 participants completed the experimental tasks,

but 61 of these participants were excluded from the analysis because they did not reach an "attentive worker" threshold, which was reached if participants' d-prime scores were at least 1.5 and they correctly answered at least half of the follow up questions about the Alice in Wonderland excerpt that they were asked to listen to. Two participants who reported learning a language other than English at home during early childhood were also excluded from analysis. This left 82 participants: 25 in the Standard American English story group, 32 participants in the raised accent story group, and 25 participants in the lowered accent story group (40 male, 40 female, 1 nonbinary, 1 preferred not to answer; five aged 18-24 years, twenty-three aged 25-34 years, thirty aged 35-44 years, twelve aged 45-54 years, eight aged 55-64 years, three aged 65-74 years, one preferred not to answer).

Stimuli

As in Exp. 2A, all auditory stimuli were synthesized using Amazon Polly, and participants heard the same 10 minute excerpt a in 2A. Front vowels in this story were either raised by one phoneme category or lowered by one phoneme category (see Table 7). Unlike 2A, 2B did not include /a/ as a front vowel in this experiment, as many dialects of English have merged /a/ and /ɔ/ into a single low back vowel.

Table 7: Pattern of Raising and Lowering for Accent Synthesis for Exp. 2B

Standard American English Accent	Raised Accent	Lowered Accent
i	i	ɪ
ɪ	i	eɪ

Table 7: Pattern of Raising and Lowering for Accent Synthesis for Exp. 2B		
Standard American English Accent	**Raised Accent**	**Lowered Accent**
eɪ	ɪ	ɛ
ɛ	eɪ	æ
æ	ɛ	æ

For the lexical decision task,"trained" words were removed, and instead only "untrained" words were used. More test words were added so that participants would hear four words in each accent type for each vowel, plus 44 nonce filler words. In all, there were 148 test words split into two groups of 74. One group contained the most and least frequent words in each vowel/accent type group of four (according to COCA) and the other group contained the second and third most frequent words in each group.

Procedure

The procedure was similar to Experiment 2A with one difference — participants completed the LDT task with one group of 74 *before* listening to the Alice in Wonderland story, and then completed the LDT task with the other group of 74 words *after* listening to the story. This change was made in order to get a baseline endorsement measure for each participant in order to measure change in endorsement rates before and after listening to accented speech. Order of presentation of LDT groups was counterbalanced across participants. Other than these changes, the headphone test, instructions, and post-experiment quiz were identical to Experiment 2A.

Results

Endorsement Rates

Logistic regression models were run to test how endorsement rates differed between accented story groups and the SAE story group. Comparison of nested models revealed that the regression with fixed effects of pre/post story LDT task, and LDT accent type, with random intercepts of subject and test word and a random slope of LDT accent type by subject best fit the data ($\chi^2(0) = 171.41$, $p < 0.0001$).

Table 8: Fixed effects for logistic regression of endorsement rate.

	Estimate	Std. Error	z value	p value
(Intercept)	**3.18430**	**0.32888**	**9.682**	**<0.0001 *****
pre/post	-0.30280	0.16225	-1.866	0.062 .
raised LDT	**-4.59553**	**0.47308**	**-9.714**	**<0.0001 *****
lowered LDT	**-4.38792**	**0.47634**	**-9.212**	**<0.0001 *****
filler LDT	**-5.71176**	**0.44603**	**-12.806**	**<0.0001 *****
pre/post:raised LDT	0.03273	0.18409	0.178	0.859
pre/post:lowered LDT	-0.08738	0.19008	-0.460	0.646
pre/post:filler LDT	0.01674	0.19346	0.087	0.931

Fixed effects (Table 8) show significant effect of LDT accent types, as well as a marginal effect of pre/post story task. Endorsement rates were significantly greater for SAE items than accented items. Additionally, endorsement rates were slightly greater overall in the pre-story than post-story LDT task. Because pre-story endorsements were greater than post-story endorsements, it is possible that participants start with broader criteria for

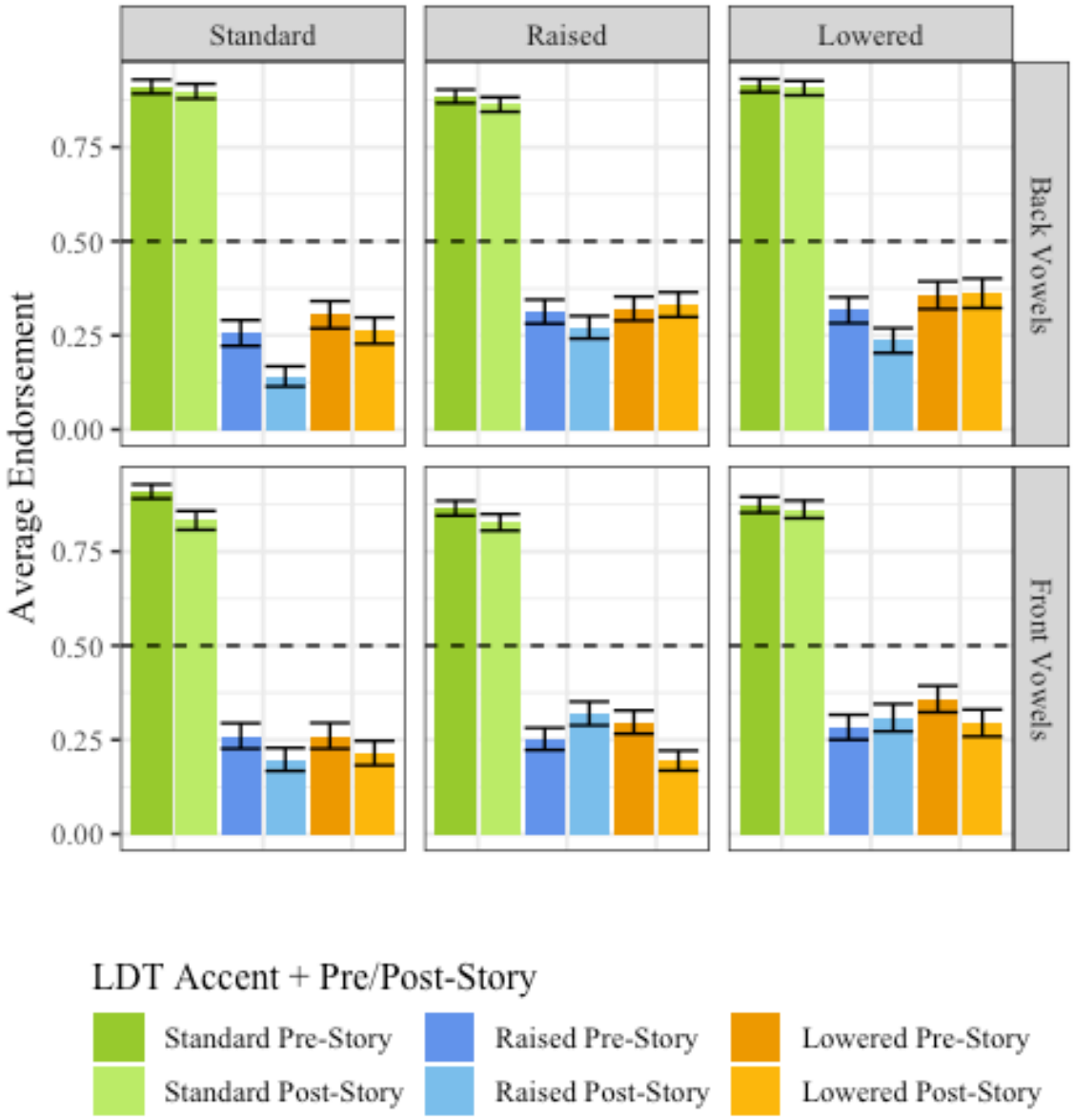

Figure 4: Endorsement rates pre- and post-story, shown by story condition and front/back vowel classification, with standard error.

lexical access, and as they gain evidence about the speaker, they start to narrow their categories and endorse less. Finally, fixed effects showed that endorsement was greater for accented items than filler nonce items, suggesting that participants may have been more likely to consider accented words as real words of English than true nonce words.

Critically, as in Experiment 2A, story condition did not improve model fit ($\chi^2(2) = 3.799$, $p = 0.150$)

Difference in Endorsement Pre- and Post-Story

To calculate difference in endorsement before and after the story, words were grouped by the original vowel of the word, LDT accent condition, and whether they were the higher or lower frequency word in the pair. The endorsement (0 or 1) of the pre-story LDT word was then subtracted from the endorsement (0 or 1) of the post-story LDT word with the same LDT condition and frequency group. These difference scores were then used in a 1-sample t-test to see if each accent group's differences were significantly different from zero for each LDT accent type for front and back vowels (see Fig. 5). There was a significant decrease in endorsement for the SAE story accent group for lowered LDT words containing front vowels ($t(250) = -2.346$, $p < 0.05$), for the SAE story accent group for raised LDT words containing back vowels ($t(174) = -2.826$, $p = 0.005$), for the raised story accent group for lowered LDT words containing front vowels ($t(217) = -2.901$, $p < 0.005$), and a marginally significant decrease in endorsement for the lowered story accent group for raised LDT back vowel words ($t(180) = -1.873$, $p = 0.06$). Overall, these results suggest that participants did not reliably change their endorsements of front vowels after hearing 10 minutes of a story in which front vowels were accented. Although participants in the raised story condition showed a trend towards increased endorsement for raised LDT front words, this trend was not significant. Participants in

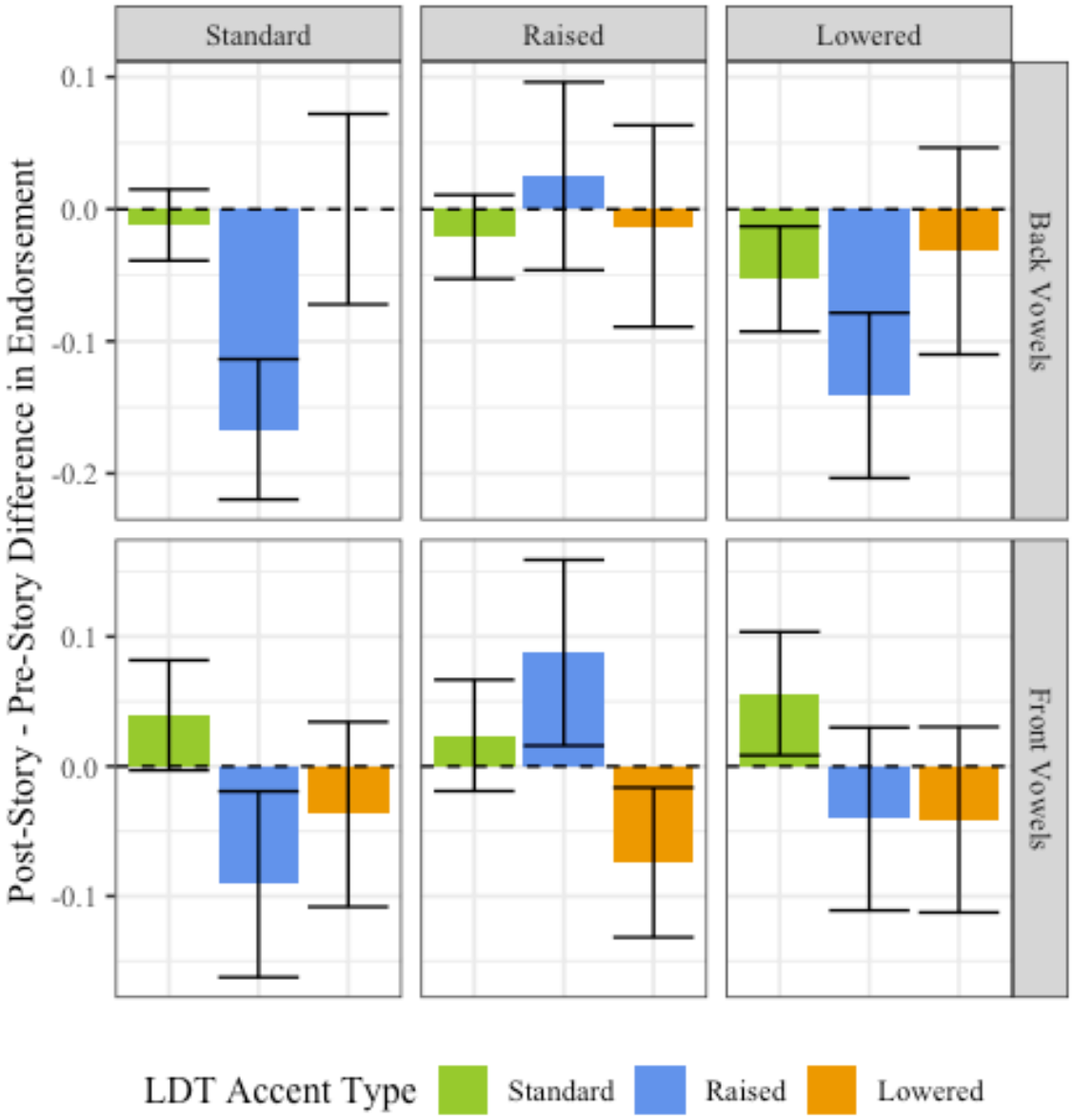

Figure 5: Difference in post- and pre-story LDT task, faceted by front/back vowel classification and story condition, with standard error.

the lowered condition showed no such trend for lowered LDT front words. For back vowel words, we observed a decrease in endorsement of raised words for those who heard the SAE story or the lowered story, but not those who heard the raised story. This suggests that subject who heard the raised story condition may have extended the vowel

raising they heard in front vowels the story to back vowels, thus keeping their criteria broad for raised vowels while other groups narrowed their criteria.

Table 9: One-sample t-tests compared against mu = 0 for each condition.

Story Accent	LDT Accent	Front/Back Vowel	t-value	95% CI	p-value
Standard	**Standard**	**Front**	-2.346	-0.146 - -0.013	**0.020 ***
Standard	**Raised**	**Front**	-1.013	-0.144 - 0.046	0.313
Standard	**Lowered**	**Front**	-0.515	-0.115 - 0.067	0.607
Standard	**Standard**	**Back**	-1.181	-0.087 - 0.022	0.239
Standard	**Raised**	**Back**	-2.826	-0.184 - -0.033	**0.005 ****
Standard	**Lowered**	**Back**	-0.126	-0.102 - 0.090	0.900
Raised	**Standard**	**Front**	-1.010	-0.094 - 0.030	0.313
Raised	**Raised**	**Front**	1.312	-0.283 - 0.141	0.191
Raised	**Lowered**	**Front**	-2.901	-0.177 - -0.034	**0.004 ****
Raised	**Standard**	**Back**	-1.271	-0.085 - 0.018	0.205
Raised	**Raised**	**Back**	-0.654	-0.114 - 0.057	0.514
Raised	**Lowered**	**Back**	-0.333	-0.106 - 0.075	0.740
Lowered	**Standard**	**Front**	-0.749	-0.086 - 0.039	0.454
Lowered	**Raised**	**Front**	0.346	-0.079 - 0.113	0.730
Lowered	**Lowered**	**Front**	-0.943	-0.138 - 0.049	0.347
Lowered	**Standard**	**Back**	-0.745	-0.074 - 0.033	0.457
Lowered	**Raised**	**Back**	-1.873	-0.170 - 0.004	**0.063 .**
Lowered	**Lowered**	**Back**	0.249	-0.091 - 0.117	0.804

Reaction Time

Analyses of reaction time for "word" responses showed similar trends as endorsement rates: comparison of nested models showed that the following model best fit the data ($\chi^2(3) = 54.94$, $p < 0.0001$, see Table 10 for fixed effects):

```
RT ~ pre/post story + LDT accent + story accent +
(1 + pre/post | subject) + (1 | word)
```

Table 10: Fixed effects for model of reaction time.

Predictors	Estimates	CI	p -value
(Intercept)	0.47	0.42 – 0.52	**<0.001** ***
Post-story	-0.03	-0.05 – -0.02	**<0.001** ***
Raised story accent	-0.05	-0.11 – 0.00	**0.052** .
Lowered story accent	-0.08	-0.14 – -0.02	**0.006** **
Raised LDT	0.16	0.12 – 0.21	**<0.001** ***
Lowered LDT	0.15	0.11 – 0.20	**<0.001** ***
Nonce LDT	0.13	0.09 – 0.18	**<0.001** ***

Participants showed faster RTs during the post-story LDT than the pre-story LDT, faster reaction times for accented story conditions than the SAE story condition, and slower reaction times for raised, lowered, and nonce LDT words compared to SAE LDT words.

Changes in RT from pre-story LDT to post-story LDT for each group and LDT type were also analyzed (Fig. 6). One-sample t-tests were conducted to determine if the difference between post- and pre-story RTs was significantly different from zero. RTs of "word" responses were significantly faster post-story in the SAE story group for SAE

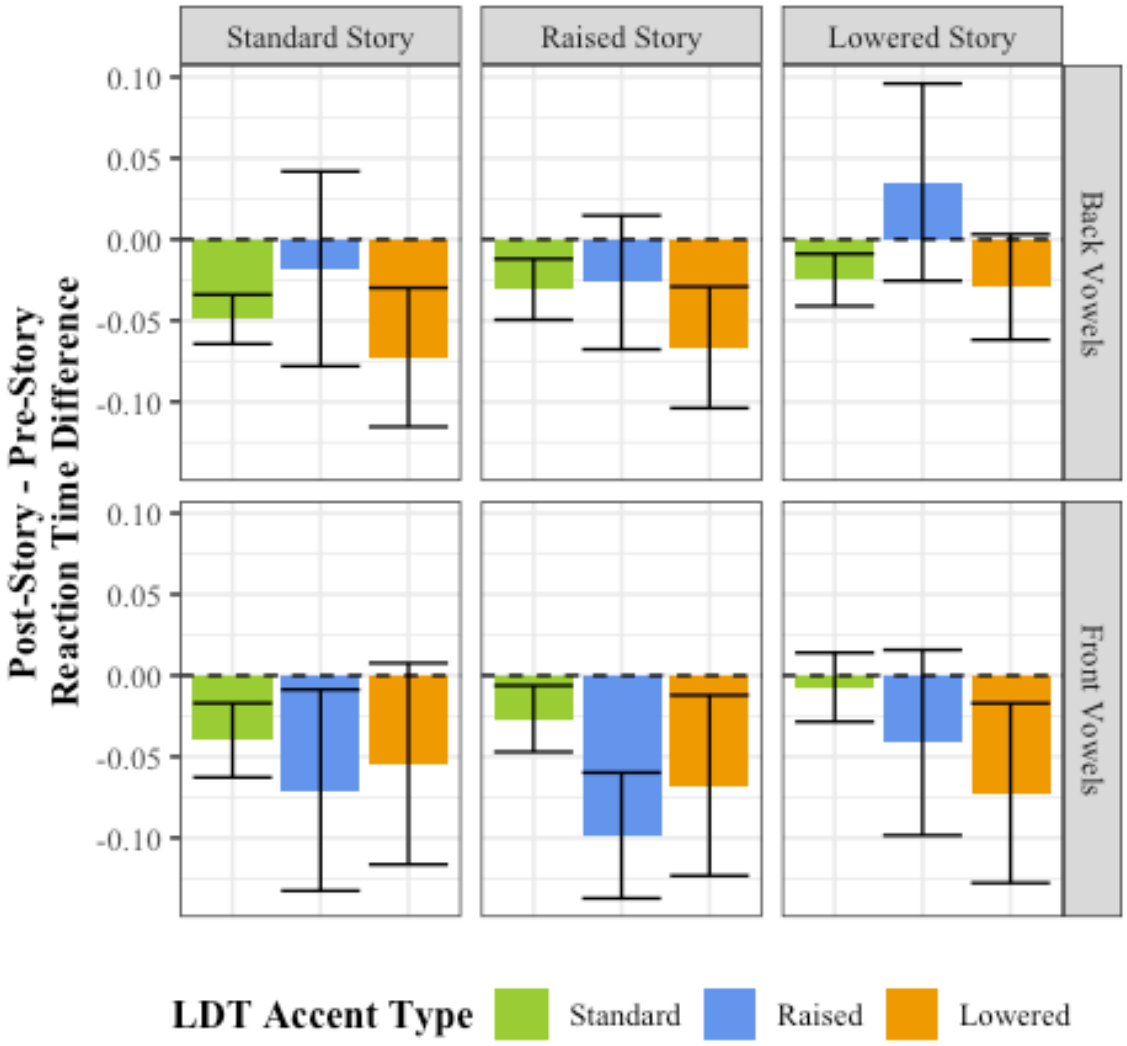

Figure 6: Difference in post- and pre-story reaction times for "correct" responses ("word" for SAE and accented items, "nonword" for nonce items), faceted by front/back vowel classification and story condition

back vowel LDT words ($t(25) = -3.24$, $p < 0.01$) and in the raised story group for raised front vowel LDT words ($t(30) = -2.54$, $p < 0.05$). These findings again demonstrate little change in behavior before and after listening to an accented story, but provide some evidence that those in the raised group may have changed their lexical processes slightly after exposure to the raised accent.

