

5-7-2016

Phonetic Retuning in Accented Speech: Is There a Role of Expectation?

Iliana Meza-Gonzalez

University of Connecticut - Storrs, iliana.meza-gonzalez@uconn.edu

Recommended Citation

Meza-Gonzalez, Iliana, "Phonetic Retuning in Accented Speech: Is There a Role of Expectation?" (2016). *Master's Theses*. 893.
https://opencommons.uconn.edu/gs_theses/893

This work is brought to you for free and open access by the University of Connecticut Graduate School at OpenCommons@UConn. It has been accepted for inclusion in Master's Theses by an authorized administrator of OpenCommons@UConn. For more information, please contact opencommons@uconn.edu.

Phonetic Retuning in Accented Speech:
Is There a Role of Expectation?

Iliana Meza González

B.A., California State University, Northridge, 2013

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the degree of

Master of Science

At the

University of Connecticut

2016

APPROVAL PAGE

Master of Science Thesis

Phonetic Retuning in Accented Speech: Is There a Role of Expectation?

Presented by

Iliana Meza González, B.A.

Major Advisor _____
Emily B. Myers

Associate Advisor _____
Rachel M. Theodore

Associate Advisor _____
Gerry Altmann

University of Connecticut

2016

Introduction

As listeners of spoken language, we are constantly faced with the challenge of perceiving a highly variable speech signal, which must somehow be transformed into an intelligible speech stream. The acoustics of these speech signals can vary greatly from one speaker to the next. Various factors such as age, sex (Xue, 1999) and vocal tract size (Cohen, Kamm & Andreou, 1995) can significantly impact the acoustic signature of an individual speaker. Other factors may vary even within a speaker, including speaking rate (Theodore, Miller, & DeSteno (2009), and affect (Williams & Stevens, 1972). Of interest is how listeners process this talker-specific variability. Idiosyncratic details of a talker's speech are accommodated by the listener in the face of extreme variability. Naturally, this leads to many questions about the mechanisms that underlie the extraordinary ability to take varying exemplars of words and phonemes and fit them into the established categories of known tokens within our native language. The question of how we perceive an intelligible and seemingly invariant speech stream despite the lack of invariance in the speech signal remains a big challenge. The specific kinds of cues we are attending to and the mechanism at work in this process are not fully understood.

To an average speaker who uses spoken language on a daily basis, this process might seem so effortless and automatic as to not be given a second thought. In reality, there is an enormous amount of variation from a number of sources that our perceptual system must take into account before delivering a clear, intelligible speech stream. Through all of this, one thing is for certain: listeners almost always *adapt* to relatively extreme variation -- meaning a speaker who is initially difficult to understand will become much more intelligible to a listener with enough time and experience (e.g. Bradlow & Bent, 2008). Recent studies have helped us understand the nuances of talker adaptation through exposure to accented speech and digitally-altered speech.

Adaptation to Accented Speech

One prime example of real-world talker adaptation is the recent research on listener intelligibility of foreign-accented speech. A foreign accent occurs as a result of second-language acquisition when the phonology of the speaker's native language (L1) imposes on their second language (L2) (Munro, 2008). From the perspective of a native listener, the accented speaker will produce many non-canonical pronunciations that may compromise intelligibility. The imposition of L1 phonology onto L2 can cause difficulty in intelligibility due to the non-standard acoustic production of phonemes in the second language (i.e., a lack of overlap between L1 and L2 phonology [e.g. Ladefoged & Maddieson, 1996]). In many cases, an accented speaker's productions will vary greatly even within a phoneme, and while no two utterances of the same phoneme will sound the same even in the context of a native speaker, these inconsistencies are present even more so in the context of a non-native speaker (Sidaras, Alexander & Nygaard, 2009). Many recent studies have focused on the question of *if* and *how* we learn to adapt to these non-native accents.