Psychoacoustic Distance

As in Experiment 2A, the effect of Euclidean distance in F1-F2 space was also tested. This distance was obtained in the same way as in Experiment 2A. The best fitting model in the analysis of endorsement was compared against a model including fixed effects terms of LDT accent, story accent, and pre/post story LDT task, and an additional term of euclidean distance, which was scaled for model fit. Random effects consisted of a random intercept of word, a random intercept of subject, and a random slope of pre/post story by subject. The model that included psychoacoustic distance was not significantly different from that without ($\chi^2(1) = 2.02$, $p = 0.15$). However, when distance is included in a linear model with difference of post/pre-story LDT endorsement as the dependent variable and a random intercept of subject, the model with a fixed effect of psychoacoustic distance and an interaction of psychoacoustic distance, LDT accent type, and front/back vowel was a significantly better fit than the null model ($\chi^2(4) = 52.22$, $p < 0.0001$). Fixed effects of the model revealed a main effect of psychoacoustic distance, a significant difference between LDT endorsement for front vs. back vowels but no significant difference between lowered and raised accent types for front vowel words (see Table 11).

Table 11: Fixed effects of model of difference post- and pre-story LDT tasks.

Predictors	Estimates	CI	t value	p value
(Intercept)	-0.03	-0.05 – -0.01	-2.90	**0.004 ****
psychoacoustic dist	-0.06	-0.10 – -0.01	-2.21	**0.027 ***
dist : raised LDT : back vowel	0.19	0.11 – 0.27	4.64	**<0.001 *****

Table 11: Fixed effects of model of difference post- and pre-story LDT tasks.				
Predictors	**Estimates**	**CI**	**t value**	**p value**
dist : lowered LDT : back vowel	0.19	0.12 – 0.26	5.45	**<0.001** ***
dist : raised LDT : front vowel	0.05	-0.01 – 0.11	1.53	0.127

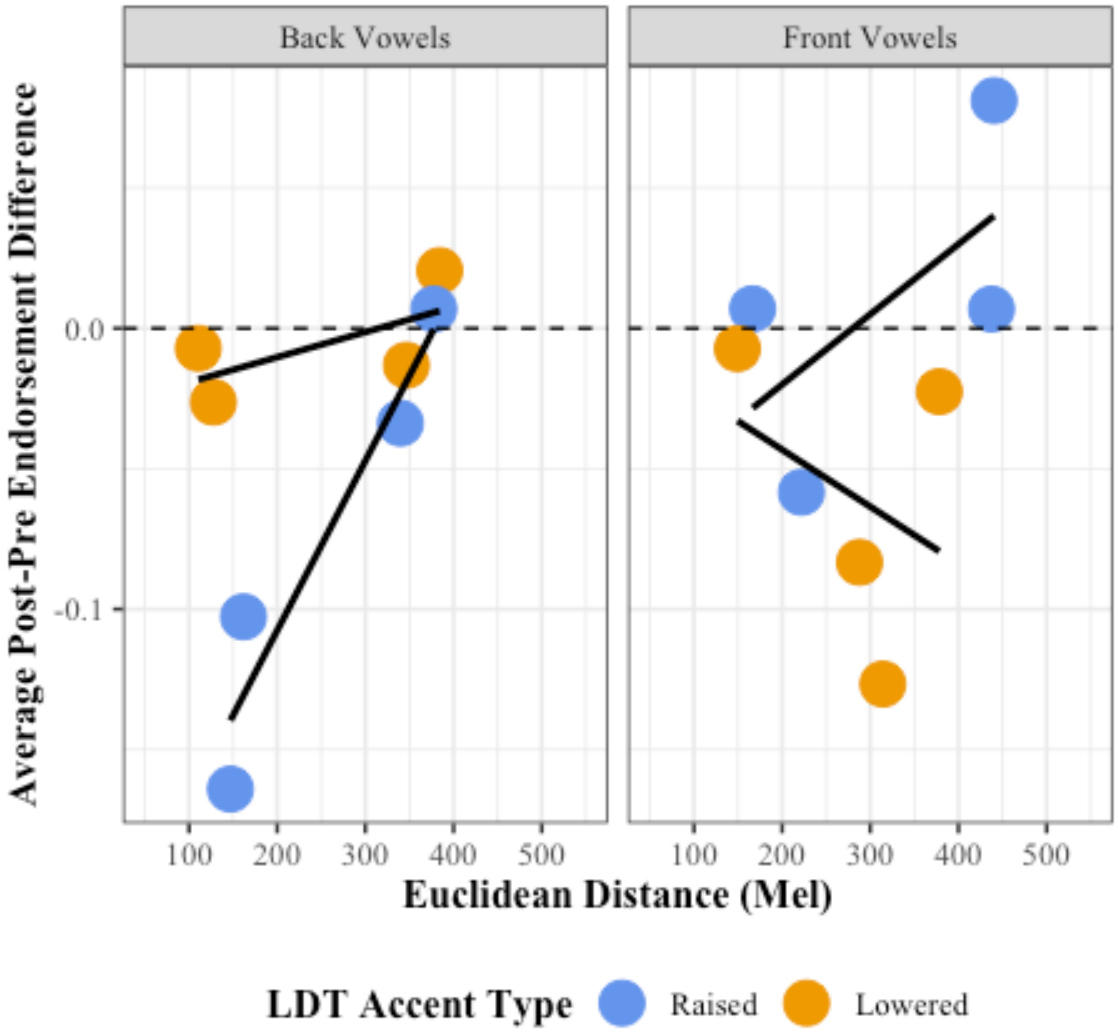

Figure 7: Difference between post-story and pre-story LDT endorsement as a function of euclidean distance between accented F1/F2 and SAE F1/F2 and faceted by front/back vowels. Linear patterns are subsetted by LDT accent type.

Discussion

Experiment 2B, like Experiment 2A, did not replicate findings by Maye et al (2008), which demonstrated that listening to an accented story caused participants to at least temporarily change their criterion for phonological-lexical mappings such that they were more likely to endorse words that matched the accent they had heard. In the current experiment, with overall low endorsement rates for non-standard pronunciations, participants were not able to access accented lexical items as accurately as Standard American English items during LDT tasks. Furthermore, participants did not appear to greatly improve performance on accented words that matched the story accent they heard. While those who heard the raised accent story endorsed lowered front vowel words somewhat less and raised front vowels somewhat more, those who heard the lowered and standard accent stories neither increased or decreased endorsement of raised or lowered front vowel words. Endorsement of raised back vowels *did* decrease for participants in the standard and lowered story condition, but not for those in raised story condition. These findings suggest first that in the preliminary LDT task, participants were more likely to accept deviations in word pronunciations, as endorsement was generally higher in the first LDT task than the second. Additionally, participants who heard the raised story *did* use phonological information from the story to change their representations somewhat — they continued to endorse raised back vowels at the same rate as they had initially, they endorsed lowered front vowels less and raised front vowels more after story exposure. Therefore, for the raised accent condition, participants appear to show some behavior consistent with a *pathway* theory of phonological representation, in which

listeners retain the boundaries of their native categories but may also incorporate tokens in a specific direction of phonological change.

Maye, Aslin, and Tanenhaus (2008) demonstrated that after hearing shifted front vowels, participants not only increased their endorsement of words that contained shifted front vowels, but also of words that contained back vowels shifted in the same direction. Experiment 2B also provided some evidence that listeners extended knowledge about front vowel shifts to back vowels, as participants who heard raised front vowels in the story did not decrease endorsement of raised back vowels, while participants who heard SAE and lowered stories did. Overall, generalizations across similar sounds has been demonstrated a number of times in the literature. For example, Nielson (2011) showed that exposure to longer VOT at one place of articulation was then generalized to result in longer VOTs at other places of articulation, and Chládková, Podlipsky, and Chionidou (2017) showed that shifts in height of front vowels affected how Greek listeners perceived back vowels. Sanker (under review) also demonstrated that exposure to shifted F1 or F2 in one vowel influences perceptual boundary shifts in other vowels.

Additionally, because stronger effects were found in the *decrease* of endorsement of raised back vowel words in the SAE condition than in the increase of endorsement of raised back vowel words in the raised story condition, the current results are also consistent with an initial *expansion* of criteria for lexical access. As participants gain experience with a speaker, their boundaries for phonological units constrict. This pattern of results is consistent with work by Schmale et al. (2012), Weatherholz (2015), and Witteman et al. (2010), among others, showing that when listeners are exposed so speech

variability, they are more likely to tolerate other types of variability as well. White and Aslin (2011) found that toddlers were also more likely to accept words in which /a/ was produced as /ɛ/ after hearing speech in which /a/ was produced as /æ/ (notably, this finding is also consistent with the *pathway* hypothesis, as /ɛ/ and /æ/ are both fronted and raised compared to /a/). Interestingly, this *expand* pattern was found for back raised vowel words but not front raised vowel words, which calls into question listeners' base tolerance for shifts of specific vowels.

The patterns of adaptation to raised LDT words in 2B found in the raised accent condition were not observed for lowered LDT words in the lowered accent condition. This fact, along with the findings of Exp. 2A, suggest that participants may be treating raised and lowered vowels differently. When looking more closely at individual vowel shifts, it appears that listeners tended to endorse tense > lax lowered vowels less than lax > tense lowered vowels. Among lowered vowels, endorsements of ɪ > eɪ and ɛ > æ were highest, suggesting that vowel shifts towards more focal, tense vowels were tolerated more than shifts away from vowel space boundaries. For back vowels, ʊ > o and ʌ > ɔ were endorsed most, showing preference for shifts towards more focal, tense, rounded vowels (towards the back-most part of vowel space). Raised front vowels showed the same pattern, with ɪ > i most highly endorsed. It's possible that this effect emerged due to the recitation style of speech in the LDT task: words were said one at a time in relatively clear speech. It may be that when listeners hear this type of speech, they are less tolerant of centralization, as clear speech typically involves an expanded vowel space (cf. van der Feest, Blanco, & Smiljanic, 2019). Differences between behavior of those in the lowered

conditions of Exp. 2A and 2B may also follow this pattern, as 2A included a æ > ɑ shift while 2B did not. On the other hand, other raised vowels did not follow this pattern. For raised back vowels, lax > tense shifts were endorsed less than tense > lax shifts, and for raised front vowels only ɪ > i was endorsed more, while the other three raised front vowels were endorsed about the same amount.

It is also possible that neighborhood density, phonotactic probability, or vowel frequency may be driving some of the observed effects. For words with higher phonological neighborhood density, participants might expect hyperarticulation of the target vowel, thereby making them less tolerant to shifted vowels (Scarborough & Zellou, 2013). They may also expect more vowel space expansion, as found in clear speech, and therefore the tense-lax directional shift distinction between some of the vowel shift observed here may be exacerbated by expectations about hyperarticulation.

Additionally, vowels, like consonants, are constrained by phonotactics. For example, English, there is some evidence that listeners can use knowledge about vowel phonotactics to segment unfamiliar words (Skoruppa, Nevins, Gillard, & Rosen, 2015). It is possible that phonotactic probability of sequences may have played a role in participants' endorsement rates—more probable sequences may have been endorsed more because they sounded like real words, even if they weren't actually recognized. To my knowledge, little work has been done on the effect of phonotactic probability of vowels on word recognition in particular, but higher frequency and predictability of phonological segments has been associated with reduced signals in production (cf. Cohen Priva & Jaeger, 2018), and informativity, or the average value of a segment's predictability across

all contexts of use, has likewise been shown to affect reduction processes in production (Cohen Priva, 2015; Seyfarth, 2014), raising the possibility that one or more of these factors underlie participants' different endorsement patterns for raised and lowered vowels. It may have been the case that the words chosen for this study included words in which shifts in vowel height went contrary to expectations about (hyper)articulation, which in turn could have made recognizing the intended lexical item more difficult. Similarly, vowel category shifts, especially lowered shifts, could have resulted in significant shifts in informativity, confounding any effects of accent adaptation.

Moreover, perceptual asymmetries in vowels, which have been demonstrated in previous literature (Polka & Bohn, 2003, 2011; Schwartz, Abry, Boë, Ménard, & Vallée, 2005; Masapollo, Polka, Ménard, Franklin, Tiede, & Morgan, 2018; Masapollo, Zhao, Franklin, & Morgan, 2019), may be at play in the types of shifts participants are more likely to tolerate or adapt to in accented speech. In a study of Australian listeners' ability to categorize vowels across five native accents of English, Shaw et al. (2018) found that listeners did not show adaptation to vowels of non-Australian varieties of English after exposure to a passage in that variety. They, too, found asymmetries in task performance across different types of vowel shifts, providing more evidence that certain vowel shifts may be asymmetric in their salience or acceptability, although more work must be done to elucidate this claim.

Finally, it may be the case that the LDT paradigm used was not sensitive enough to detect accent adaptation, especially given the variety of listening conditions in which the MTurk participants may have been doing the task, as well as the fact that the

difference measure between pre-story and post-story performance was not matched by lexical item, but by vowel and relative frequency. Because LDT provides only binary responses, it may not be fine-grained enough to show the degrees of processing differences actually taking place when listeners are exposed to accented speech. Therefore, while accent adaptation was not observed for lowered vowels in Exp. 2B, listeners may still be adapting somewhat to the lowered-vowel accent they heard.

Conclusion

To assess whether listeners from a wide variety of ages and backgrounds could adapt to a novel, synthetic accent after listening to a 10 minute story, two experiments were conducted on MTurk with Standard American English and two synthesized accents, one in which all front vowels had been raised by one vowel category in American English vowel space, and one in which all front vowels had been lowered by one vowel category in American English vowel space. Experiment 2A tested for differences across listeners of different accent conditions, but found no significant effects of story accent overall. There was, however, a significant difference between raised and lowered words in the LDT task. 2B tested within participants and found some evidence of adaptation to the raised story accent but not the lowered story accent. The overall lack of evidence for accent adaptation in the current study may be due to tasks demands and limited sensitivity of the lexical decision paradigm. Furthermore, the asymmetries observed between raised and lowered accent types may be the result of underlying phonological

processes, such as frequency, predictability, informativity, or asymmetric perceptual assimilation.

Chapter 4: Rapid Adaptation to a Novel Accent: Eye tracking

Introduction

Experiment 2 failed to show strong effects of adaptation to a novel accent, even after about 10 minutes of exposure. However, this lack of effects may be due in part to the inability of lexical decision tasks to get at fine-grained distinctions between participants' treatment of different test conditions, as it is a binary response variable. Therefore, the current study uses a paradigm originally designed by White and Morgan (2008) to test toddlers' sensitivity to mispronunciations of familiar words. In this paradigm, participants hear a word and see two images, one of the familiar item that matches the familiar word, and one of an unfamiliar item that the participant would not have a name for. White and Morgan wanted to see if mispronunciations in familiar words would cause toddlers to fail to access the familiar word in their lexicon and map what they perceived as a novel word onto a novel item. This paradigm proved to be a good measure of toddlers' sensitivity to the sounds of their language, as they demonstrated graded looking times to the familiar image as a function of the severity of the mispronunciation. Indeed, this method has shown to be a sensitive measure of the degree of lexical access (or disruption thereof) in both adults and toddlers and across different word segments and phoneme classes (White & Morgan, 2008; Ren & Morgan, 2011; Franklin, Tin, Gasdaska, & Morgan, in preparation; Franklin & Morgan, 2020; White & Aslin, 2011).

For example. Franklin, Tin, Gasdaska, and Morgan (in preparation) showed that this paradigm successfully measures adults' sensitivity to mispronunciations with the

same type of graded sensitivity that toddlers had shown previously. Furthermore, Franklin et al. demonstrated that while consonants constrain lexical access more strongly than vowels, listeners still show graded sensitivity to vowel mispronunciations. Therefore, while the paradigm in Experiment 2 may not be sensitive enough to reliably test vowel-based accent adaptation, this paradigm may be.

The purpose of the current experiment was to test whether participants are sensitive to minimal exposure of a novel accent and whether participants *expand*, *shift*, or create a *pathway* within their phonological representations. To do so, an eye tracking procedure was paired with White and Morgan's (2008) paradigm. Exposure to the novel accent took place only within the carrier sentence of each test word, "where's the…" or "where're the…" in which the /ɛ/ in "where" was pronounced as [i]. Although this is short exposure, previous research has provided evidence that listeners can adapt to accents with less than a minute of exposure (Clarke & Garrett, 2004), are sensitive to even minute sub-phonemic deviations from their native norms (Harada 2006), and can even detect accented speech from a single monosyllabic word (Park, 2013). Therefore, the sentential context in the current experiment may be enough for participants to detect the specific accented deviation tested in this experiment.

I hypothesize that if participants in this study treat the accented pronunciations the same as the mispronounced words, the *expand* hypothesis would be best supported. However, if participants are more likely to accept the accented words as referring to the target item than the other mispronunciations, the *shift* and *pathway* hypotheses will be better supported. Furthermore, if the accented words are still treated as less likely to refer

to the target item than correctly pronounced tokens, the *pathway* hypothesis will better explain this data than the *shift* hypothesis.

Methods

Subjects

30 participants were recruited from the Brown University community. Of these, two were excluded from analysis (early bilinguals). The 28 remaining participants were monolingual speakers of American English (mean age 23.7, 19 female). In a follow up questionnaire, participants were asked about their experience with accents, including their ease of understanding accented speech and their current and childhood experience with accented speakers.

Stimuli

An "accented" sentential context was devised in order to direct participants to the possibility that the speaker of the experiment did not have a Standard American English accent. The sentential context "where's the…" or "where're the…" was manipulated so that in locations where a listener would generally hear /ɛ/ in SAE, they would instead hear /i/. Therefore, the sentential contexts sounded like /wirz ðə/ or /wirɚ ðə/.

Test words were selected from words whose medial vowels could be manipulated along one, two, and three phonological dimensions (height, backness, and tenseness, not necessarily in that order) to produce nonce words. Additionally, these words were easily imageable (e.g. "car" but not "peace"). Finally, words were balanced so that each of the

possible orders of feature mispronunciation (e.g., one-feature height mispronunciation, two-feature height and backness mispronunciation, and three-feature height, backness, and tenseness mispronunciation) included 12 words per mispronunciation order for a total of 72 words (see appendix for word list). Correctly pronounced words and mispronounced words were divided into four experiment conditions such that for each mispronunciation group, participants heard 3 correctly pronounced words, 3 words mispronounced along one phonological dimension, 3 mispronounced along two phonological feature dimensions, and 3 mispronounced along three phonological feature dimensions. Each participant only heard one version of each word.

An additional 19 words with medial vowel /ɛ/ were manipulated so that participants would always hear /i/ instead of /ɛ/ in order to match the accentedness of the sentential context. Finally, 12 real words and 24 nonce words were included as filler trials.

All auditory stimuli were synthesized using Amazon Polly's Neural Joanna voice. SSML was used in order to input all words (correctly pronounced and mispronounced) in IPA.

For each word, two images were paired: one depicting a referent of the familiar test word or filler word and one distracter image (see Fig. 1). In nonword trials, the target image was novel, while the distractor was familiar. In all trials, one image was familiar and one was novel, and no images were used for more than one trial for each participant. The presentation of familiar and unfamiliar images was based on previous studies that have shown that participants may perceive a mispronounced token as a novel label for an unfamiliar (and therefore nameless) item (Franklin et al, in preparation). Moreover, this

design avoids the possibility raised in studies that have used pairs of familiar items that participants may look at the target because the distractor matched the mispronunciation less, and thus may not be the best measure of sensitivity to phonological change.

Target images were prototypical representations of each target word. Distractor images were images of objects that did not have readily accessible labels. Target and distractor images were paired for overall impressionistic visual similarity, and all images were presented to each participant only once during the experiment. Images were displayed side by side such that half the time, the target was presented on the left side of the screen and the other half of the time the target was presented on the right side of the screen. These presentations were counterbalanced so that half the participants saw the images in the same configuration and half the participants saw the images in the opposite configuration.

Procedure

Participants were tested using an SMI remote binocular eye tracker paired with SMI Experiment Center and iViewX software. The study was conducted in a dimmed, sound attenuated booth in the Metcalf Infant Research Lab at Brown University. Subjects were told they would simultaneously hear a word and see two images at a time on a computer screen and were asked to look at whichever image best fit the word they heard.

Participants completed four warm-up trials and received on-screen instructions before starting the test trials. Each trial began when subjects looked at a center oriented fixation cross on a computer monitor. When they had fixated for 500 ms, the test images

appeared. At the same time, participants heard the test token (either correctly pronounced or mispronounced) through stand-alone speakers. The images remained on the screen for 4 seconds before the next trial began. Each participant completed 127 trials (72 test trials, 19 accent trials and 36 filler trials).

Monocular (left eye) gaze was recorded at 120 Hz and downsampled to 60 Hz. Raw gaze data were preliminarily analyzed using in-house software to ascertain looking time to each of the images in each trial, as well as longest gaze fixation on each image, latency to each image, and number of gaze switches in each trial.

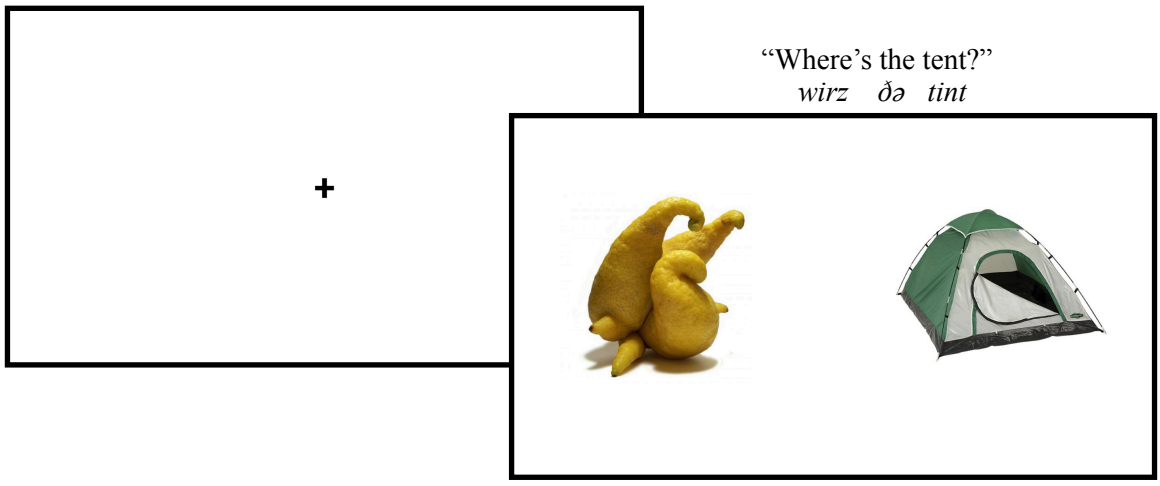

Figure 1: An example trial for "tent." Participants saw a fixation cross on the center of the screen, then two images side-by-side and heard the auditory stimulus.

Results

All analyses were carried out using the lme4 package (version 1.1-19) in R version 3.5.1.

Proportion of Time Looking to Target

In the results reported below, the dependent measure used was proportional looking time to the target image (i.e., the referent of the correctly pronounced label). Proportion looking to target was computed as follows:

$$\frac{\text{Proportion of Time Looking to Target}}{\text{Proportion of Time Looking Anywhere on the Screen}}$$

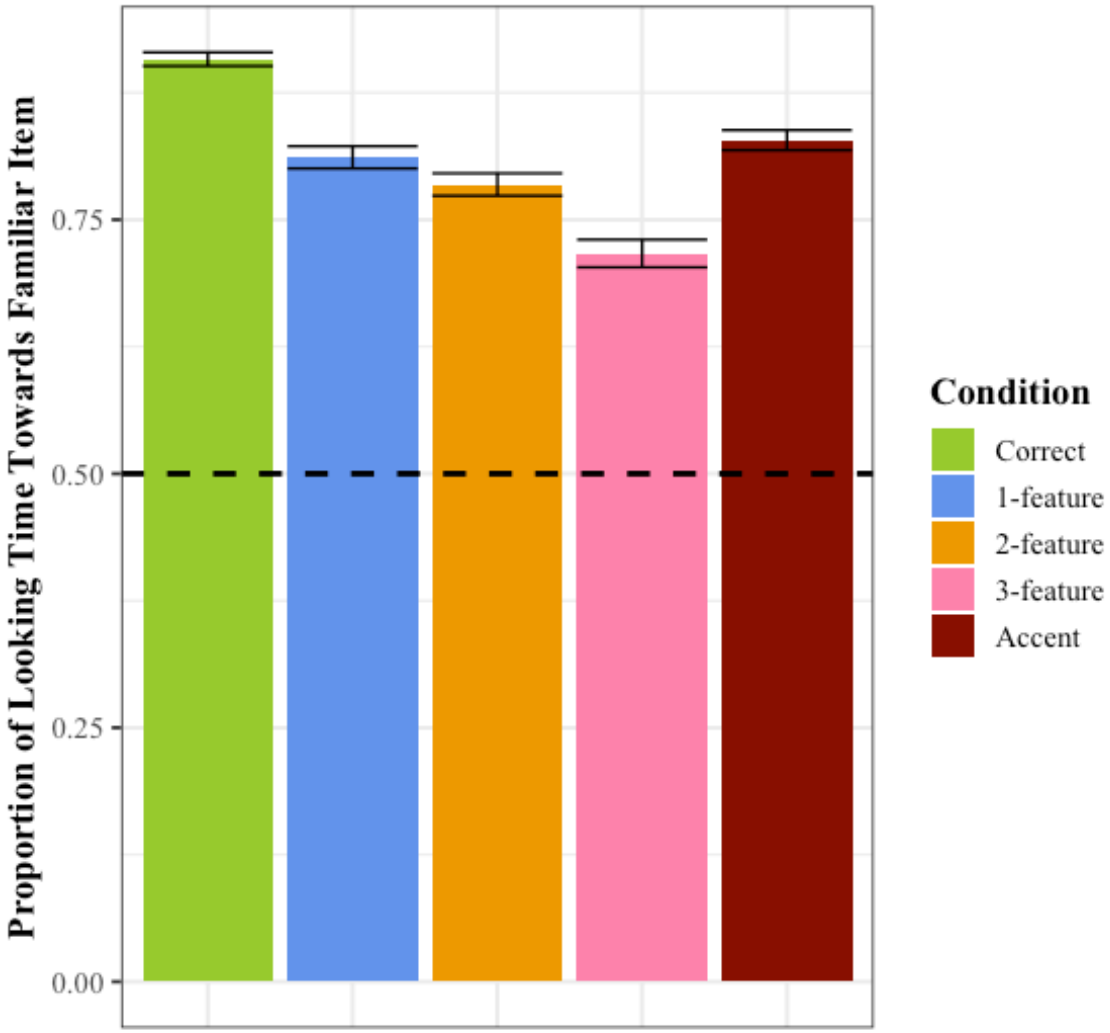

Figure 2: Average PLT for each condition with standard error. Dotted line is at chance.

Proportion of time looking to target (PLT) was computed over the entirety of each trial as well as frame-by-frame for each trial. Both of these measures were used in data analysis.

To test for a main effect of condition, a one-way Anova was conducted with PLT as the response variable. The test revealed a main effect of condition ($F(4,2533) = 36.64$, $p < 0.0001$). Pairwise comparisons using t-tests with pooled SD were then conducted. Correctly pronounced words had significantly higher looking times than all other conditions (1-feature: $p < 0.0001$, 2-feature: $p < 0.0001$, 3-feature: $p < 0.0001$, accent: $p < 0.0001$). 1- and 2-feature conditions were not significantly different from each other ($p = 0.18$), or the accent condition ($p = 0.91$, 0.18, respectively). The 3-feature condition was significantly different from all of these ($p < 0.0001$ for all). All conditions were all significantly different from chance (0.5), with $p < 0.0001$.

Time Course

Looking time was also analyzed across the time course of the trial (Mirman, 2014). The correctly pronounced condition was treated as the baseline, and parameters were estimated on mispronounced conditions. Looking patterns were modeled with second-order polynomials and fired effects of condition, as well as their interaction, random intercepts of subject and word and random slopes of first and second order polynomials by subject. This model was significantly better fit than the model that did not include condition ($\chi^2(15) = 14402$, $p < 0.0001$).

Table 1: Fixed effects of the logistic regression on time course of looking. ot1 and ot2 are linear and quadratic orthogonal polynomials.

	Estimate	Std. Error	z value	p value
(Intercept)	1.49953	0.09280	16.159	**< 2e-16** ***
ot1	28.45567	0.39815	71.470	**< 2e-16** ***
ot2	-13.98957	0.38396	-36.435	**< 2e-16** ***
1-feature	-0.66937	0.01280	-52.278	**< 2e-16** ***
2-feature	-0.80047	0.01270	-63.037	**< 2e-16** ***
3-feature	-1.15168	0.01240	-92.886	**< 2e-16** ***
accented	-0.59597	0.12598	-4.731	**2.24e-06** ***
filler	-0.89015	0.09914	-8.979	**< 2e-16** ***
ot1:1-feature	-5.92107	0.16872	-35.094	**< 2e-16** ***
ot1:2-feature	-6.26549	0.16892	-37.091	**< 2e-16** ***
ot1:3-feature	-8.39682	0.16509	-50.862	**< 2e-16** ***
ot1:accented	-1.64041	0.17695	-9.270	**< 2e-16** ***
ot1:filler	-9.33960	0.14759	-63.283	**< 2e-16** ***
ot2:1-feature	4.08811	0.16784	24.357	**< 2e-16** ***
ot2:2-feature	5.31510	0.16648	31.927	**< 2e-16** ***
ot2:3-feature	6.81125	0.15723	43.321	**< 2e-16** ***
ot2:accented	3.63409	0.17581	20.670	**< 2e-16** ***
ot2:filler	4.09961	0.14595	28.090	**< 2e-16** ***

All conditions significantly differ from the correctly pronounced condition in intercept, slope, and curve. To test differences between mispronunciation and accent conditions, the same model was used with data from only two conditions at a time. Results showed that the accented condition had different looking patterns from the correctly pronounced condition ($\chi^2(3) = 2837.7$, $p < 0.0001$), but also from the 1-feature

($\chi^2(3) = 468.35$, $p < 0.0001$), 2-feature ($\chi^2(3) = 812.73$, $p < 0.0001$), and 3-feature ($\chi^2(3) = 4076.20$, $p < 0.0001$) mispronunciation conditions. Additionally, 1- and 2-feature mispronunciation conditions were significantly different from each other ($\chi^2(3) = 185.30$, $p < 0.0001$), as were 2- and 3-feature mispronunciation conditions ($\chi^2(3) = 1378.50$, $p < 0.0001$). Within these comparisons, all conditions significantly differed from each other in intercept, linear, and quadratic terms, except for accented and 1-feature mispronunciation's quadratic terms, which were only marginally significant ($z = -1.79$, $p = 0.07$), and 1- and 2-feature mispronunciations' linear terms, which were not significantly different ($z = -1.20$, $p = 0.23$), see Appendix for all fixed effects. These results demonstrate that although participants did not treat accented words in the same way as correctly pronounced SAE words, they also did not treat them like mispronounced words, with some delay in early looking followed by overall higher looking times once PLT plateaued after about 1.75 seconds. 1- and 2-feature mispronunciations show some differences in the speed of looking, but plateau at the same PLT. 3-feature mispronunciations garnered reliably less looking than the other conditions, but was still well above chance.

In the early part of the the trial, it also appeared that the accented words showed an early delay in looking, even before the target word was uttered. To test if looking was significantly lower for accented words than the other words at this early stage of the trial, PLT was calculated from 200 ms after the onset of the trial to 1000 ms after the onset of the trial for each trial for each subject. 200 ms was chosen as the starting point of this analysis because that is roughly the amount of time it takes for humans to program a

saccade, so the earliest time participants could look at the images is 200 ms after trial onset. 1000 ms was chosen as the offset of this analysis because at that point, less than half the words had been fully uttered. An ANOVA was used to test whether there was a main effect of condition on PLT for this slice of time, and the result showed a marginal effect of condition ($F(5,3540) = 2.131$, $p = 0.059$). This finding confirmed that looking to the the target was not significantly different for the images of accented words and those of the words in the other conditions before the utterance of the target word.

Psychoacoustic Distance

Rather than categorizing vowels by phonological features, it is possible to conceptualize differences between correctly pronounced and mispronounced/accented vowels through measures of psychoacoustic distance in F1-F2 vowel space. To obtain this measure, first and second formant values (F1 and F2) in Hertz were obtained from the midpoint of vowels in test words using Praat (Boersma & Weenick, 1992, ver 6.1). SAE versions of each of the accented words were also synthesized using Amazon Polly in the same manner as the test words, and the F1 and F2 values at the midpoints of these words were also obtained. These values were then converted to Mel in order to get a measure of psychoacoustic distance and account for differences in distances across higher or lower frequencies. Finally, Euclidean distance in F1-F2 acoustic space was calculated using the following equation:

$$d(p,q) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

where q are coordinates in F1-F2 acoustic space for the correctly pronounced SAE vowels, and p are coordinates in F1-F2 acoustic space for mispronounced and accented vowels. To test the effects of psychoacoustic distance above and beyond condition, a linear model with PLT as the outcome, condition and psychoacoustic distance as fixed effects, and random intercepts of subject and word was run. Compared to the model that did not include psychoacoustic distance, this model fit the data significantly better ($\chi^2(1) = 6.44$, $p = 0.01$). Fixed effects of the model reveal that while mispronunciation conditions are significantly different from the correctly pronounced condition even with psychoacoustic distance added to the model, the accented condition is not significantly different (Table 2).

A linear regression was also conducted to test if psychoacoustic distance differed significantly by test condition with a random intercept of word. In this analysis, test condition was a significant predictor of psychoacoustic distance ($\chi^2(4) = 3209.88$, $p < 0.0001$), and each condition significantly differed in their psychoacoustic distance, with accented words being most distant from their SAE counterparts (Table 3). These results suggest that participants treated the accent condition differently from mispronunciation conditions, and while PLT for mispronunciation conditions varied as a function of psychoacoustic distance, PLT for accented words did not follow the same pattern, as PLT was higher despite greater distance for accented words. This finding in particular provides strong evidence for participants' incorporation of /i/ items into their existing /ɛ/ category during the experiment as a result of exposure to the similarly-accented sentential context. This demonstrates *pathway*-like directionality of change, since less acoustically

distant items that did not match the sentence accent were associated with more variable PLT.

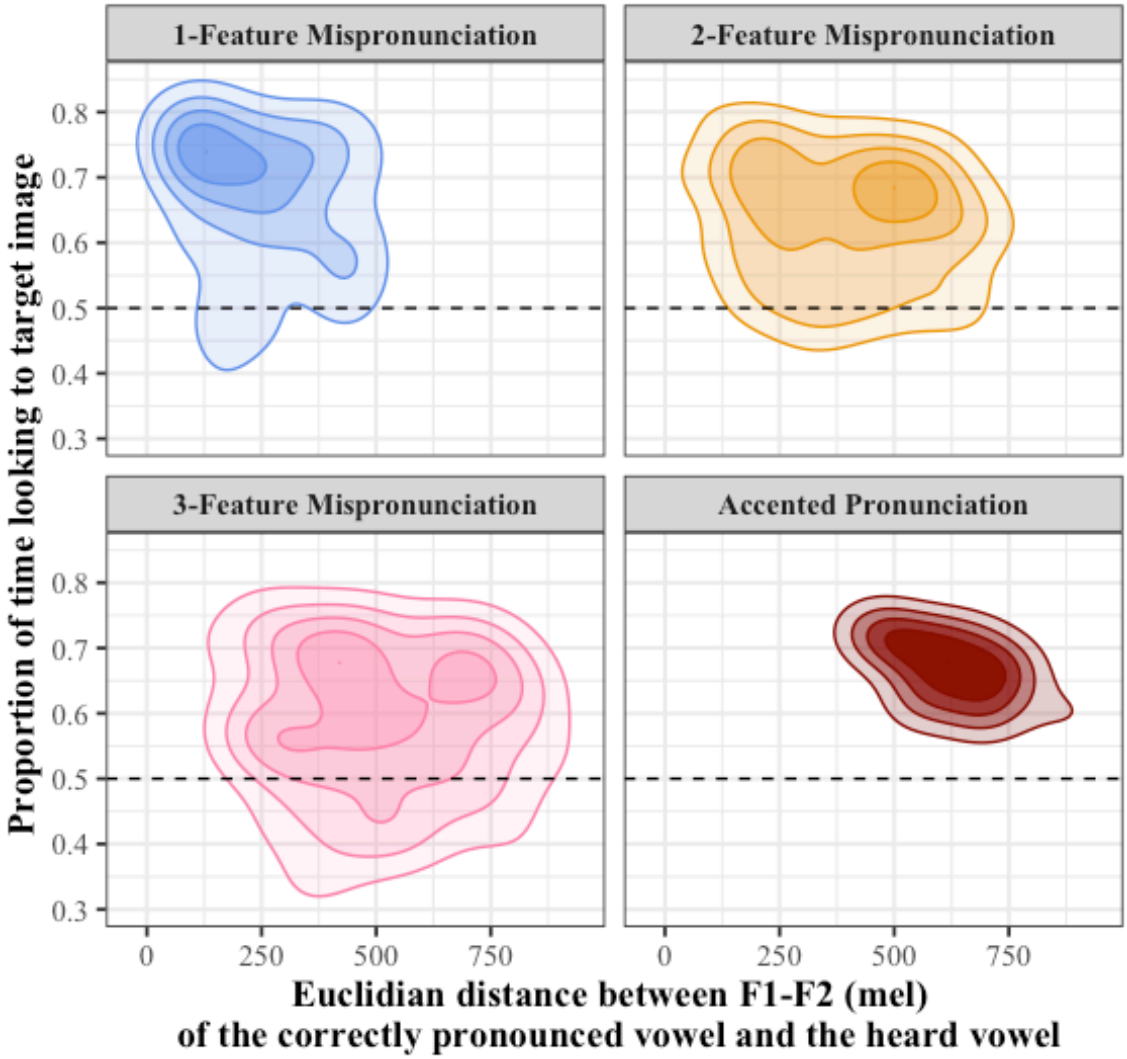

Figure 4: PLT as a function of F1-F2 distance of mispronounced/accented stimuli from correctly/ SAE pronounced stimuli. Dotted line is chance looking.

Table 2: Fixed effects of linear model with condition and psychoacoustic distance as predictors of PLT and random intercepts of subject and word

	Estimates	CI	t value	p value
(Intercept)	0.73	0.69 – 0.76	39.29	**<0.001** ***
1-feature	-0.05	-0.08 – -0.03	-3.94	**<0.001** ***

Table 2: Fixed effects of linear model with condition and psychoacoustic distance as predictors of PLT and random intercepts of subject and word

	Estimates	CI	t value	p value
2-feature	-0.06	-0.09 – -0.03	-4.03	**<0.001** ***
3-feature	-0.10	-0.14 – -0.07	-5.74	**<0.001** ***
accented	-0.02	-0.07 – 0.03	-0.94	0.345
psychacous dist	0.00	-0.00 – -0.00	-2.54	**0.011** .