In one such study, Bradlow and Bent (2008), investigated whether adaptation to a novel Mandarin speaker's accent was dependent on the identity and number of Mandarin-accented speakers presented during exposure. One group of listeners was exposed to a single speaker while the other group was exposed to five speakers. Participants were tasked with transcribing a number of spoken sentences, from which an intelligibility score was calculated based on the correct identification of several keywords. The study found that exposing a listener to one speaker increased intelligibility ratings for that speaker, but not for a new speaker. This finding seemed to suggest that adjustment to accented speech may be talker-specific when exposed to a single speaker. In contrast, listeners who were exposed to multiple Mandarin-accented speakers achieved high intelligibility scores even on Mandarin-accented speakers they had never heard before. Bradlow and Bent explained that exposure to multiple talkers of the same accent might help listeners adjust their criteria for certain deviant phonemes across the accent

once they have learned the systematic deviations particular to that accent. It follows that those listeners who heard multiple talkers with the same non-native accent may have formed a kind of Mandarin-accented representation (due to the fact that this was the common feature that linked all the speakers) while the single-talker listeners did not. This Mandarin-accented representation (or any other accented representation, for that matter) might include disambiguating features of deviant phonemes that are initially confusing for a listener but are adapted to through lexically-driven listening experience. This representation is presumably what guides the speech perception of new talkers with the same accent. If this is the case, it may be that listeners need exposure to multiple talkers of an accent in order to differentiate what is talker-specific from what is accent-specific. This explanation offered by Bradlow and Bent, although telling, still leaves a number of uncertainties. Namely, the mechanism by which this adaptation is made possible remains unexplained. For instance, is adaptation the result of phonemic category shift (i.e. bottom-up information) or rather, lexical context (i.e. top-down information)?

Taking this one step further, Baese-Berk, Bradlow, & Wright (2013) also investigated the generalizability of accent adaptation by exposing listeners to either five speakers with the same accent or five speakers, each with a different accent. Results showed that those listeners who heard multiple accents (Thai, Korean, Hindi, Romanian and Mandarin) exhibited significantly higher intelligibility for a novel accent (Slovakian) than those who had heard a single accent (Mandarin). In some cases, a novel speaker's intelligibility scores from a multiple-accent-exposed participant equaled those of a participant who had been exposed to that speaker alone; in other words, hearing multiple speakers of an accent improves adaptation to a new speaker (of the same accent) to the same degree as exposure to that speaker alone. Baese-Berk and colleagues explain that this accent-independent adaptation may be due to sound characteristics of English that systematically cause similar difficulties among speakers of other languages. It is this systematic variation, then, which listeners learn to focus on when deciphering accented speech. In essence, the claim is that certain phonological properties of

non-native accents tend to be shared even across speakers of different language backgrounds. The ability to attend to these specific shared features would lead to better performance on new accents, assuming they also shared those same deviations.

Based on these explanations on how listeners adjust to accented speech, we might assume one of three hypotheses. First, listeners may relax or loosen their criteria for what counts as an acceptable phoneme of a particular category. In this case, listeners would rely heavily on top-down information from lexical context that allows them to accept and resolve even non-standard pronunciations. Second, listeners may be adjusting to these specific deviant properties of the accented speakers' acoustics. This is likely to occur if the adaptation mechanism uses low-level acoustic and phonetic cues to cope with variability in the speech signal. Third, listeners may be increasing their skill in a more general way, attending to the accent as whole without the need to make any phonemic adjustments. Perhaps the suprasegmental, phonological or other general characteristics of a person's speech create a top-down expectation that allows for the listener to consciously accept a greater amount of variability in any given number of phonemic categories. That is, the third possibility suggests that listeners simply become more skillful at using top-down cues to resolve any phonetic ambiguities in the input, without specifically adapting to the acoustic properties of that input.