Table 3: Fixed effects of linear model with condition as a predictor of psychoacoustic distance and a random intercept of word

Predictors	Estimates	CI	p
(Intercept)	0.18	-22.49 – 22.84	0.988
1-feature	257.23	243.97 – 270.48	**<0.001** ***
2-feature	390.42	377.13 – 403.70	**<0.001** ***
3-feature	511.01	497.74 – 524.28	**<0.001** ***
accented	615.82	568.76 – 662.88	**<0.001** ***

Changes across the experiment

An additional analysis was conducted to test how PLT may have changed across the course of the experiment. In this analysis, nested linear models were compared, and the best fitting model with PLT as the dependent variable included condition, trial number, and an interaction of these terms as fixed effects, with random intercepts of subject and word. ($\chi^2(5) = 46.237$, $p < 0.0001$). Additionally, in order to compare trials at equivalent times across the experiment, trials were also divided into 5 bins, with 25-26 trials in each bin (i.e., the first bin included trials 1-26 and the last bin included trials 103-127). Welch two-sample t-tests comparing first and last trial bins (the first 26 and

last 25 trials of the experiment) reveal significant increase in PLT for all conditions except for filler nonce words (correct: $t(177.36) = -2.30$, $p = 0.02$; 1-feature: $t(148.53) = -2.63$, $p < 0.01$; 2-feature: $t(195.7) = -3.08$, $p < 0.005$; 3-feature: $t(203.57) = -3.86$, $p < 0.0005$); accent: $t(163.51) = -2.82$, $p < 0.01$; filler: $t(375.5) = 0.753$, $p = 0.45$).

Conditions were also compared in bins 1, 3, and 5 using Welch two-sample t-tests (see Tables 3A-C). Results of these tests demonstrate similar initial starting and ending PLT for 1-feature mispronounced, 2-feature mispronounced, and accented items. 3-feature mispronounced items started off with PLT at chance, but PLT increased to similar

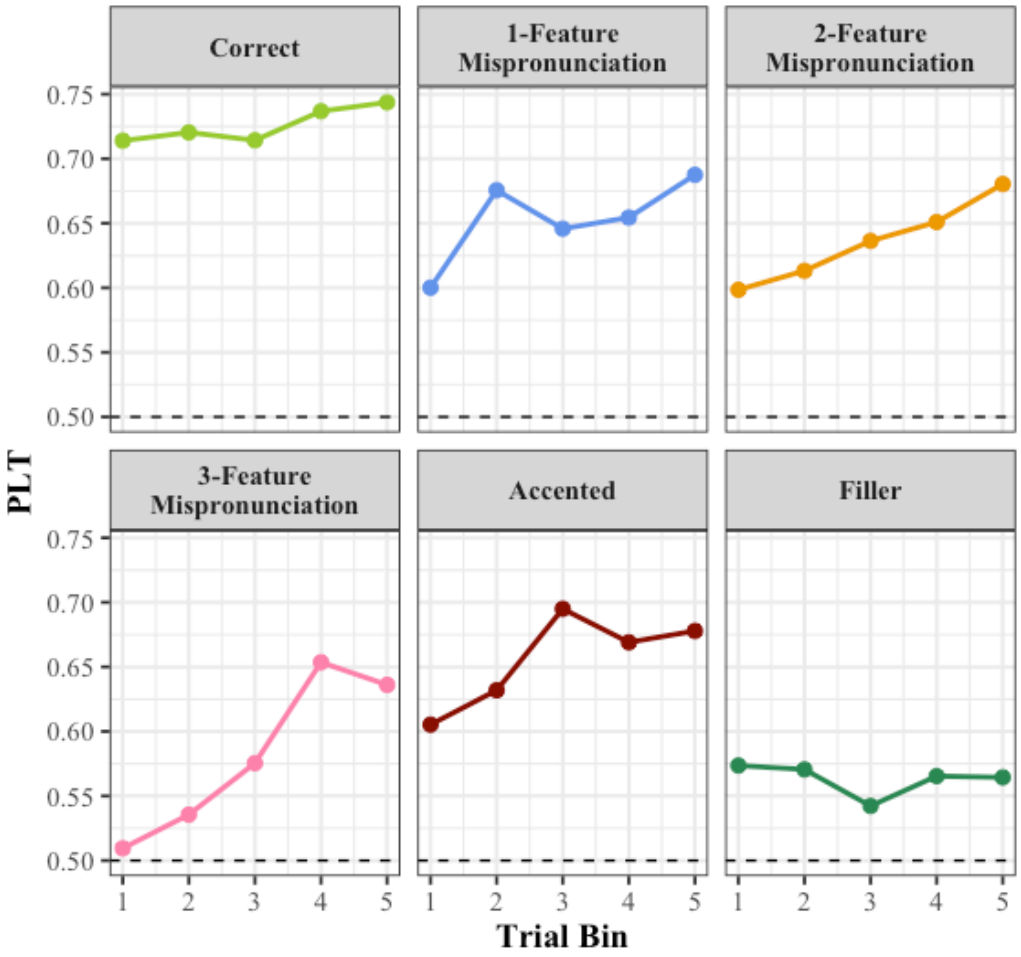

Figure 5: PLT across the course of the experiment, grouped into 5 bins of 25-26 trials. Dotted line = chance.

levels of the other test conditions. Across all time points, correctly pronounced items showed significantly greater looking times than the other conditions, except for a peak in PLT for accented items around the middle of the experiment, suggesting a rapid adaptation to accented words followed by a plateau at around 70% of time looking at the target, while PLT continued to increased for correct and mispronounced items.

Table 4A: t-tests for trial bin 1

	t value	df	p value
correct vs 1-feature	3.75	109.41	**< 0.0005** ***
correct vs 2-feature	5.01	151.70	**< 0.0001** ***
correct vs. 3-feature	7.78	150.59	**< 0.0001** ***
correct (SAE) vs accent	4.35	138.63	**< 0.0001** ***
1-feature vs 2-feature	0.29	172.96	0.77
1-feature vs 3-feature	2.75	191.32	**< 0.01** **
1-feature vs accent	0.05	178.19	0.96
2-feature vs 3-feature	2.77	218.94	**<0.01** **
2-feature vs accent	-0.25	206.60	0.80
3-feature vs accent	-2.90	216.97	**<0.005** **

Table 4B: t-tests for trial bin 3

	t value	df	p value
correct vs 1-feature	2.78	188.21	**< 0.01** **
correct vs 2-feature	3.29	201.01	**< 0.005** **
correct vs. 3-feature	4.60	165.72	**< 0.0001** ***
correct (SAE) vs accent	1.03	201.97	0.30
1-feature vs 2-feature	0.26	201.81	0.79

Table 4B: t-tests for trial bin 3

	t value	df	p value
1-feature vs 3-feature	2.06	191.30	**0.04 ***
1-feature vs accent	-1.84	194.66	0.07 .
2-feature vs 3-feature	1.93	181.15	0.06 .
2-feature vs accent	-2.27	207.36	**0.02 ***
3-feature vs accent	-3.78	171.11	**< 0.0005 *****

Table 4C: t-tests for trial bin 5

	t value	df	p value
correct vs 1-feature	2.90	184.25	**< 0.005 ****
correct vs 2-feature	3.12	156.32	**< 0.005 ****
correct vs. 3-feature	4.38	134.82	**< 0.0001 *****
correct (SAE) vs accent	3.91	195.58	**< 0.0005 *****
1-feature vs 2-feature	0.30	192.65	0.77
1-feature vs 3-feature	1.88	170.11	0.06 .
1-feature vs accent	0.47	204.53	0.64
2-feature vs 3-feature	1.58	168.99	0.12
2-feature vs accent	0.12	172.83	0.91
3-feature vs accent	-1.64	145.34	0.10

Finally, when compared to chance (0.5) in a one sample t-test, PLT was significantly above chance ($p < 0.001$) in all conditions except for 3-feature mispronunciations in trial bins 1 and 2 ($t(115) = 0.43$, $p = 0.67$; $t(86) = 1.25$, $p = 0.21$). This suggests that only 3-feature mispronunciations were divergent enough to fully disrupt lexical access, while other conditions did not.

Discussion

The current study examines how sensitive listeners are to vowel-based accent differences and vowel mispronunciations of familiar words, and how such mismatches in pronunciations affect lexical access. To demonstrate that the speaker had a non-standard accent, words were presented in the carrier sentence "Where's the X?" where the /ɛ/ in *where* was pronounced as [i]. The results indicate that participants treated the ɛ > i accented words differently than words containing vowels mispronounced along phonological dimensions of height, backness, and/or tenseness. Overall, participants looked more at accented items than mispronounced items, including 1-feature mispronunciations, which were closer to targets in terms of shared phonological features and psychoacoustic space than accented items. On the other hand, participants still looked less at accented items that correctly (SAE) pronounced items. These results suggest that while participants were able to use carrier sentence information about the speaker's accent in order to inform their looking behavior, the information was not enough to increase PLT to the level of the correctly pronounced SAE words.

This pattern best supports the *pathway* hypothesis, in which a category is broadened in a specific direction to accept new exemplars—lexical access is fastest and strongest for SAE items, but with some exposure to evidence of an alternative form, listeners can form a "pathway" between the new phonetic version of a lexical item and that item's phonological representation in the mind of the listener. Because PLT was greater for accented items than mispronounced items but less than SAE items, it is possible that a weak pathway between [i] and lexical items with /ɛ/ formed for

participants as they heard /wirz ðə/ in every trial. It is also possible that accented items were still looked at less than SAE items because mispronounced items disrupted stronger pathway formation—although participants heard evidence for an /ɛ/ > [i] accent, they also heard other items that did not follow this pattern, especially test words such as *desk*, *jet*, *check*, *chef*, *keg*, *gem*, *leg,* and *sled*, which were /ɛ/ words in the mispronunciation conditions, and therefore were never realized as [i].

On the other hand, because PLT increased for all mispronunciation conditions across the course of the experiment, participants may have also *expanded* their phonological representations of lexical items, as variability stemming from mispronunciations may have caused participants to allow even for words pronounced along three phonological dimensions as possible lexical candidates. Adaptation to accented words was more rapid across the course of the experiment, however, suggesting that representational space may have been changing in multiple ways—rapidly creating a pathway between new accented items and older, codified phonological forms while also slowly expanding generally.

Notably, only 3-feature mispronunciations reliably disrupted lexical access, and only for about the first half of the experiment. This shows that even moderate deviances from listeners' native vowel categories do not strongly impact lexical access, at least in the paradigm used in this experiment. Indeed, while participants *did* look more to unfamiliar images in nonce word filler trials (coded as the target image in these cases), they did not look at unfamiliar images at the same rate that participants looked at familiar images in any of the test conditions. Moreover, looking patterns for nonce items did not

change across the course of the experiment. Together, these observations provide evidence for an expansion of acceptable categories for word-medial vowels rather than increased knowledge about the experimental paradigm. Participants were reticent to assign words to unfamiliar items (with overall lower PLT for nonce fillers, although they are still above chance), which may have helped to drive the phonological expansion.

The interpretation of differing adaptation mechanisms for accented and mispronounced words is supported by the findings on the effects of psychoacoustic distance as well. While mispronunciation conditions followed trends of psychoacoustic distance, in which farther vowels also differed along more phonological features, which in turn also corresponded to looking time, accented words did not follow this pattern: accented words were farther in psychoacoustic measure from their SAE counterparts than even 3-feature mispronunciations (although accented words differed from SAE words along 2 phonological dimensions, height and tenseness), but had a significantly higher PLT than all mispronunciation conditions (which was especially apparent in time course analyses). Participants adapted more quickly and easily to accented words than to mispronunciations *despite* greater distance, suggesting the sentential context provided enough information for participants to reshape their phonological-lexical boundaries to include [i] as a potential candidate for words containing /ɛ/. If participants had only expanded their categories, there would not have been greater looking times for accented words than mispronounced words, and might have instead seen more similar looking patterns for accented words as mispronounced words.

The current experiment also demonstrates that the methods used in Experiment 2 were not sensitive enough to clearly detect accent adaptation, suggesting that the participants in Experiment 2 may have been reshaping their phonological representations, despite a lack of effects of accent condition. Thus, the fine-grained eye tracking measures used in this experiment provides a better measure of accent adaptation, and future work exploring phonological adaptation should employ this more sensitive measure to gain a better understanding of the mechanisms at work during adaptation to phonological variation.

While the current experiment showed in phonological representational space for vowels, it is not obvious that the same degree of expansion would be observed for consonants. Previous research on vowel and consonant perception has demonstrated that in many languages, including English, consonants and vowels affect lexical access differently (Cutler et al., 2000; van Ooijen, 1996; New, Araujo, & Nazzi, 2008; New & Nazzi, 2015; Oh et al., 2015; Franklin et al., in preparation). For example, in a study by Cutler et al. (2000), participants were asked to reconstruct real words from non-words by changing either one consonant or one vowel. Both Spanish-speaking and Dutch-speaking participants responded more quickly and accurately when they had to change vowels than when they had to change consonants, providing evidence that vowels may be more mutable than consonants in lexical items. In this case, it may be easier for listeners to adapt to vowel-based accent differences than consonant-based accent differences, although to my knowledge this hypothesis has not yet been directly studied.

While many accents in English differ mainly in vowel realizations, consonant differences are often more salient, possibly because consonants constrain lexical access more strongly, or possibly because listeners tend to perceive consonants more categorically, while they tend to perceive vowels more continuously (Ganong, 1980; Liberman et al., 1954; Fry et al., 1962; Pisoni, 1973; Kronod, Coppess, & Feldman, 2016). For example, speakers of rhotic varieties of English, like SAE, may not remember vowel shifts accurately in RP British English (see Exp. 1B), but they would likely remember differences final r pronunciations between their native variety and RP British English. Therefore, it could also be the case that participants adapt rapidly to accented items, but do not expand categories to mispronounced items to the same degree as vowels. More research is needed to elucidate mechanisms of consonant-based accent adaptation.

While participants' increased PLT for accented items may have been due to information about the speaker's accent in the sentential context, it is possible that the accented vowel shift tested here may have been ideal for adaptation, as it the direction of change went from less focal to more focal within the vowel space. The NRV framework (Polka & Bohn, 2003, 2011) suggests that listeners may have an underlying perceptual bias that favors vowels closer to the periphery of F1/F2 space. Although this bias has not been tested with the vowel shift used here or in the context of lexical access, it may be the case that lower-level perceptual biases may account for the effects reported in this paper. However, because accented, 1-feature mispronounced and 2-feature

mispronounced words all had similar looking times early on in the experiment, it is unlikely that NRV-based perceptual biases are driving the observed looking behaviors.

Additionally, due to COVID-19, I was not able to conduct a comparable study with a lowered /wærz/ sentence context or an unaccented (SAE-accented) sentence context. While the current study does provide compelling evidence that participants adapted to the accent heard in the sentence context, a future study with identical test words but an SAE sentence context or differently accented context would provide stronger evidence for accent adaptation. Although I interpret the current findings as two mechanisms simultaneously working to adapt to variable speech, it is possible that the rise in PLT over time for mispronounced items may have resulted, at least in part, from the knowledge that the speaker was not speaking in SAE. Because listeners received only a little information about the speaker's accent, they may have been able to quickly adapt to the shift they heard in the carrier sentence while generally expanding overall to account for other possible accent-related shifts. Indeed, a number of studies have shown that listeners were more likely to accept deviations in pronunciations when exposed to speech in which other deviations occurred (Weatherholtz, 2015; Witteman et al., 2010; Eisner et al., 2010; White & Aslin, 2011; Maye et al. 2008). This interpretation aligns with the *Shift and Expand* model posited by Hitczenko and Feldman (2016) and supported by ideal adapter model of speech adaptation put forth by Kleinschmidt & Jaeger (2015).

Conclusion

The experiment reported here demonstrates that native American English listeners are sensitive to changes in vowels arising both from accented speech and from shifts along phonological dimensions, but none of these changes reliably constrain lexical access. Overall, participants appear to be more likely to look at accented words than mispronounced words, suggesting that speech variability systematically due to speaker accent is treated differently by listeners during lexical processing.

Chapter 5: General Discussion

Through the experiments reported in this dissertation, I have tested how listeners remember familiar and adapt to novel accents. In Experiment 1A, participants demonstrated that they were familiar with a feature of Spanish-Accented English in which American English /ɪ/ is pronounced like /i/. Experiment 1B showed that although listeners are also familiar with a difference between Standard American English (SAE) and RP British English in which /æ/ is pronounced like /ɑ/, they have not retained the phonotactic rules for this alternation and will accept even inauthentic instances of /ɑ/ as more British-accented.

In Experiment 2, I tested if exposure to a novel accent would shift listeners' vowel categories such that they would accept similarly accented words in a lexical decision task. Experiment 2A measured differences between subjects and Experiment 2B measured differences within subjects across pre-exposure and post-exposure lexical decision tasks. I found little evidence for accent adaptation in this experiment overall, although Experiment 2B did provide some evidence that while participants did not shift or expand their categories to accommodate *more* for novelly-accented items, they also did not shrink their categories and reduce endorsement for phonological changes that matched the accent they had heard in exposure.

Finally, in Experiment 3, participants heard words that were either pronounced in SAE, had vowels mispronounced along one, two, or the phonological feature dimensions, or included an accented vowel that matched the novel accent in the carrier sentence "where's the…" This experiment demonstrated that listeners *were* able to adapt to a

novel accent with very little exposure, and although they looked less at accented items than SAE items, they looked more at accented items than mispronounced items. Participants also showed an increase in looking to mispronounced items over the course of the experiment, potentially providing evidence that exposure to an accented carrier sentence resulted in expansion of categories across the entire vowel space.

These studies were devised to test how phonological representations change as listeners encounter new, accented forms of words—do listeners *shift* their phonological categories towards a new mean or category prototype? Do they *expand* their categories to include more possible exemplars, including those for which they have no evidence? Or do participants maintain the boundaries of their original category but incorporate new forms selectively through a *pathway* between old and new exemplars?

Experiment 1 provided evidence for the *shift* hypothesis, and Experiments 2 and 3 provided evidence first and foremost for the *pathway* hypothesis. In Experiment 1A, participants correctly identified tokens of the word "fish" that contained /i/ to be Spanish-accented and also correctly did not identify tokens that contained /ɛ/ to be Spanish-accented. They also showed categorical effects, as they did not identify /ɪ/ tokens as more Spanish-accented than /ɛ/ tokens, although they are closer acoustically to /i/. Because participants did not endorse /ɪ/ tokens, it appears that within their representations of Spanish-accented English, the phonemic boundaries of the vowel in "fish" have shifted from /ɪ/ to /i/. Experiment 1B similarly supported evidence for a shift from /æ/ to /ɑ/ in Americans' representations of British-accented English, both when that shift is authentic (e.g., before voiceless fricatives) and when it is not (e.g., before stop consonants).

Experiment 1B does not provide evidence for or against the expand hypothesis, as only two vowels were tested. Together, the results of Experiment 1A and 1B demonstrate that at a pre-lexical level, listeners shift their phonemic categories to new means such that they overgeneralize to inauthentic environments, potentially dependent on the saliency of the shift. While these experiments provide compelling evidence for phonemic category *shifts* as an explanation of accent adaptation, Experiment 1 does not require word recognition for participants to successfully complete the task. While listeners may have shifted long-term representations of accented vowel spaces, different mechanisms may be at play during online lexical processing.

To test how accent-based shifts can affect lexical access, Experiment 2 sought to replicate a study by Maye, Aslin, and Tanenhaus (2008) in which participants were exposed to a novel, synthesized accent and their endorsement for SAE and accented words in a lexical decision task measured. Neither Experiment 2A nor 2B succeeded in replicating Maye et al.'s (2008) study, which demonstrated increased endorsement of accented words after exposure to a matching accent. In Experiment 2A, between-subject measures of word endorsement were not significantly different for participants who heard different accents during exposure, and in Experiment 2B, differences in pre- and post-exposure endorsement provided a small amount of evidence that participants who heard the accent in which front vowels were raised by one position in American English vowel space decreased their endorsement of words with lowered front vowels and did not decrease their endorsement of raised back vowels (as participants who heard the SAE and lowered accents did). Experiment 2B's findings suggest that some initial expansion of

phonological category representations may have occurred, but exposure to the speaker's accent caused representations to shrink asymmetrically. I concluded that failure to replicate Maye et al.'s results may have been due largely to increased task demands as compared to the original study (specifically that this study was conducted on MTurk and therefore listening environment was not lab-quality) as well as limited sensitivity of the lexical decision task for testing fine-grained distinctions. Additionally, our experiment did not test the same words twice, as Maye et al. did. The increased variability due to comparing different words in pre- and post-exposure lexical decision tasks may have also contributed to the current lack of results.

Therefore, in Experiment 3, I employed eye-tracking measures as well as a looking paradigm devised by White and Morgan (2008) in order to get a more sensitive measure of accent adaptation. In this experiment, participants looked more at images corresponding to 'accented' words that matched an accented carrier sentence than to images of 'mispronounced' words that did not match the carrier sentence. Additionally, looking to the target image increased over the course of the experiment for all of these conditions. These results support both *pathway* and *expansion* hypotheses: participants adapted more rapidly to the 'accented' words than mispronounced words, demonstrating detailed adaptation to that specific shift. Meanwhile, participants also gradually adapted to all mispronounced conditions, providing evidence for an overall expansion of phonological representation categories. The *shift* hypothesis, on the other hand, is not supported by these results, as participants looked most to target images for correctly pronounced words, and looking for these items increased across the experiment as well.

Although the results of these three studies support varying hypotheses, they can still be incorporated into a cohesive schema. Experiment 1 demonstrates that salient features of accents are represented at a pre-lexical level. When listeners encounter a new speaker or are faced with variability, they appear to expand categories, and, with increased speaker information (as in Experiment 2) narrow categories to eliminate unlikely variability, while they might continue to expand their categories when variability is more global or when speaker information is limited (as in Experiment 3). Finally, as listeners begin to form a representation of an accent's salient phonological features, they may "overlay" the accent representation on their native representation, allowing for a *pathway* between new and old representations to form during speech processing. However, because these experiments use different tasks, more empirical testing is needed to test whether this meta interpretation is correct.

It is also not yet clear how these processes interact at different levels. On one hand, listeners clearly remember features of somewhat familiar accents, and can access this information to make decisions about accentedness. For example, previous research demonstrates that listeners are able to group accented speakers along broad sociocultural dimensions (Bent, Atagi, Akbik, & Bonifield, 2016; Atagi & Bent, 2015; McCullough & Clopper, 2016). On the other hand, listeners in online processing are likely guided by lexical and semantic processing (Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Kraljic & Samuel, 2005) and are likewise sensitive to phonological distance between prototypical and divergent lexical forms. Previous work has also shown that listeners are good at rapidly adapting to novel accents that vary along a

number of category and cue dimensions (Baese-berk et al., 2013; Bradlow & Bent, 2008; Clarke & Garrett, 2004; Sidaras et al., 2009) and temporarily recalibrating their perceptual boundaries given new input (Eisner & McQueen, 2006; Reinisch & Holt, 2014; Vroomen et al., 2007). In Experiment 1 reported in this thesis, task performance was not modulated by reported experience with the test accent, and in Experiments 2 and 3, reported ease of understanding accented speech was also not predictive of performance, raising the possibility that listeners rely very little on long-term memories about accents during online perception of accented speech, and that processing accented speech may employ the same mechanisms as other types of speech variability.

The findings in this set of studies are consistent with conclusions by Kleinschmidt and Jaeger (2015), which state that an 'ideal adapter' uses a combination of flexible variance and flexible mean, and the weight of these two components is dependent on the situation, namely whether the listener has greater confidence in their prior beliefs about category variances or means (Kleinschmidt & Jaeger, 2015). It follows that listeners can dynamically weight cues during accent perception given both their prior knowledge and their updated beliefs about the informativity of these cues and change the shape of their categories accordingly. In the case of accented speech, this means that with some exposure to a novel accent, listeners can "hone in" on mismatches between their native categories and accented categories in real time, and except in the case of extreme divergences, can quickly and reliably adapt to accented speech. The research in this thesis adds to this claim, with the caveat that adaptation is not always 100% accurate—in Experiment 1B, participants accepted /bɑt/ as British-accented, and in Experiment 2B,

participants who heard raised front vowels in exposure showed greater endorsement of raised back vowels than participants who heard SAE or lowered front vowels, demonstrating a tendency to generalize to phonologically similar environments despite a lack of evidence for shifts or mismatches in those environments. In fact, previous work provides evidence for generalization across vowel space. For example, Sanker (under review) showed how exposure to shifted F1 or F2 in a single vowel can affect other vowels with matching characteristics (e.g., exposure to lowered /ɪ/ resulted in lowered perceptual shifts in the boundary between other high and mid vowels, and Chládková, Podlipsky, and Chionidou (2017) demonstrated that exposure to shifted perceptual boundaries for front vowels in Greek caused shifted perception of back vowels in test. Such research suggests that listeners bootstrap perception processes using phonetic and phonemic variability and apply their generalizations not only to new exemplars of the same vowel category, but also those of related categories.

Although this dissertation focuses on adult perception of accent, it is worthwhile to briefly address how such perceptual processes might develop. Early in linguistic development, babies appear to have rather limited abilities to contend with cross-speaker variation. For example, 7.5-month-olds trained on new word tokens in a male voice cannot recognize those tokens when they are produced by a female voice (Houston & Jusczyk, 2000) and word tokens trained with a happy affect are not recognized when produced in a neutral affect (Singh, Morgan, & White, 2004). Similarly, infants trained on new words in Spanish-accented (Schmale & Seidl, 2009) or Canadian-accented English (Schmale, Cristià, Seidl, & Johnson, 2010) are not able to recognize these words

in a Midwestern American English accent. These sorts of effects continue into childhood, with children up to seven demonstrating difficulty recognizing familiar words in unfamiliar accents (Nathan, Wells, & Donlan, 1998).

Conversely, other research has shown that toddlers, like adults, are able to accommodate to unfamiliar phonological variants, but their performance is context-dependent, namely with exposure to the accent prior to test (van Heugten & Johnson, 2014, 2016; White & Aslin, 2011), exposure to multiple accented speakers prior to test (Potter & Saffran, 2017), or magnitude of phonetic distinction (Escudero, Best, Kitamura, & Mulak, 2014; Franklin & Morgan, 2020), suggesting the ability to rapidly adapt to phonological variability begins to develop very early in life, although these abilities improve and refine into adulthood (Adank, Evans, Stuart-Smith, & Scott, 2007; Munro & Derwing, 1995). Moreover, five- to six-year-olds can successfully discriminate between their native accent and a non-native accent but not their native accent and another regional accent (e.g. Floccia, Butler, Girard, & Goslin, 2009; Girard, Floccia, & Goslin, 2008; Wagner, Clopper, & Pate, 2014; Jones, Yan, Wagner, & Clopper, 2017; Weatherhead, White, & Friedman, 2018), demonstrating that while children are somewhat sensitive the degree of mismatch between their native accent and another accent, their sensitivity is not as fine-grained or detailed as adults', although Weatherhead, Friedman, and White (2019) showed that four- and five-year-olds *can* distinguish their native accent from both regional and non-native foreign accents, even when only prosodic information was presented to the children. While it is evident that young listeners, like older listeners, can use variability in the speech signal to guide their

linguistic perception, it is not obvious whether they can flexibly *shift, expand,* or form *pathways* between their native prototypical categories and new exemplars dependent on the type of input, how much they rely on information in long-term memory to guide perception, or how much they generalize to similar forms compared to adults.

The present work raises a number of questions for future research. First, continued research should be conducted to elucidate how lexical access interacts with accent perception, as divergent phonological forms can affect word recognition (e.g., Franklin et al., in preparation) but lexical information can also guide phonological perception (e.g. Ganong, 1980). In other words, the relative effects of bottom-up and top-down processing of accented speech, as well as the level of representation (phonological or lexical) at which the majority of accent processing occurs, is not yet well known. Are listeners shifting phonological representations while expanding lexical ones?

Additionally, Experiment 1 showed that listeners remember aspects of familiar accents but falls short of demonstrating whether these memories affect lexical access. Previous research shows a benefit from exposure to accented speech in the short term (Bradlow & Bent, 2008) and a familiarity effect for strongly accented words (Witteman, Weber, & McQueen, 2013), but it may also be the case that such information is used to adapt more rapidly when encountering a familiar accent.

Finally, work using novel, synthesized accents should be conducted with toddlers and young children to get a better idea of how young listeners use various types of variability during speech processing. While natural accents can differ from a listener's native variety along a number of dimensions at once (including consonant and vowel

differences, prosody differences, and speech rate differences), synthesized accents can be controlled to carefully test a single dimension at a time. Through methods such as these, we can gain a better understanding of how systematic variability of a single dimension affects processing in the early stages of language learning, and thus outline a more complete trajectory of language learning.

Conclusion

In classical models of word recognition, it is often assumed that pre-lexical processing includes normalizing incoming speech to remove "irrelevant" variability, allowing for subsequent matching of phonological units in the signal and in the listener's representational system (e.g. Lahiri & Marslen-Wilson, 1991; McClelland & Elman, 1986; Norris, 1994). More recently, however, research has demonstrated that listeners *are* highly sensitive to variability in the speech signal, and can leverage the information it provides to streamline processing. The results of the studies in this dissertation further this claim, showing that at various levels of linguistic processing and in a variety of tasks, participants don't simply "hear through" variability, but remember it and use it to guide linguistic performance. Moreover, listeners appear to flexibly change their phonological representations given the type of accent exposure they receive and the type of information they must glean from it.

References

Adank, P., Evans, B., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of Familiar and Unfamiliar Native Accents Under Adverse Listening Conditions. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), 520-529. https://doi.org/0.1037/a0013552

Adank, P. & Janse, E. (2010). Comprehension of a novel accent by young and older listeners. *Psychology and Aging*, 25(3), 736-740. https://doi.org/10.1037/a0020054

Arnon, I. & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, 56, 349 - 371. https://doi.org/ 10.1177/0023830913484891

Atagi, E. & Bent, T. (2015). Relationship between listeners' nonnative speech recognition and categorization abilities. *The Journal of the Acoustical Society of America 137(1)*, EL44-EL50. https://doi.org/10.1121/1.4903916

Baese-Berk, M. M., Bradlow, A. R. & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America*, 133(2), EL174-EL180. https://doi.org/10.1121/1.4789864

Bailey, T. & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development*, 17(2), 1265-1282. https://doi.org/10.1016/S0885-2014(02)00116-8

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01

Bent, T., Atagi, E., Akbik, A., & Bonifield, E. (2016). Classification of regional dialects, international dialects, and nonnative accents. *Journal of Phonetics, 58,* 104-117. https://doi.org/10.1016/j.wocn.2016.08.004

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

Bent, T., Baese-Berk, M., Borrie, S. A., McKee, M. (2016). Individual differences in the perception of regional, nonnative, and disordered speech varieties. *The Journal of the Acoustical Society of America*, 140(5), 3775-3786. https://doi.org/10.1121/1.4966677

Blumstein, S (1986). On Acoustic Invariance in Speech. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and Variability in Speech Process* (178-201). Lawrence Erlbaum Associates, Inc.

Boersma, P. & Weenick, D. (1992). PRAAT: Doing phonetics by computer.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native. speech. *Cognition*, *106*(2), 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

Bradlow, A.R., Akahane-Yamada, R., Pisoni, D.B. & Tohkura, Y. Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics* 61, 977–985 (1999). https://doi.org/10.3758/BF03206911

Brown, V. A., Fox, N. P., & Strand, J. F. (2020, September 10). "Where Are The . . .Fixations?": Grammatical Number Cues Guide Anticipatory Fixations To Upcoming Referents And Reduce Lexical Competition. https://doi.org/10.31234/osf.io/pge7k

Chistovich, L. A. (1960). Classification Of rapidly repeated speech sounds. Akusticheskij Zhurnal, 6, 392–398. (Translated in Soviet Physics-Acoustics,

Chládková, K., Podlipsky, V. J., & Chionidou, A. (2017). Perceptual adaptation of vowels generalizes across the phonology and does not require local context. Journal of Experimental Psychology: Human Perception and Performance, 43(2), 414–427.

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

Clopper, C. & Bradlow, A. (2009). Free classification of American English dialects by native and non-native listeners. *Journal of Phonetics*, 37(4), 436-451. https://doi.org/10.1016/j.wocn.2009.07.004

Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language and speech*, *47*(Pt 3), 207–239. https://doi.org/10.1177/00238309040470030101

Clopper, C. G., & Pisoni, D. B. (2007). Free classification of regional dialects of American English. *Journal of Phonetics*, 35, 421–438. https://doi.org/10.1016/j.wocn.2006.06.001

Cohen Priva, U. (2015). Informativity affects consonant duration and deletion rates. *Laboratory Phonology, 6(2), 243–278. http://doi.org/10.1515/lp-2015-0008*

Cohen Priva, U. & Jaeger, T. F. (2018). The interdependence of frequency, predictability, and informativity in the segmental domain. *Linguistics Vanguard, 4*(s2), 1-18.

Cooper, A. & Bradlow, A. (2016). Linguistically guided adaptation to foreign-accented speech. *The Journal for the Acoustical Society of America*, 140(5), EL378-EL384

Cutler, A. (2012). Native Listening: Language Experience and the Recognition of Spoken Words. Cambridge, Massachusetts; London, England: The MIT Press. Retrieved May 22, 2021, from http://www.jstor.org/stable/j.ctt5hhjc1

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., &Van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition*, 28(5), 746-755. https://doi.org/10.3758/BF03198409

Cutler, A., Smits, R., & Cooper, N. (2005). Vowel perception: Effects of non-native language vs. non-native dialect. *Speech Communication, 47*, 32-42. https://doi.org/10.1016/j.specom.2005.02.001

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A. G., Taylor, K., and McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise- vocoded sentences. *Journal of Experimental Psychology*, 134(2), 222–241. https://doi.org/10.1037/0096-3445.134.2.222

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human*

Perception and Performance, 23(3), 914–927. https://doi.org/10.1037/0096-1523.23.3.914

Eisner, F. & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time (L). *Journal of the Acoustical Society of America, 119*(4), 1950-3. https://doi.org/10.1121/1.2178721

Eisner, F., Weber, A., & Melinger, A. (2010). Generalization of learning in pre-lexical adjustments to word-final devoicing. *The Journal of the Acoustical Society ofAmerica*, 128(4), 2323–2323.

Escudero, P., Best, C., Kitamura, C., & Mulak, K. (2014). Magnitude of phonetic distinction predicts success at early word learning in native and non-native accents. *Frontiers in Psychology, 5.* https://doi.org/10.3389/fpsyg.2014.01059

Evans, B. G., and Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. J*ournal of the Acoustical Society of America*, 115, 352–361. https://doi.org/10.1121/1.1635413

Evans, B. & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *The Journal of the Acoustical Society of America*, 121(6), 3814-3826. https://doi.org/10.1121/1.2722209

Flege, J., Bohn, O., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics,* 25(4), 437-470. https://doi.org/10.1006/jpho.1997.0052

Floccia, C., Butler, J., Girard, F., & Goslin, J. (2009). Categorization of regional and foreign accent in 5- to 7-year-old British children. *International Journal of Behavioral Development, 33(4),* 366-275. https://doi.org/10.1177/0165025409103871

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance*, 32(5), 1276-1293. https://doi.org/10.1037/0096-1523.32.5.1276

Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Delle Luche, C., Durrant, S., White, L., Goslin, J., Vihman, M. (2014). British English infants segment words only with exaggerated infantdirected speech stimuli. *Cognition*, 148, 1-9. https://doi.org/10.1016/j.cognition.2015.12.004

Fox, N. P., & Blumstein, S. E. (2016). Top-down effects of syntactic sentential context on phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance*, 42(5), 730–741. https://doi.org/10.1037/a0039965

Franklin L. & Morgan, J. L. (2020, July 6-9). Toddlers show adult-like sensitivity to consonants and vowels in early word representations [vICIS 2020 Congress Poster Presentation].

Franklin, L. R., Tin, S., Gasdaska, B., & Morgan, J. L. (in preparation). Consonantal and vocalic representation during lexical access, Brown University.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, 5(4), 171–189. https://doi.org/10.1177/002383096200500401

Ganong, W.F. (1980) Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception Performance,* 6, 110–125.

Gilbert, R., Chandrasekaran, B., and Smiljanic, R. (2014). Recognition memory in noise for speech of varying intelligibility. *Journal of the Acoustical Society of America*, 135(1). 389-399.

Girard, F., Floccia, C., & Goslin, J. (2008). Perception and awareness of accents in young children. *Developmental Psychology 26(3)*, 409-433. https://doi.org/10.1348/026151007X251712

Goldinger, S. (1998). Echoes of Echoes? An Episodic Theory of Lexical Access. Psychological Review, 105(2), 251-279. https://doi.org/10.1037/0033-295X.105.2.251

Greenspan, S. & Nusbaum, H., & Pisoni, D. (1988). Perceptual Learning of Synthetic Speech Produced by Rule. *Journal of experimental psychology: Learning, memory, and cognition,* 14, 421-33. https://doi.org/10.1037/0278-7393.14.3.421.

Grieser, D., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology, 25*(4), 577–588. https://doi.org/10.1037/0012-1649.25.4.577

Gureckis, T. M., Martin, J., McDonnell, J. et al. psiTurk: An open-source framework for conducting replicable behavioral experiments online. *Behav Res Methods,* 48, 829–842 (2016). https://doi.org/10.3758/s13428-015-0642-8

Harada, T. (2006). The acquisition of single and geminate stops by English-speaking children in a Japanese immersion program. Studies in Second Language Acquisition, 28, 601–632.

Havy, M., Serres, J., & Nazzi, T. (2014). A Consonant/Vowel Asymmetry in Word-form Processing: Evidence in Childhood and in Adulthood. *Language and Speech*, 57(2), 254–281. https://doi.org/10.1177/0023830913507693

Hitczenko, K. & Feldman, N. H. (2016). Modeling Adaptation to a Novel Accent. Proceedings of the 38th Annual Conference of the Cognitive Science Society.

Hochmann, J. R., Benavides-Varela, S., Nespor, M., & Mehler, J. (2011). Consonants and vowels: Different roles in early language acquisition. *Developmental Science*. https://doi.org/10.1111/j.1467-7687.2011.01089.x

Hosseinzadeh, N., Kambuziya, A., Shariati, M. (2015). British and American Phonetic Varieties. *Journal of Language Teaching and Research*, 6(3), 647-655. http://dx.doi.org/10.17507/jltr.0603.23

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. J*ournal of Experimental Psychology: Human Perception and Performance, 26(5),* 1570–1582. https://doi.org/10.1037/0096-1523.26.5.1570

Huang, Y. & Ferreira, F. (2020). The application of signal detection theory to acceptability judgements. *Frontiers in Psychology, 11*(73), 1-11. http://doi.org/10.3389/fpsyg.2020.00073

Jones, Z., Yan, Q., Wagner, L., & Clopper, C. (2017). The development of dialect classification across the lifespan. *Journal of Phonetics,* 60, 20-37. https:/doi.org/10.1016/j.wocn.2016.11.001

Kleinschmidt, D. F., and Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review,* 122(2), 148-203. https://doi.org/10.1037/a0038695

Kleinschmidt, D., Weatherholtz, K., & Jaeger, F. (2018). Sociolinguistic perception as inference under uncertainty. *Topics in Cognitive Science*, 10(4), 818-834. https://doi.org/10.1111/tops.12331

Kraljic, T., Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51(2), 141–178. https://doi.org/10.1016/j.cogpsych.2005.05.001

Kraljic, T., Samuel, A. G., Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, 194(4), 332–338. https://doi.org/10.1111/j.1467-9280.2008.02090.x

Kronrod, Y., Coppess, E., & Feldman, N. (2016). A unified account of categorical effects in phonetic perception. *Psychonomic Bulletin and Review*, 23(6), 1681-1712. https://doi.org/10.3758/s13423-016-1049-y

Lahiri, A., Gewirth, L., Blumstein, S. E. (1984). A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross‐language study. *The Journal of the Acoustical Society of America*, 76, 391. https://doi.org/10.1121/1.391580

Lahiri, A. & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition, 38*. 245-294. http://doi.org/10.1016/0010-0277(91)90008-R.

Liberman, A., Delattre, P., Cooper, F. & Gerstman, L. (1954). The role of consonant–vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs,* 68, 1–13.

Luce, P. A., & Lyons, E. A. (1998). Specificity of memory representations for spoken words. Memory and Cognition, 26(4), 708–715. https://doi.org/10.3758/BF03211391

Lüdecke, D (2020). sjPlot: Data visualization for statistics in social science. R package version 2.8.6, https://CRAN.R-project.org/package=sjPlot

Luthra, S., Correia, J. M., Kleinschmidt, D. F., Mesite, L. & Myers, E. B. (in press). Lexical information guides retuning of neural patterns in perceptual learning for speech. *Journal of Cognitive Neuroscience*.

Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance, 33*(2), 391–409. https://doi.org/10.1037/0096-1523.33.2.391

Masapollo, M., Polka, L., Ménard, L., Franklin, L., Tiede, M., & Morgan, J. (2018) Asymmetries in unimodal visual vowel perception: The roles of oral-facial kinematics, orientation, and configuration. *Journal of Experimental Psychology: Human Perception and Performance*, 44(7), 1103-1118. http://dx.doi.org/10.1037/xhp0000518

Masapollo, M., Zhao, C., Franklin, L. & Morgan, J. (2019). Asymmetric discrimination of nonspeech tonal analogues of vowels. J*ournal of Experimental Psychology: Human Perception and Performance, 45*(2), 285-300. https://doi.org/10.1037/xhp0000603

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, 32(3), 543–562. https://doi.org/10.1080/03640210802035357

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

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

McGarr, N. S. (1983). The Intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech, Language, and Hearing Research,* 26(3), 451-458. https://doi.org/10.1044/jshr.2603.451

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. https://doi.org/10.1037/0278-7393.31.2.306

McQueen, J. M., Cutler, A., Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, 30, 1113–1126. https://doi.org/10.1207/s15516709cog0000_79

Mehler, J., Peña, M., Nespor, M., & Bonatti, L. (2006). The ‘‘soul’’ of language does not use statistics: Reflections on vowels and consonants. *Cortex*, 42, 846-854. https://doi.org/10.1016/S0010-9452(08) 70427-1

Miller, G. A. and Nicely, P. (1955). An Analysis of Perceptual Confusions among some English Consonants, *Journal of the Acoustical Society of America,* 27(2), 1955.

Mirman, D. (2014). Growth Curve Analysis and Visualization Using R Analysis and Visualization Using R. Taylor & Francis Group, LLC.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics, 47*(4), 379–390. https://doi.org/10.3758/BF03210878

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Nathan, L., Wells, B., & Donlan, C. (1998). Children's comprehension of unfamiliar regional accents: a preliminary investigation. *Journal of child language*, *25*(2), 343–365. https://doi.org/10.1017/s0305000998003444

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition*, 98(1), 13–30. https://doi.org/10.1016/j.cognition.2004.10.005

Nazzi, T., Floccia, C., Moquet, B., & Butler, J. (2009). Bias for consonantal information over vocalic information in 30-month-olds: Cross-linguistic evidence from French and English. *Journal of Experimental Child Psychology*, 102(4), 522–537. https://doi.org/10.1016/j.jecp.2008.05.003

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

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. Lingue E Linguaggio, 2, 203–230.

New, B., Araùjo, V., & Nazzi, T. (2008). Differential processing of consonants and vowels in lexical access through reading. *Psychological Science*, 19, 1223-1227. https://doi.org/10.1111/j.1467-9280.2008.02228.x

New, B. & Nazzi, T. (2015). The time course of consonant and vowel processing during word recognition. *Language, Cognition, and Neuroscience*, 29(2), 147-157. https://doi.org/10.1080/01690965.2012.735678

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*(3), 189–234. https://doi.org/10.1016/0010-0277(94)90043-4

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

Nygaard, L. C., Burt, S. A., & Queen, J. S. (2000). Surface form typicality and asymmetric transfer in episodic memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1228–1244. https://doi.org/10.1037/0278-7393.26.5.1228

Nygaard, L. C. and Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics, 60*(3), 355-376. doi:10.3758/bf03206860

Oh, Y., Coupé, C., Marsico, E., & Pellegrino, F. (2015). Bridging phonological system and lexicon: Insights from a corpus study of functional load. *Journal of Phonetics, 53*, 153-176. https://doi.org/10.1016/j.wocn.2015.08.003

Pallier, C., Sebastian-Gallés, N., Dupoux, E., Christophe, A., Mehler, J. (1998). Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory and Cognition*, 26(4), 844-851. https://doi.org/10.3758/BF03211403

Park, H. (2013). Detecting foreign accent in monosyllables: The role of L1 phonotactics. *Journal of Phonetics*, 41, (78-87). http://dx.doi.org/10.1016/j.wocn.2012.11.001

Parker, M. & Borrie, S. A. (2018). Judgments of Intelligence and Likability of Young Adult Female Speakers of American English: The Influence of Vocal Fry and the

Surrounding Acoustic-Prosodic Context. *Journal of Voice*, 32(5), 538-545. https://doi.org/10.1016/j.jvoice.2017.08.002

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics,* 13(2), 253–60.

Polka, L., & Bohn, O. (2003). Asymmetries in vowel perception. Speech Communication, 41, 221–231. http://dx.doi.org/10.1016/S0167-6393 (02)00105-X

Polka, L. & Bohn, O. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetic*, 39(4), 467-478. https://doi.org/10.1016/j.wocn.2010.08.007

Potter, C. E., & Saffran, J. R. (2017). Exposure to multiple accents supports infants' understanding of novel accents. *Cognition, 166,* 67–72. https://doi.org/10.1016/j.cognition.2017.05.031

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Reinisch, E., & Holt, L. L. (2014). Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of experimental psychology. Human perception and performance*, *40*(2), 539–555. https://doi.org/10.1037/a0034409

Ren, J., & Morgan, J. (2011). Sub-segmental details in early lexical representation of consonants. *Proceedings of the 17th International Congress of Phonetic Sciences*

(ICPhS XVII), (August). Retrieved from http://titan.cog.brown.edu/research/infant_lab/Ren_sub segmental details early lexical.pdf

Sanker, C. (under review). *Perceptual learning and convergence are generalized phonologically across vowels.*

Scarborough, R. & Zellou, G. (2013). Clarity in communication: "clear" speech authenticity and lexical neighborhood density effects in speech production and perception. *The Journal of the Acoustical Society of America, 134*(5), 3793-3807. https://doi.org/10.1121/1.4824120

Schmale, R., Cristia, A., & Seidl, A. (2012). Toddlers recognize words in an unfamiliar accent after brief exposure. *Developmental Science*, 15(6), 732–738. https://doi.org/10.1111/j.1467-7687.2012.01175.x

Schwab, E. C., Nusbaum, H. C., & Pisoni, D. B. (1985). Some effects of training on the perception of synthetic speech. *Human Factors, 27*(4), 395–408.

Schwartz, J. L., Abry, C., Boë, L. J., Ménard, L., & Vallée, N. (2005). Asymmetries in vowel perception, in the context of the dispersion- focalization theory. Speech Communication, 45, 425–434. http://dx.doi .org/10.1016/j.specom.2004.12.001

Seyfarth, Scott. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. Cognition 133(1). 140–155. doi:10.1016/j.cognition.2014.06.013.

Shaw, J. A., Best, C. T., Docherty, G., Evans, B., Foulkes, P., Hay, J., & Mulak, K. E. (2018). Resilience of English vowel perception across regional accent variation. *Laboratory Phonology*, 9(1), 1-36. https://doi.org/10.5334/labphon.87

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, 125(5), 3306–16. https://doi.org/10.1121/1.3101452

Singh, L., Morgan, J. L., & White, K. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language, 51(2)*, 173-189. https://doi.org/10.1016/j.jml.2004.04.004

Sjerps, M. J., & McQueen, J. M. (2010). The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 36*(1), 195–211. https://doi.org/10.1037/a0016803

Skoruppa, K., Nevins, A., Gillard, A., & Rosen, S. (2015). The role of vowel phonotactics in native speech segmentation. *Journal of Phonetics, 49*, 67-76.

Stevens, K. N. (1989). On the quantal nature of speech. Journal of Phonetics, 17, 3–46. Stevens,

Strand, E. A. , & K. Johnson . (1996). Gradient and visual speaker normalization in the perception of fricatives. In D. Gibbon (Ed.), *Natural language processing and speech technology: Results of the 3rd KONVENS conference*, Bielefeld, October 1996 (pp. 14-26). Berlin: Mouton.

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

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

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, 52(3), 330–346. https://doi.org/10.1016/j.jml.2004.11.004

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76(2), 147-166. https://doi.org/10.1016/S0010-0277(00)00081-0

Toro, J. M., Nespor, M., Mehler, J., & Bonatti, L. (2008). Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychological Science*, 19, 137-144. https://doi.org/10.1111/j.1467-9280.2008.02059.x

Van Bezooijen, R., & Gooskens, C. (1999). Identification of language varieties: The contribution of different linguistic levels. *Journal of Language and Social Psychology*, 18(1), 31–48. https://doi.org/10.1177/0261927X99018001003

Van der Feest, S. V. H., Blanco, C. P., & Smiljanic, R. (2019). Influence of speaking style adaptations and semantic context on the time course of word recognition in quiet and in noise. *Journal of Phonetics*, 73, 158–177. https://doi.org/10.1016/j.wocn.2019.01.003

Van Engen, Kristin, Jasmine Phelps, Rajka Smiljanic and Bharath Chandrasekaran (2014). Sentence intelligibility in N-talker babble: effects of context, modality, and speaking style. *Journal of Speech, Language, and Hearing Research*. doi:10.1044/JSLHR-H-13-0076

van Heugten, M., & Johnson, E. K. (2014). Learning to contend with accents in infancy: Benefits of brief speaker exposure. *Journal of Experimental Psychology: General, 143*(1), 340–350. https://doi.org/10.1037/a0032192

van Heugten, M., & Johnson, E. K. (2016). Toddlers' Word Recognition in an Unfamiliar Regional Accent: The Role of Local Sentence Context and Prior Accent Exposure. *Language and Speech*, *59*(3), 353–363. https://doi.org/10.1177/0023830915600471

Van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition*, 24, 573-583.

Vroomen, J., van Linden, S., de Gelder, B., & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory-visual speech perception: Contrasting build-up courses. Neuropsychologia, 45(3), 572–7. doi: 10.1016/j.neuropsychologia.2006.01 .031

Wagner, L., Clopper, C. G., & Pate, J. K. (2014). Children's perception of dialect variation. *Journal of child language*, *41*(5), 1062–1084. https://doi.org/10.1017/S0305000913000330

Weatherhead, D., White, K., & Friedman, O. (2018). Children's accent-based inferences depend on geographic background. *Journal of Experimental Child Psychology, 175*, 108-116. https://doi.org/10.1016/j.jecp.2018.05.004

Weatherhead, D., Friedman, O., and White, K. (2019). Preschoolers are sensitive to accent distance. *Journal of Child Language, 46(6),* 1-15. https://doi.org/10.1017/S0305000919000369

Weatherholtz, K. (2015). *Perceptual learning of systemic cross- category vowel variation.* Doctoral dissertation, The Ohio State University.

Wells, J. C. (1982). *Accents of English Volume 2: An Introduction*. Cambridge: C.U.P.

White, K. & Aslin, R (2011). Adaptation to novel accents by toddlers. *Developmental Science*, 14(2), 372-384. https://doi.org/10.1111/j.1467-7687.2010.00986.x

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, 59(1), 114–132. https://doi.org/10.1016/j.jml.2008.03.001

White, K. S., Yee, E., Blumstein, S. E., & Morgan, J. L. (2013). Adults show less sensitivity to phonetic detail in unfamiliar words, too. *Journal of Memory and Language*, 68(4), 362–378. https://doi.org/10.1016/j.jml.2013.01.003

Witteman, M. J., Weber, A., & McQueen, J. M. (2010). Rapid and long-lasting adaptation to foreign-accented speech. *The Journal of the Acoustical Society of America*, 128(4), 2486. https://doi.org/10.1121/1.3508921

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, and Psychophysics,* 75**,** 537–556 (2013). https://doi.org/10.3758/s13414-012-0404-y

Witteman, M. J., Bardhan, N. P., Weber, A., McQueen, J. M. (2015) Automaticity and stability of adaptation to a foreign-accented speaker. *Language and Speech*, 58(2), 168-189. https://doi.org/10.1177/0023830914528102

Woods, K. J. P., Siegel, M. H., Traer, J., McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, and Psychophysics*, 79(7), 2064-2072. https://doi.org/10.3758/s13414-017-1361-2

Yan, Q. (2015). The perceptual categorization of Enshi Mandarin regional varieties. *Journal of Linguistic Geography, 3*(1), 1-19. https://doi.org/10.1017/jlg.2015.3

Appendix

Appendix A: Test items for Experiment 2A and 2B

Table A1: Test items for Experiment 2A

trained/ untrained	accent condition	orig vowel	heard vowel	word	phon
trained	R	a	æ	salt	sælt
trained	R	a	æ	sharp	ʃærp
trained	R	a	æ	giant	ʤæjənt
trained	SE	a	a	start	start
trained	SE	a	a	garden	gardən
trained	SE	a	a	bark	bark
trained	L	æ	a	glass	glas
trained	L	æ	a	back	bak
trained	L	æ	a	manner	manɚ
trained	R	æ	ɛ	catch	kɛʧ
trained	R	æ	ɛ	after	ɛftɚ
trained	R	æ	ɛ	Attack	ətɛk
trained	SE	æ	æ	cat	kæt
trained	SE	æ	æ	fan	fæn
trained	SE	æ	æ	scatter	skædɚ
trained	L	eɪ	ɛ	cake	kɛk
trained	L	eɪ	ɛ	place	plɛs
trained	L	eɪ	ɛ	afraid	əfrɛd
trained	R	eɪ	ɪ	change	ʧɪnʤ
trained	R	eɪ	ɪ	race	rɪs
trained	R	eɪ	ɪ	age	ɪdʒ
trained	SE	eɪ	eɪ	Table	teɪbəl

Table A1: Test items for Experiment 2A

trained/ untrained	accent condition	orig vowel	heard vowel	word	phon
trained	SE	eɪ	eɪ	bank	beɪŋk
trained	SE	eɪ	eɪ	strange	streɪndʒ
trained	L	ɛ	æ	help	hælp
trained	L	ɛ	æ	wet	wæt
trained	L	ɛ	æ	wretched	rætʃəd
trained	SE	ɛ	ɛ	Test	tɛst
trained	SE	ɛ	ɛ	best	bɛst
trained	SE	ɛ	ɛ	pepper	pɛpɚ
trained	R	ɛ	eɪ	shelf	ʃeɪlf
trained	R	ɛ	eɪ	french	freɪntʃ
trained	R	ɛ	eɪ	accept	əkseɪp
trained	SE	ɚ	ɚ	girl	gɚl
trained	SE	ɚ	ɚ	bird	bɚd
trained	SE	ɚ	ɚ	turn	tɚn
trained	L	i	ɪ	field	fɪld
trained	L	i	ɪ	key	kɪ
trained	L	i	ɪ	queen	kwɪːn
trained	SE	i	i	speak	spik
trained	SE	i	i	eat	iːt
trained	SE	i	i	tears	tiːrz
trained	L	ɪ	eɪ	sister	seɪstɚ
trained	L	ɪ	eɪ	shrink	ʃɻeɪŋk
trained	L	ɪ	eɪ	Inch	eɪntʃ
trained	R	ɪ	i	drink	dɻiŋk
trained	R	ɪ	i	Liquid	likwəd

Table A1: Test items for Experiment 2A

trained/ untrained	accent condition	orig vowel	heard vowel	word	phon
trained	R	ɪ	i	think	θiŋk
trained	SE	ɪ	ɪ	kitchen	kɪʧən
trained	SE	ɪ	ɪ	sit	sɪt
trained	SE	ɪ	ɪ	stick	stɪk
trained	L	o	ʌ	motion	mʌʃən
trained	L	o	ʌ	Nose	nʌz
trained	L	o	ʌ	door	dʌr
trained	SE	o	o	shore	ʃor
trained	SE	o	o	course	kors
trained	SE	o	o	Portion	porʃən
trained	R	o	ʊ	Home	hʊm
trained	R	o	ʊ	stove	stʊv
trained	R	o	ʊ	normal	nʊrməl
trained	SE	ɔ	ɔ	Wrong	rɔŋ
trained	SE	ɔ	ɔ	dog	dɔg
trained	SE	ɔ	ɔ	bottle	bɔdəl
trained	R	ɔ	ʌ	smaller	smʌlɚ
trained	R	ɔ	ʌ	frog	frʌg
trained	R	ɔ	ʌ	Knock	nʌk
trained	L	u	ʊ	Room	rʊm
trained	L	u	ʊ	reduce	rədʊs
trained	L	u	ʊ	Soup	sʊp
trained	SE	u	u	pool	pul
trained	SE	u	u	blue	blu
trained	SE	u	u	confuse	kənfjuz

Table A1: Test items for Experiment 2A

trained/ untrained	accent condition	orig vowel	heard vowel	word	phon
trained	L	ʊ	o	look	lok
trained	L	ʊ	o	soot	sot
trained	L	ʊ	o	bushes	boʃəz
trained	SE	ʊ	ʊ	Foot	fʊt
trained	SE	ʊ	ʊ	cook	kʊk
trained	SE	ʊ	ʊ	wolf	wʊlf
trained	R	ʊ	u	Wood	wud
trained	R	ʊ	u	sugar	ʃugɚ
trained	R	ʊ	u	sure	ʃur
trained	L	ʌ	ɔ	other	ɔðɚ
trained	L	ʌ	ɔ	huff	hɔf
trained	L	ʌ	ɔ	Trouble	trɔbəl
trained	SE	ʌ	ʌ	once	wʌns
trained	SE	ʌ	ʌ	covered	kʌvəd
trained	SE	ʌ	ʌ	rush	rʌʃ
trained	R	ʌ	o	become	bəkom
trained	R	ʌ	o	Gloves	glovz
trained	R	ʌ	o	dutchess	doʧəs
untrained	SE	a	a	star	star
untrained	SE	a	a	afar	əfar
untrained	SE	a	a	pipe	paɪp
untrained	R	a	æ	bounce	bæʊns
untrained	R	a	æ	father	fæðɚ
untrained	R	a	æ	dice	dæɪs
untrained	L	æ	a	map	map

Table A1: Test items for Experiment 2A

trained/ untrained	accent condition	orig vowel	heard vowel	word	phon
untrained	L	æ	a	flask	flask
untrained	L	æ	a	lantern	lantɚn
untrained	R	æ	ɛ	cab	kɛb
untrained	R	æ	ɛ	sack	sɛk
untrained	R	æ	ɛ	cracker	krɛkɚ
untrained	SE	æ	æ	cash	kæʃ
untrained	SE	æ	æ	badge	bædʒ
untrained	SE	æ	æ	candle	kændəl
untrained	L	eɪ	ɛ	cage	kɛdʒ
untrained	L	eɪ	ɛ	base	bɛs
untrained	L	eɪ	ɛ	baker	bɛkɚ
untrained	R	eɪ	ɪ	safe	sɪf
untrained	R	eɪ	ɪ	skater	skɪtɚ
untrained	R	eɪ	ɪ	await	əwɪt
untrained	SE	eɪ	eɪ	cave	keɪv
untrained	SE	eɪ	eɪ	tape	teɪp
untrained	SE	eɪ	eɪ	lace	leɪs
untrained	L	ɛ	æ	bell	bæl
untrained	L	ɛ	æ	keg	kæg
untrained	L	ɛ	æ	check	ʧæk
untrained	R	ɛ	eɪ	Hen	heɪn
untrained	R	ɛ	eɪ	jet	ʤeɪt
untrained	R	ɛ	eɪ	gem	ʤeɪm
untrained	SE	ɛ	ɛ	Shed	ʃɛd
untrained	SE	ɛ	ɛ	desk	dɛsk

Table A1: Test items for Experiment 2A

trained/ untrained	accent condition	orig vowel	heard vowel	word	phon
untrained	SE	ɛ	ɛ	sled	slɛd
untrained	L	i	ɪ	peas	pɪz
untrained	L	i	ɪ	cheese	ʧɪz
untrained	L	i	ɪ	teeth	tɪθ
untrained	SE	i	i	sheep	ʃip
untrained	SE	i	i	meat	mit
untrained	SE	i	i	beach	biʧ
untrained	L	ɪ	eɪ	clip	kleɪp
untrained	L	ɪ	eɪ	chips	ʧeɪps
untrained	L	ɪ	eɪ	mitten	meɪtən
untrained	R	ɪ	i	pig	pig
untrained	R	ɪ	i	gift	gift
untrained	R	ɪ	i	winter	wintɚ
untrained	SE	ɪ	ɪ	fist	fɪst
untrained	SE	ɪ	ɪ	ship	ʃɪp
untrained	SE	ɪ	ɪ	pin	pɪn
untrained	L	o	ʌ	hope	hʌp
untrained	L	o	ʌ	toast	tʌst
untrained	L	o	ʌ	sword	sʌrd
untrained	R	o	ʊ	stone	stʊn
untrained	R	o	ʊ	toad	tʊd
untrained	R	o	ʊ	toes	tʊz
untrained	SE	o	o	goat	got
untrained	SE	o	o	comb	kom
untrained	SE	o	o	soap	sop