While studies on accented speech have given us more insight on the talker specificity of adaptation to variability in speech, they give us little insight into the acoustic cues that we use to help us adapt. Because the variability found in accented speech spans a number of linguistic dimensions, it becomes even harder to pinpoint these potential acoustic cues. In order to better hone in on the specific cues and mechanisms of perceptual adaptation, the variability in the stimuli must be adequately controlled to appear only in the linguistic dimension to be tested and reduced elsewhere as much as possible. To this end, a number of researchers have used digitally-altered phonemes embedded in native-accented speech to better understand the perceptual learning that enables an intelligible speech stream.

Lexically-Guided Phonetic Retuning

An important step in understanding the stability of phonemic representations that makes an intelligible speech stream possible is to first understand the factors that can induce changes in these representations. Given that perception of a non-native accent involves learning a mapping between non-standard acoustic tokens and existing representations in one's native phonology, a clue for how this process may occur comes from a phenomenon which many researchers call *perceptual learning for speech* (e.g. Eisner & McQueen, 2005; Norris, McQueen, & Cutler, 2003; Samuel & Kraljic, 2009). The broad study of perceptual learning has yielded better understanding about the quality of exposure needed to cause a shift in our established phonemic categories as well as the extent and relative longevity of the shift (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Kraljic & Samuel, 2005).

It is important to note that phonetic categories must be -- somewhat paradoxically -- both reliably *stable* and highly *flexible*. The *stability* is necessitated internally -- by the linguistic brain's propensity to attribute specific meanings to specific phonetic sounds in order to give rise to meaning in spoken language. The *flexibility* is necessitated externally -- by the vast variability in speakers' acoustic signatures (e.g. sex, age, pitch, rate, etc.). We see evidence of our mind reconciling these two aspects of speech every day, as our brains are able to converge the acoustics of multiple utterances of the same word onto a single known token. One common way of exploiting this flexibility to induce perceptual learning in a speech context has been the use of ambiguous phonemes to induce lexically-guided *phonetic retuning*.

Studies on phonetic retuning commonly feature an ambiguous mixture of two phonemes (e.g. /s/ and /f/) inserted into word-medial or word-final positions. In isolation, these ambiguous phonemes will not sound quite like either of the original phonemes from which they were created, but rather an ambiguous blend with acoustic properties of both. The paradigm relies on the propensity of top-down lexical expectations to guide our perception of ambiguous

phonemes, such that we end up perceiving these phonemes as fully intact when they are featured as part of a real word (see Kraljic & Samuel, 1999 for a review; see Samuel, 1981 for evidence of lexically-guided phonemic restoration). Through this repeated exposure, a listener's will "learn" (through adaptation) to perceive the ambiguous phoneme as the original phoneme they were led to expect. This results in an observable shift in the phonetic boundary of the target phoneme, which almost always occurs without the listener's conscious awareness. This process is described with more detail in the following paragraphs.

A seminal study by Norris, McQueen, and Cutler (2003), first showed the lexically-guided shift in phonemic boundaries using this mechanism. Norris and colleagues exposed Dutch listeners to /s/- or /f/-final words in which the final phoneme had been replaced by a digitally-altered ambiguous phoneme (/ʔ/) halfway between /s/ and /f/. Participants were exposed to either an /s/-group in which the ambiguous phoneme, /ʔ/, replaced word-final /s/ or an /f/-group in which the same ambiguous phoneme replaced word-final /f/. After exposure, listeners were asked to categorize several trials of five tokens along an /ɛs/ - /ɛf/ continuum. Results showed that those in the /s/ condition categorized more sounds as /s/ while those in the /f/ condition categorized more sounds as /f/. These results point to a clear shift in one direction for each exposure group. Control conditions in this study showed that listeners who heard /ʔ/-final nonwords did not show a significant phoneme shift effect as did those who were exposed to real lexical items. With these findings, Norris and colleagues established the imperative role of lexical information as a driving factor in phonetic retuning, followed by others who have also failed to show a perceptual learning effect in the context of nonwords (e.g. Kraljic & Samuel, 2006; Eisner & McQueen, 2005).