Table A1: Test items for Experiment 2A

trained/ untrained	accent condition	orig vowel	heard vowel	word	phon
untrained	R	ɔ	ʌ	ball	bʌl
untrained	R	ɔ	ʌ	cloth	klʌθ
untrained	R	ɔ	ʌ	mop	mʌp
untrained	SE	ɔ	ɔ	clock	klɔk
untrained	SE	ɔ	ɔ	pot	pɔt
untrained	SE	ɔ	ɔ	tongs	tɔŋz
untrained	L	u	ʊ	juice	ʤʊs
untrained	L	u	ʊ	booth	bʊθ
untrained	L	u	ʊ	stoop	stʊp
untrained	SE	u	u	hoop	hup
untrained	SE	u	u	spoon	spun
untrained	SE	u	u	moon	mun
untrained	L	ʊ	o	hoof	hof
untrained	L	ʊ	o	put	pot
untrained	L	ʊ	o	woman	womən
untrained	R	ʊ	u	hook	huk
untrained	R	ʊ	u	bullet	bulət
untrained	R	ʊ	u	took	tuk
untrained	SE	ʊ	ʊ	hood	hʊd
untrained	SE	ʊ	ʊ	pull	pʊl
untrained	SE	ʊ	ʊ	good	gʊd
untrained	L	ʌ	ɔ	tub	tɔb
untrained	L	ʌ	ɔ	dove	dɔv
untrained	L	ʌ	ɔ	club	klɔb
untrained	R	ʌ	o	duck	dok

Table A1: Test items for Experiment 2A

trained/ untrained	accent condition	orig vowel	heard vowel	word	phon
untrained	R	ʌ	o	bus	bos
untrained	R	ʌ	o	mug	mog
untrained	SE	ʌ	ʌ	hug	hʌg
untrained	SE	ʌ	ʌ	jug	ʤʌg
untrained	SE	ʌ	ʌ	nut	nʌt
filler		i	i	nirp	nirp
filler		i	i	leeth	liθ
filler		i	i	reeted	ritəd
filler		ɪ	ɪ	fip	fɪp
filler		ɪ	ɪ	tib	tɪb
filler		ɪ	ɪ	ayif	əyɪf
filler		eɪ	eɪ	Keisher	keɪʃɚ
filler		eɪ	eɪ	Vetch	veɪtʃ
filler		eɪ	eɪ	frape	freɪp
filler		ɛ	ɛ	bep	bɛp
filler		ɛ	ɛ	nedʒ	nedge
filler		ɛ	ɛ	shen	ʃɛn
filler		æ	æ	zaften	zæftən
filler		æ	æ	hab	hæb
filler		æ	æ	tham	θæm
filler		a	a	daf	daf
filler		a	a	sall	Sal
filler		a	a	ras	ras
filler		u	u	poov	puv
filler		u	u	bule	bjul

Table A1: Test items for Experiment 2A

trained/ untrained	accent condition	orig vowel	heard vowel	word	phon
filler		u	u	wookən	wuk
filler		ʊ	ʊ	Rooses	rʊsəz
filler		ʊ	ʊ	thun	θʊn
filler		ʊ	ʊ	clug	klʊg
filler		o	o	corb	korb
filler		o	o	shomess	ʃomɛs
filler		o	o	dofe	dof
filler		ʌ	ʌ	luz	lʌz
filler		ʌ	ʌ	vud	vʌd
filler		ʌ	ʌ	fust	fʌst
filler		ɔ	ɔ	vont	vɔnt
filler		ɔ	ɔ	pov	pɔv
filler		ɔ	ɔ	coth	kɔθ
filler		ɚ	ɚ	jerd	ʤɚd
filler		ɚ	ɚ	gurt	gɚt
filler		ɚ	ɚ	shirps	ʃɚps

Table A2: Test items for 2B

accent condition	original vowel	heard vowel	word	phon
SE	æ	æ	badge	bæʤ
SE	æ	æ	candle	kændəl
SE	æ	æ	cash	kæʃ
SE	æ	æ	staff	stæf
L	ɛ	æ	bell	bæl
L	ɛ	æ	check	ʧæk

Table A2: Test items for 2B

accent condition	original vowel	heard vowel	word	phon
L	ɛ	æ	keg	kæg
L	ɛ	æ	press	præs
filler	æ	æ	hab	hæb
filler	æ	æ	prack	præk
filler	æ	æ	tham	θæm
filler	æ	æ	zaften	zæftən
SE	eɪ	eɪ	cave	keɪv
SE	eɪ	eɪ	fable	feɪbl
SE	eɪ	eɪ	fame	feɪm
SE	eɪ	eɪ	shame	ʃeɪm
R	ɛ	eɪ	gem	ʤeɪm
R	ɛ	eɪ	guess	geɪs
R	ɛ	eɪ	hen	heɪn
R	ɛ	eɪ	jet	ʤeɪt
L	ɪ	eɪ	chips	ʧeɪps
L	ɪ	eɪ	clip	kleɪp
L	ɪ	eɪ	mist	meɪst
L	ɪ	eɪ	mitten	meɪtən
filler	eɪ	eɪ	blaib	bleɪb
filler	eɪ	eɪ	frape	freɪp
filler	eɪ	eɪ	keisher	keɪʃɚ
filler	eɪ	eɪ	veitch	veɪtʃ
SE	ɛ	ɛ	death	dɛθ
SE	ɛ	ɛ	desk	dɛsk
SE	ɛ	ɛ	rest	rɛst
SE	ɛ	ɛ	shed	ʃɛd

Table A2: Test items for 2B

accent condition	original vowel	heard vowel	word	phon
R	æ	ɛ	cab	kɛb
R	æ	ɛ	cracker	krɛkɚ
R	æ	ɛ	grass	grɛs
R	æ	ɛ	map	mɛp
L	eɪ	ɛ	baker	bɛkɚ
L	eɪ	ɛ	cage	kɛdʒ
L	eɪ	ɛ	fake	fɛk
L	eɪ	ɛ	game	gɛm
filler	ɛ	ɛ	bep	bɛp
filler	ɛ	ɛ	getch	gɛtʃ
filler	ɛ	ɛ	nedge	nedʒ
filler	ɛ	ɛ	shen	ʃɛn
filler	ɚ	ɚ	gurt	gɚt
filler	ɚ	ɚ	jerd	dʒɚd
filler	ɚ	ɚ	shirps	ʃɚps
filler	ɚ	ɚ	sirl	sɚl
SE	i	i	east	ist
SE	i	i	teach	tiʧ
SE	i	i	weed	wid
SE	i	i	week	wik
R	ɪ	i	gift	gift
R	ɪ	i	pig	pig
R	ɪ	i	risk	risk
R	ɪ	i	winter	wintɚ
filler	i	i	keetch	kitʃ
filler	i	i	leeth	liθ

Table A2: Test items for 2B

accent condition	original vowel	heard vowel	word	phon
filler	i	i	nirp	nirp
filler	i	i	reeted	ritəd
SE	ɪ	ɪ	pin	pɪn
SE	ɪ	ɪ	print	prɪnt
SE	ɪ	ɪ	ship	ʃɪp
SE	ɪ	ɪ	wrist	rɪst
R	eɪ	ɪ	await	əwɪt
R	eɪ	ɪ	rate	rɪt
R	eɪ	ɪ	safe	sɪf
R	eɪ	ɪ	skater	skɪtɚ
L	i	ɪ	cheese	ʧɪz
L	i	ɪ	lean	lɪn
L	i	ɪ	peas	pɪz
L	i	ɪ	teeth	tɪθ
filler	ɪ	ɪ	ayif	əyɪf
filler	ɪ	ɪ	fip	fɪp
filler	ɪ	ɪ	splim	splɪm
filler	ɪ	ɪ	tib	tɪb
SE	o	o	load	lod
SE	o	o	pose	poz
SE	o	o	soap	sop
SE	o	o	toes	toz
R	ʌ	o	bus	bos
R	ʌ	o	duck	dok
R	ʌ	o	mug	mog
R	ʌ	o	sun	son

Table A2: Test items for 2B

accent condition	original vowel	heard vowel	word	phon
L	ʊ	o	butcher	bʊtʃɚ
L	ʊ	o	hoof	hof
L	ʊ	o	woman	womən
L	ʊ	o	wool	wol
filler	o	o	corb	korb
filler	o	o	dofe	dof
filler	o	o	shomess	ʃomɛs
filler	o	o	snoke	snok
SE	ɔ	ɔ	crop	krɔp
SE	ɔ	ɔ	flock	flɔk
SE	ɔ	ɔ	job	dʒɔb
SE	ɔ	ɔ	pot	pɔt
L	ʌ	ɔ	dove	dɔv
L	ʌ	ɔ	drug	drɔg
L	ʌ	ɔ	numb	nɔm
L	ʌ	ɔ	tub	tɔb
filler	ɔ	ɔ	coth	kɔθ
filler	ɔ	ɔ	pov	pɔv
filler	ɔ	ɔ	vont	vɔnt
filler	ɔ	ɔ	yotch	jɔtʃ
SE	u	u	hoop	hup
SE	u	u	moon	mun
SE	u	u	spoon	spun
SE	u	u	truth	truθ
R	ʊ	u	bull	bul
R	ʊ	u	hook	huk

Table A2: Test items for 2B

accent condition	original vowel	heard vowel	word	phon
R	ʊ	u	took	tuk
R	ʊ	u	woof	wuf
filler	u	u	bule	bjul
filler	u	u	looch	lutʃ
filler	u	u	poov	puv
filler	u	u	wookən	wuk
SE	ʊ	ʊ	hood	hʊd
SE	ʊ	ʊ	push	pʊʃ
SE	ʊ	ʊ	shook	ʃʊk
SE	ʊ	ʊ	stood	stʊd
R	o	ʊ	goat	gʊt
R	o	ʊ	hold	hʊld
R	o	ʊ	stoke	stʊk
R	o	ʊ	toad	tʊd
L	u	ʊ	juice	ʤʊs
L	u	ʊ	proof	prʊf
L	u	ʊ	soothe	sʊð
L	u	ʊ	stoop	stʊp
filler	ʊ	ʊ	clug	klʊg
filler	ʊ	ʊ	rooses	rʊsəz
filler	ʊ	ʊ	thun	θʊn
filler	ʊ	ʊ	tunt	tʊnt
SE	ʌ	ʌ	blood	blʌd
SE	ʌ	ʌ	club	klʌb
SE	ʌ	ʌ	stuff	stʌf
SE	ʌ	ʌ	tug	tʌg

Table A2: Test items for 2B

accent condition	original vowel	heard vowel	word	phon
R	ɔ	ʌ	ball	bʌl
R	ɔ	ʌ	cloth	klʌθ
R	ɔ	ʌ	hawk	hʌk
R	ɔ	ʌ	pod	pʌd
L	o	ʌ	hose	hʌs
L	o	ʌ	roll	rʌl
L	o	ʌ	sword	sʌrd
L	o	ʌ	zone	zʌn
filler	ʌ	ʌ	fust	fʌst
filler	ʌ	ʌ	luz	lʌz
filler	ʌ	ʌ	swut	swʌt
filler	ʌ	ʌ	vud	vʌd

Appendix B: All pairwise comparisons for Experiment 2A and 2B

Table B1: One-sample t-tests of dprime compared against mu = 0 for each condition, LDT type, front/back classification, and trained/untrained for Experiment 2A (between subject analysis).

Pre/Post-Story LDT	Story Accent	LDT Accent	Front/Back Vowel	t-value	95% CI min	p-value
Trained	**Standard**	**Standard**	**Front**	33.252	2.365	< 0.001 ***
Trained	**Standard**	**Raised**	**Front**	19.884	0.967	< 0.001 ***
Trained	**Standard**	**Lowered**	**Front**	12.361	0.777	< 0.001 ***
Trained	**Standard**	**Standard**	**Back**	33.205	2.237	< 0.001 ***
Trained	**Standard**	**Raised**	**Back**	18.396	1.196	< 0.001 ***
Trained	**Standard**	**Lowered**	**Back**	5.249	0.201	< 0.001 ***
Trained	**Raised**	**Standard**	**Front**	19.627	2.114	< 0.001 ***

Table B1: One-sample t-tests of dprime compared against mu = 0 for each condition, LDT type, front/back classification, and trained/untrained for Experiment 2A (between subject analysis).

Pre/Post-Story LDT	Story Accent	LDT Accent	Front/Back Vowel	t-value	95% CI min	p-value
Trained	**Raised**	**Raised**	**Front**	11.625	0.782	< 0.001 ***
Trained	**Raised**	**Lowered**	**Front**	11.267	0.714	< 0.001 ***
Trained	**Raised**	**Standard**	**Back**	17.266	1.924	< 0.001 ***
Trained	**Raised**	**Raised**	**Back**	9.597	0.862	< 0.001 ***
Trained	**Raised**	**Lowered**	**Back**	3.298	0.112	< 0.01 **
Trained	**Lowered**	**Standard**	**Front**	23.112	2.277	< 0.001 ***
Trained	**Lowered**	**Raised**	**Front**	21.089	1.061	< 0.001 ***
Trained	**Lowered**	**Lowered**	**Front**	13.587	0.996	< 0.001 ***
Trained	**Lowered**	**Standard**	**Back**	22.503	2.115	< 0.001 ***
Trained	**Lowered**	**Raised**	**Back**	16.649	1.223	< 0.001 ***
Trained	**Lowered**	**Lowered**	**Back**	7.228	0.395	< 0.001 ***
Untrained	**Standard**	**Standard**	**Front**	39.310	2.295	< 0.001 ***
Untrained	**Standard**	**Raised**	**Front**	11.451	0.575	< 0.001 ***
Untrained	**Standard**	**Lowered**	**Front**	6.736	0.334	< 0.001 ***
Untrained	**Standard**	**Standard**	**Back**	28.313	2.088	< 0.001 ***
Untrained	**Standard**	**Raised**	**Back**	7.530	0.346	< 0.001 ***
Untrained	**Standard**	**Lowered**	**Back**	7.792	0.425	< 0.001 ***
Untrained	**Raised**	**Standard**	**Front**	18.041	1.976	< 0.001 ***
Untrained	**Raised**	**Raised**	**Front**	7.492	0.454	< 0.001 ***
Untrained	**Raised**	**Lowered**	**Front**	3.927	0.164	< 0.001 ***
Untrained	**Raised**	**Standard**	**Back**	17.872	1.813	< 0.001 ***
Untrained	**Raised**	**Raised**	**Back**	7.120	0.320	< 0.001 ***
Untrained	**Raised**	**Lowered**	**Back**	4.389	0.204	< 0.001 ***
Untrained	**Lowered**	**Standard**	**Front**	23.787	2.238	< 0.001 ***
Untrained	**Lowered**	**Raised**	**Front**	10.322	0.537	< 0.001 ***

Table B1: One-sample t-tests of dprime compared against mu = 0 for each condition, LDT type, front/back classification, and trained/untrained for Experiment 2A (between subject analysis).

Pre/Post-Story LDT	Story Accent	LDT Accent	Front/Back Vowel	t-value	95% CI min	p-value
Untrained	**Lowered**	**Lowered**	**Front**	5.158	0.267	< 0.001 ***
Untrained	**Lowered**	**Standard**	**Back**	21.963	1.973	< 0.001 ***
Untrained	**Lowered**	**Raised**	**Back**	9.515	0.458	< 0.001 ***
Untrained	**Lowered**	**Lowered**	**Back**	8.759	0.493	< 0.001 ***

Table B2: One-sample t-tests of dprime compared against mu = 0 for each condition, LDT type, front/back classification, and pre/post-story task for Experiment 2B (within subject analysis).

Pre/Post-Story LDT	Story Accent	LDT Accent	Front/Back Vowel	t-value	95% CI min	p-value
Pre-Story	**Standard**	**Standard**	**Front**	23.426	1.991	<0.001 ***
Pre-Story	**Standard**	**Raised**	**Front**	4.913	0.314	<0.001 ***
Pre-Story	**Standard**	**Lowered**	**Front**	5.107	0.309	<0.001 ***
Pre-Story	**Standard**	**Standard**	**Back**	29.218	2.086	<0.001 ***
Pre-Story	**Standard**	**Raised**	**Back**	4.035	0.276	<0.001 ***
Pre-Story	**Standard**	**Lowered**	**Back**	5.976	0.378	<0.001 ***
Pre-Story	**Raised**	**Standard**	**Front**	23.748	1.925	<0.001 ***
Pre-Story	**Raised**	**Raised**	**Front**	5.074	0.266	<0.001 ***
Pre-Story	**Raised**	**Lowered**	**Front**	7.347	0.441	<0.001 ***
Pre-Story	**Raised**	**Standard**	**Back**	26.459	2.068	<0.001 ***
Pre-Story	**Raised**	**Raised**	**Back**	6.01	0.438	<0.001 ***
Pre-Story	**Raised**	**Lowered**	**Back**	6.461	0.457	<0.001 ***
Pre-Story	**Lowered**	**Standard**	**Front**	33.021	1.915	<0.001 ***
Pre-Story	**Lowered**	**Raised**	**Front**	2.793	0.138	<0.01 **
Pre-Story	**Lowered**	**Lowered**	**Front**	5.996	0.394	<0.001 ***
Pre-Story	**Lowered**	**Standard**	**Back**	28.239	1.981	<0.001 ***
Pre-Story	**Lowered**	**Raised**	**Back**	3.537	0.240	<0.001 ***

Table B2: One-sample t-tests of dprime compared against mu = 0 for each condition, LDT type, front/back classification, and pre/post-story task for Experiment 2B (within subject analysis).

Pre/Post-Story LDT	Story Accent	LDT Accent	Front/Back Vowel	t-value	95% CI min	p-value
Pre-Story	**Lowered**	**Lowered**	**Back**	6.127	0.415	<0.001 ***
Post-Story	**Standard**	**Standard**	**Front**	20.981	1.833	<0.001 ***
Post-Story	**Standard**	**Raised**	**Front**	5.640	0.275	<0.001 ***
Post-Story	**Standard**	**Lowered**	**Front**	5.467	0.292	<0.001 ***
Post-Story	**Standard**	**Standard**	**Back**	29.449	2.116	<0.001 ***
Post-Story	**Standard**	**Raised**	**Back**	2.452	0.061	0.011 *
Post-Story	**Standard**	**Lowered**	**Back**	5.547	0.358	<0.001 ***
Post-Story	**Raised**	**Standard**	**Front**	19.833	1.853	<0.001 ***
Post-Story	**Raised**	**Raised**	**Front**	5.900	0.460	<0.001 ***
Post-Story	**Raised**	**Lowered**	**Front**	3.519	0.152	<0.001 ***
Post-Story	**Raised**	**Standard**	**Back**	21.695	1.983	<0.001 ***
Post-Story	**Raised**	**Raised**	**Back**	5.095	0.365	<0.001 ***
Post-Story	**Raised**	**Lowered**	**Back**	7.006	0.518	<0.001 ***
Post-Story	**Lowered**	**Standard**	**Front**	15.696	1.756	<0.001 ***
Post-Story	**Lowered**	**Raised**	**Front**	4.663	0.295	<0.001 ***
Post-Story	**Lowered**	**Lowered**	**Front**	5.556	0.321	<0.001 ***
Post-Story	**Lowered**	**Standard**	**Back**	18.701	1.881	<0.001 ***
Post-Story	**Lowered**	**Raised**	**Back**	4.185	0.242	<0.001 ***
Post-Story	**Lowered**	**Lowered**	**Back**	5.701	0.437	<0.001 ***

Table B3: Front Vowel - pairwise comparisons across story accent conditions for standard LDT for Exp. 2A

	diff	lwr	upr	p adj
raised:trained-standard:trained	-0.002908099	-0.04127334	0.0354571412	0.9999356

Table B3: Front Vowel - pairwise comparisons across story accent conditions for standard LDT for Exp. 2A

	diff	lwr	upr	p adj
lowered:trained-standard:trained	-0.012390858	-0.04783339	0.0230516720	0.9190815
standard:untrained-standard:trained	-0.027472657	-0.06180337	0.0068580507	0.2017281
raised:untrained-standard:trained	**-0.048172938**	**-0.08643339**	**-0.0099124861**	**0.0045223**
lowered:untrained-standard:trained	**-0.036241432**	**-0.07161915**	**-0.0008637113**	**0.0409839**
lowered:trained-raised:trained	-0.009482759	-0.04881149	0.0298459753	0.9833753
standard:untrained-raised:trained	-0.024564558	-0.06289436	0.0137652463	0.4481126
raised:untrained-raised:trained	**-0.045264839**	**-0.08715086**	**-0.0033788200**	**0.0253444**
lowered:untrained-raised:trained	-0.033333333	-0.07260367	0.0059370060	0.1494350
standard:untrained-lowered:trained	-0.015081800	-0.05048597	0.0203223692	0.8298769
raised:untrained-lowered:trained	-0.035782080	-0.07500860	0.0034444392	0.0972283
lowered:untrained-lowered:trained	-0.023850575	-0.06027091	0.0125697604	0.4226813
raised:untrained-standard:untrained	-0.020700281	-0.05892520	0.0175246383	0.6355701
lowered:untrained-standard:untrained	-0.008768775	-0.04410806	0.0265705147	0.9810974
lowered:untrained-raised:untrained	0.011931506	-0.02723647	0.0510994785	0.9539451

Table B4: Front vowel words - pairwise comparisons across story accent conditions for raised LDT for Exp. 2A

	diff	lwr	upr	p adj
raised:trained-standard:trained	0.017941888	-0.07803911	0.11392288	0.9948381
lowered:trained-standard:trained	0.035715892	-0.05368307	0.12511485	0.8649520
standard:untrained-standard:trained	**-0.145564060**	**-0.23410313**	**-0.05702499**	**0.0000425**
raised:untrained-standard:trained	**-0.115751073**	**-0.21292545**	**-0.01857670**	**0.0090320**
lowered:untrained-standard:trained	**-0.149130438**	**-0.23929067**	**-0.05897021**	**0.0000370**
lowered:trained-raised:trained	0.017774004	-0.08003990	0.11558790	0.9954812
standard:untrained-raised:trained	**-0.163505949**	**-0.26053456**	**-0.06647733**	**0.0000240**
raised:untrained-raised:trained	**-0.133692962**	**-0.23866078**	**-0.02872514**	**0.0038835**
lowered:untrained-raised:trained	**-0.167072327**	**-0.26558249**	**-0.06856216**	**0.0000206**
standard:untrained-lowered:trained	**-0.181279952**	**-0.27180274**	**-0.09075717**	**0.0000002**
raised:untrained-lowered:trained	**-0.151466965**	**-0.25045215**	**-0.05248178**	**0.0001924**
lowered:untrained-lowered:trained	**-0.184846331**	**-0.27695537**	**-0.09273729**	**0.0000002**
raised:untrained-standard:untrained	0.029812987	-0.06839628	0.12802225	0.9545792
lowered:untrained-standard:untrained	-0.003566378	-0.09484106	0.08770830	0.9999976
lowered:untrained-raised:untrained	-0.033379365	-0.13305263	0.06629390	0.9318519

Table B5: Front vowel words - pairwise comparisons across story accent conditions for lowered LDT for Exp. 2A

	diff	lwr	upr	p adj
raised:trained-standard:trained	0.035019455	-0.059931393	0.12997030	0.9001394
lowered:trained-standard:trained	0.086441775	-0.001648409	0.17453196	0.0579927
standard:untrained-standard:trained	**-0.158497244**	**-0.244172404**	**-0.07282208**	**0.0000021**
raised:untrained-standard:trained	**-0.152480545**	**-0.248601001**	**-0.05636009**	**0.0000926**
lowered:untrained-standard:trained	**-0.168927913**	**-0.257069212**	**-0.08078661**	**0.0000008**
lowered:trained-raised:trained	0.051422319	-0.045897704	0.14874234	0.6598941
standard:untrained-raised:trained	**-0.193516699**	**-0.288656279**	**-0.09837712**	**0.0000001**
raised:untrained-raised:trained	**-0.187500000**	**-0.292144391**	**-0.08285561**	**0.0000051**
lowered:untrained-raised:trained	**-0.203947368**	**-0.301313662**	**-0.10658108**	**0.0000000**
standard:untrained-lowered:trained	**-0.244939019**	**-0.333232600**	**-0.15664544**	**0.0000000**
raised:untrained-lowered:trained	**-0.238922319**	**-0.337383812**	**-0.14046083**	**0.0000000**
lowered:untrained-lowered:trained	**-0.255369688**	**-0.346058231**	**-0.16468114**	**0.0000000**
raised:untrained-standard:untrained	0.006016699	-0.090290196	0.10232359	0.9999753
lowered:untrained-standard:untrained	-0.010430669	-0.098775248	0.07791391	0.9994301
lowered:untrained-raised:untrained	-0.016447368	-0.114954595	0.08205986	0.9969717

Table B6: Front vowel words - pairwise comparisons across story accent conditions for filler LDT for Exp. 2A

	diff	lwr	upr	p adj
raised-standard	0.036702421	-0.01858560	0.09199045	0.2646551
lowered-standard	-0.006009135	-0.05648921	0.04447094	0.9579245
lowered-raised	-0.042711556	-0.09905903	0.01363592	0.1772413

Table B7: Back vowel words - pairwise comparisons across story accent conditions for standard LDT for Exp. 2A

	diff	lwr	upr	p adj
raised:trained-standard:trained	0.001350093	-0.05528892	0.05798910	0.9999998
lowered:trained-standard:trained	-0.003988173	-0.05642070	0.04844436	0.9999344
standard:untrained-standard:trained	-0.027033366	-0.07789657	0.02382984	0.6541757
raised:untrained-standard:trained	-0.033842744	-0.09034147	0.02265598	0.5263920
lowered:untrained-standard:trained	-0.042153025	-0.09435321	0.01004716	0.1931137
lowered:trained-raised:trained	-0.005338266	-0.06349480	0.05281827	0.9998341
standard:untrained-raised:trained	-0.028383459	-0.08512919	0.02836227	0.7109253
raised:untrained-raised:trained	-0.035192837	-0.09704038	0.02665470	0.5834778
lowered:untrained-raised:trained	-0.043503119	-0.10145026	0.01444403	0.2665218
standard:untrained-lowered:trained	-0.023045192	-0.07559299	0.02950260	0.8116314
raised:untrained-lowered:trained	-0.029854571	-0.08787449	0.02816534	0.6852200
lowered:untrained-lowered:trained	-0.038164852	-0.09200781	0.01567811	0.3302843

Table B7: Back vowel words - pairwise comparisons across story accent conditions for standard LDT for Exp. 2A

	diff	lwr	upr	p adj
raised:untrained-standard:untrained	-0.006809379	-0.06341509	0.04979633	0.9993759
lowered:untrained-standard:untrained	-0.015119660	-0.06743562	0.03719630	0.9631373
lowered:untrained-raised:untrained	-0.008310281	-0.06612031	0.04949975	0.9985208

Table B8: Back vowel words - pairwise comparisons across story accent conditions for raised LDT for Exp. 2A

	diff	lwr	upr	p adj
raised:trained-standard:trained	-0.03150162	-0.13629686	0.07329363	0.9563783
lowered:trained-standard:trained	0.01528512	-0.08252512	0.11309536	0.9977886
standard:untrained-standard:trained	**-0.31165655**	**-0.40712021**	**-0.21619289**	**0.0000000**
raised:untrained-standard:trained	**-0.26225754**	**-0.36760057**	**-0.15691451**	**0.0000000**
lowered:untrained-standard:trained	**-0.28533108**	**-0.38286110**	**-0.18780107**	**0.0000000**
lowered:trained-raised:trained	0.04678673	-0.06077379	0.15434726	0.8167482
standard:untrained-raised:trained	**-0.28015494**	**-0.38558612**	**-0.17472375**	**0.0000000**
raised:untrained-raised:trained	**-0.23075592**	**-0.34520931**	**-0.11630254**	**0.0000002**
lowered:untrained-raised:trained	**-0.25382947**	**-0.36113523**	**-0.14652370**	**0.0000000**
standard:untrained-lowered:trained	**-0.32694167**	**-0.42543296**	**-0.22845038**	**0.0000000**

Table B8: Back vowel words - pairwise comparisons across story accent conditions for raised LDT for Exp. 2A

	diff	lwr	upr	p adj
raised:untrained-lowered:trained	**-0.27754266**	**-0.38563696**	**-0.16944836**	**0.0000000**
lowered:untrained-lowered:trained	**-0.30061620**	**-0.40111161**	**-0.20012079**	**0.0000000**
raised:untrained-standard:untrained	0.04939901	-0.05657667	0.15537469	0.7685926
lowered:untrained-standard:untrained	0.02632547	-0.07188754	0.12453848	0.9733605
lowered:untrained-raised:untrained	-0.02307354	-0.13091434	0.08476726	0.9903235

Table B9: Back vowel words - pairwise comparisons across story accent conditions for lowered LDT for Exp. 2A

	diff	lwr	upr	p adj
raised:trained-standard:trained	0.034385113	-0.067353613	0.13612384	0.9292160
lowered:trained-standard:trained	0.076950253	-0.017853919	0.17175442	0.1882745
standard:untrained-standard:trained	0.080067844	-0.012053916	0.17218960	0.1307187
raised:untrained-standard:trained	0.065182215	-0.037850039	0.16821447	0.4629913
lowered:untrained-standard:trained	0.096605846	0.001935507	0.19127619	0.0423536
lowered:trained-raised:trained	0.042565139	-0.061460798	0.14659108	0.8525051
standard:untrained-raised:trained	0.045682731	-0.055904588	0.14727005	0.7946520
raised:untrained-raised:trained	0.030797101	-0.080778926	0.14237313	0.9697310

Table B9: Back vowel words - pairwise comparisons across story accent conditions for lowered LDT for Exp. 2A

	diff	lwr	upr	p adj
lowered:untrained-raised:trained	0.062220733	-0.041683251	0.16612472	0.5265221
standard:untrained-lowered:trained	0.003117592	-0.091524079	0.09775926	0.9999990
raised:untrained-lowered:trained	-0.011768038	-0.117059407	0.09352333	0.9995635
lowered:untrained-lowered:trained	0.019655594	-0.077468555	0.11677974	0.9925166
raised:untrained-standard:untrained	-0.014885629	-0.117768379	0.08799712	0.9984720
lowered:untrained-standard:untrained	0.016538002	-0.077969606	0.11104561	0.9962129
lowered:untrained-raised:untrained	0.031423631	-0.073747251	0.13659451	0.9574821

Table B10: Back vowel words - pairwise comparisons across story accent conditions for filler LDT for Exp. 2A

	diff	lwr	upr	p adj
raised-standard	0.04533975	-0.02150557	0.11218508	0.2496509
lowered-standard	0.02256919	-0.03918801	0.08432640	0.6672370
lowered-raised	-0.02277056	-0.09164730	0.04610618	0.7180417

Appendix C: Supplementary Material for Experiment 3

Table C1: Stimuli with pronunciations in IPA

word	condition	correct	1-feature	2-feature	3-feature
badge	mispro	bædʒ	bɪdʒ	bʊdʒ	budʒ

Table C1: Stimuli with pronunciations in IPA

bag	filler	bæg			
bank	mispro	beɪŋk	bɛŋk	bɪŋk	bʊŋk
base	mispro	beɪs	bis	bɪs	bʊs
beach	mispro	bitʃ	beɪtʃ	bɛtʃ	bʌtʃ
bell	accent	bil			
belt	accent	bilt			
bip	filler	bɪp			
boot	filler	but			
brain	mispro	breɪn	brɛn	brɪn	brʊn
bule	filler	bjul			
bus	mispro	bʌs	bɛs	bɪs	bis
bush	mispro	bʊʃ	buʃ	biʃ	beɪʃ
cab	mispro	kæb	kɪb	kib	kub
cage	mispro	keɪdʒ	kodʒ	kudʒ	kʊdʒ
cash	mispro	kæʃ	kɛʃ	keɪʃ	kuʃ
cat	mispro	kæt	kɛt	kʊt	kut
cave	mispro	keɪv	kɛv	kɔv	kʊv
cent	accent	sint			
check	mispro	tʃɛk	tʃæk	tʃʊk	tʃuk
cheese	mispro	tʃiz	tʃɪz	tʃʊz	tʃʌz
chef	mispro	ʃɛf	ʃeɪf	ʃof	ʃuf
chest	accent	tʃist			
chips	mispro	tʃɪps	tʃɛps	tʃʌps	tʃops
clip	mispro	klɪp	klɛp	kleɪp	klop
clock	mispro	klɔk	klɛk	klæk	klik
cloth	mispro	klɔθ	klɛθ	klɪθ	kliθ
comb	mispro	kom	kɔm	kɛm	kɪm
corb	filler	korb			
corn	filler	korn			

Table C1: Stimuli with pronunciations in IPA

cot	mispro	kɔt	kɛt	keɪt	kit
cuffs	mispro	kʌfs	kɛfs	keɪfs	kifs
daf	filler	dæf			
deck	accent	dik			
desk	mispro	dɛsk	dɔsk	dʊsk	dusk
dog	mispro	dɔg	dɛg	dæg	dig
doll	filler	dɔl			
dove	mispro	dʌv	dʊv	duv	div
dress	accent	dris			
duck	mispro	dʌk	dʊk	dæk	dik
edge	accent	idʒ			
egg	accent	ig			
elf	accent	ɛ	i	ilf	
fan	mispro	fæn	fɛn	fʊn	fun
fence	accent	fins			
fip	filler	fɪp			
fish	mispro	fɪʃ	fiʃ	feɪʃ	foʃ
fist	mispro	fɪst	fʊst	fɔst	fost
frog	filler	frɔg			
gat	filler	gæt			
gauze	mispro	gɔz	goz	guz	giz
gem	mispro	dʒɛm	geɪm	dʒom	dʒum
gift	mispro	gɪft	gift	guft	goft
glove	mispro	glʌv	glɛv	gleɪv	gliv
goat	mispro	got	gut	gʊt	gɪt
hap	filler	hæp			
hat	filler	hæt			
hen	accent	hin			

Table C1: Stimuli with pronunciations in IPA

hog	mispro	hɔg	hog	hug	hig
hoof	mispro	hʊf	hɪf	hɛf	heɪf
hook	mispro	hʊk	huk	hik	heɪk
hoop	mispro	hup	hʊp	hʌp	hɛp
jet	mispro	dʒɛt	dʒʌt	dʒot	dʒut
jud	filler	dʒʊd			
jug	mispro	dʒʌg	dʒog	dʒug	dʒig
juice	mispro	dʒus	dʒos	dʒɔs	dʒæs
keg	mispro	kɛg	kʌg	kʊg	kug
kell	filler	kɛl			
knot	filler	nɔt			
lape	filler	leɪp			
leg	mispro	lɛg	lʌg	log	lug
lock	filler	lɔk			
map	mispro	mæp	mɛp	meɪp	mup
meat	mispro	mit	mut	mʊt	mɔt
moke	filler	mok			
mop	mispro	mɔp	mʊp	mɪp	mip
mosque	mispro	mɔsk	mɛsk	meɪsk	misk
moth	mispro	mɔθ	mʊθ	muθ	miθ
mug	mispro	mʌg	mog	mug	mig
neck	accent	nik			
nest	accent	nist			
net	accent	nit			
nirp	filler	nɚp			
peas	mispro	piz	pɪz	pæz	pʌz
pig	mispro	pɪg	pig	peɪg	pog
pin	mispro	pɪn	pʊn	pun	pon

Table C1: Stimuli with pronunciations in IPA

piv	filler	pɪv			
pot	mispro	pɔt	pot	peɪt	pit
reet	filler	rit			
roob	filler	rub			
rose	filler	roz			
sack	mispro	sæk	sɛk	sʊk	suk
safe	mispro	seɪf	sɛf	sæf	sʊf
sall	filler	sal			
sheep	mispro	ʃip	ʃup	ʃʊp	ʃʌp
shelf	accent	ʃilf			
shell	accent	ʃil			
ship	mispro	ʃɪp	ʃʊp	ʃup	ʃop
shome	filler	ʃom			
skate	mispro	skeɪt	skot	skʌt	skʊt
sled	mispro	slɛd	slæd	slʊd	slud
smoke	mispro	smok	smeɪk	smɛk	smɪk
spoon	mispro	spun	spin	speɪn	spɛn
steak	mispro	steɪk	stok	stuk	stʊk
steps	accent	stips			
stone	mispro	ston	stɔn	stʊn	stɪn
tape	mispro	teɪp	tip	tup	tʊp
teeth	mispro	tiθ	teɪθ	toθ	tɔθ
tent	accent	tint			
thoon	filler	θun			
tib	filler	tɪb			
toad	mispro	tod	teɪd	tid	tɪd
toast	mispro	tost	tust	tist	tɪst
toes	mispro	toz	tɔz	tɛz	tæz

Table C1: Stimuli with pronunciations in IPA

tongs	mispro	tɔŋz	toŋz	teɪŋz	tiŋz
top	filler	tɔp			
tub	mispro	tʌb	tob	teɪb	tib
vase	filler	veɪs			
verch	filler	vəɻtʃ			
vint	filler	vɪnt			
web	accent	wib			
weg	filler	wɛg			
wheel	filler	wil			
wolf	mispro	wʊlf	wɔlf	wælf	weɪlf
wood	mispro	wʊd	wʌd	wæd	weɪd
yif	filler	jɪf			
zaft	filler	zæft			

Table C2: Fixed effects of time course model comparing accented condition and 1-feature mispronunciation condition

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.88555	0.10728	8.255	< 0.0001 ***
ot1	23.32515	0.82674	28.213	< 0.0001 ***
ot2	-8.95613	0.48987	-18.283	< 0.0001 ***
accented	0.08130	0.01175	6.918	< 0.0001 ***
ot1:accented	4.20010	0.18899	22.224	< 0.0001 ***
ot2:accented	-0.33311	0.18663	-1.785	0.0743 .

Table C3: Fixed effects of time course model comparing accented condition and 2-feature mispronunciation condition

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.85201	0.11397	7.475	< 0.0001 ***
ot1	24.51322	0.76549	32.023	< 0.0001 ***

ot2	-6.86443	0.60085	-11.425	< 0.0001 ***
accented	0.21078	0.01176	17.929	< 0.0001 ***
ot1:accented	4.51455	0.19039	23.712	< 0.0001 ***
ot2:accented	-1.48669	0.18812	-7.903	< 0.0001 ***

Table C4: Fixed effects of time course model comparing accented condition and 3-feature mispronunciation condition

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	0.47409	0.11935	3.972	< 0.0001 ***
ot1	22.07700	0.61275	36.029	< 0.0001 ***
ot2	-5.65200	0.54363	-10.397	< 0.0001 ***
accented	0.58304	0.01151	50.643	< 0.0001 ***
ot1:accented	6.99226	0.18633	37.527	< 0.0001 ***
ot2:accented	-2.87835	0.18177	-15.835	< 0.0001 ***

Table C5: Fixed effects of time course model comparing 1-feature mispronunciation condition and 2-feature mispronunciation condition

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	0.91074	0.09970	9.135	< 0.0001 ***
ot1	23.17329	0.74692	31.025	< 0.0001 ***
ot2	-8.36977	0.48420	-17.286	< 0.0001 ***
2-feature	-0.11898	0.01126	-10.57	< 0.0001 ***
ot1:2-feature	-0.22137	0.18496	-1.197	0.231
ot2:2-feature	1.20670	0.18257	6.610	< 0.0001 ***

Table C6: Fixed effects of time course model comparing 2-feature mispronunciation condition and 3-feature mispronunciation condition

	Estimate	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	0.83341	0.10692	7.795	< 0.0001 ***
ot1	23.61546	0.50037	47.196	< 0.0001 ***
ot2	-6.93181	0.45741	-15.154	< 0.0001 ***
3-feature	-0.36427	0.01097	-33.217	< 0.0001 ***
ot1:3-feature	-2.38608	0.17904	-13.327	< 0.0001 ***

ot2:3-feature	1.36000	0.17359	7.834	< 0.0001 ***