Going a step further, Kraljic and Samuel (2005) demonstrated the importance of acoustic information (namely, spectral similarity between fricatives) in transferring these phonetic shifts from one talker to another. Using the target phonemes /s/ and /f/ Kraljic and Samuel replicated the methods and findings from Norris et al. (2003) and additionally conducted an experimental

condition which switched the gender of the voice from exposure to test. Listeners who were lexically exposed to a male voice were subsequently tested in a female voice and vice versa. The authors found that listeners who heard a female voice and tested on a male voice still showed a phonetic category shift while those who heard a male voice and tested on a female voice did not. Upon further analysis of the acoustic properties of the /s/ and /ʃ/ phonemes used in the study, the authors found that the spectral mean of the female exposure items (5432 Hz for /s/; 5383 Hz for /ʃ/) fell relatively close to that of the male test items (4943 Hz). This transfer of learning was particularly revealing because it seems counterintuitive that any perceptual shifts learned in the voice of one talker should ever generalize to the voice of another talker. Kraljic and Samuel explained that the similarity of acoustic information between exposure and test words is therefore likely to be one of the cues that helps drive and maintain phonetic retuning. Based on this information, Kraljic and Samuel's findings may point to a mechanism that relies on creating a bottom-up criterion for the perception of phonetic segments that require perceptual learning, rather than one that relies on building a talker identity that includes this characteristic.

Eisner and McQueen (2005) lent further support to the idea that acoustic (i.e. phonetic) similarity between the tokens that cause phonetic retuning and the tokens to which they generalize is a crucial reason behind the transfer of phonetic retuning within a speaker or across speakers. Following the methodology of Norris et al. (2003), Eisner and McQueen additionally included one experimental condition which switched the vowels in the test items with those of an unfamiliar talker while maintaining the fricatives from the talker used in exposure. Again, the authors found a robust phonetic retuning effect like the one seen in original test items, despite listeners reporting a noticeable change in speaker. Eisner and McQueen attributed this effect to the perceptual system's recognition of talker identity (in the fricatives) rather than the acoustic information itself. However, it is worth noting that the phonetic information of the shifted phoneme remained the same in both the exposure and test phases despite the change in vocalic acoustics; recall that results from Kraljic & Samuel (2005) pointed toward the potentially

primary role of phonetic similarity in regard to the generalization and transfer of phonetic retuning from one speaker to another. In the context of their study, Eisner and McQueen's "speaker identity" explanation may be inextricable from a "phonetic information" explanation. Yet, given such findings as those from Kraljic and Samuel (2005) above, explaining this study's transfer of phonetic retuning from one speaker to another through low-level acoustic information rather than talker identity would not be unlikely.

But while the two studies above added indispensable knowledge to our understanding of how spectral information in the speech signal affects perceptual changes of fricative phonemes, Kraljic and Samuel (2006) used stop consonants to demonstrate that phonetic retuning can transfer not only to new speakers, but also to new phonemes. This time, listeners were exposed to words whose /d/ or /t/ phonemes had been replaced with a halfway ambiguous phoneme. Listeners then categorized phonemes along a /d--/t/ continuum as well as a /b--/p/ continuum. The researchers chose these two pairs of phonemes since they are both differentiated by the same temporal cue (i.e. voice onset time) and are thus a comparable way to measure feature generalization. As expected, results showed that phonetic retuning occurred for the /d/ and /t/ phonemes and surprisingly, the effect also carried over to the previously unheard /b/ and /p/ phonemes. Results were significant for both the exposure voice and a new, previously unheard voice. The authors suggest that while these results may seem to contradict previous findings, it is probably the case that fricatives tend to be more talker-specific because they are spectrally differentiated (i.e. they are made phonetically distinct by their differing means and distributions of acoustic energy) and therefore carry more information about talker identity than stop consonants, which are temporally differentiated (i.e. they are made phonetically distinct by differences in voice onset time). These explanations are not mutually exclusive, but are in fact complementary to earlier research. Recall that explanations above considered the possibility that phonetic retuning is crucially driven by low-level acoustic information, such that shared acoustic similarity between tokens is likely to be a reliable predictor of whether the learning will

generalize, rather than talker identity. Kraljic and Samuel were able to point out just how critical these acoustic features are for retuning phonemic categories since their results were neither phoneme-specific, nor talker-specific, but rather were closely linked by the similarity of their phonetic *features*. This growing body of perceptual learning in speech literature has gradually added more and more clues toward the uncovering of the exact mechanism(s) that our speech perception system uses to cope with the natural variability in speech.

The Role of Expectation

Not to be overlooked in all of this is the role of a listener's expectation when it comes to the perception of speech and perceptual changes. Acoustic cues are not likely to be working alone when it comes to phonetic retuning or perceptual learning. For one, it is important to note that in Norris et al. (2003) and in many similar studies, the ambiguous phonemes are not inserted into word-initial positions. The importance of lexical cues preceding the ambiguous sound has been demonstrated by previous work. A study by Jesse and McQueen (2011) indicated that the retuning of phonetic categories was not significant for ambiguous sounds presented in word-initial positions. The authors explained that retuning may be dependent on the timing of the ambiguous sound relative to the disambiguating information (in this case, the lexical cue) such that ambiguous sounds heard before any lexical information has been presented, might not have sufficient basis on which to retune. As such, the expectation created as a speaker's production unfolds seems to be crucial in guiding a listener's perception of clear and unambiguous subsequent phonemes. This work suggests that lexical expectation must be engaged before the ambiguous phoneme in order for phonetic retuning to occur. This is not to be confused with other lexically-guided phenomena such as the Ganong effect, which does not induce a shift or recalibration of the phonetic boundary in the target ambiguous phoneme (see Ganong, 1980). This might be due to the location of the ambiguous phoneme; the Ganong effect is seen in word-initial phonemes while phonetic retuning is seen in word-medial and word-

final contexts.

However, expectations that arise at the lexical level may come easy to a listener when the speaker shares his or her native accent. Listening to a foreign-accented speaker may create a different kind of expectation -- a more generalized expectation that the speaker's production of words and phonemes will deviate substantially from one's own. It can be hypothesized that a large degree of difference in phonology between an accented speaker's native language and their second language might cause the listener to expect highly deviant pronunciations of phonemes and words. This expectation could potentially lead a listener to be more "lenient" as to which sounds they will accept as members of their native phonetic categories (i.e. accept a wider variation of sounds into established categories for an accented speaker than they would for a native speaker).

This potentially crucial component of the listener's expectation (particularly when it comes to a talker's non-native accent) is one that has thus far been overlooked in the phonetic retuning literature. However, expectation has been shown to affect other domains of speech perception including syntactic processing (Hanulikova, van Alphen, van Goch & Weber, 2012), and perception of mispronounced phonemes (Schmid & Yeni-Komshian (1999) so it would not be unfounded to hypothesize that expectation may affect perceptual learning down to the phonetic level. And while the phonetic retuning effect is generally regarded as automatic and preattentive due to its robustness, we ask whether a difference in the phonological productions of the speaker can lead to a greater, more generalized retuning effect in the listener.

The Current Study

The current study seeks to explore whether listener expectation might activate different mechanisms for adaptation to the variability in foreign accented speech than for native accented speech. Recently, Reinisch and Holt (2014) combined two traditionally separate lines of research -- 1) adaptation to accented speech and 2) phonetic retuning through ambiguous

phonemes. Reinisch and Holt used Dutch-accented English speech which contained ambiguous /s/-/f/ phonemes to test whether phonetic retuning would still occur, even in the context of accented speech. Results showed that listeners exhibited a robust retuning effect for both male and female Dutch-accented speakers of English. Additionally, the effect for the first female Dutch-accented speaker generalized, or transferred, to a new female speaker of the same accent even when listeners knew they were listening to two different speakers. At first, this may seem to stand in contrast to findings which have shown adaptation to fricatives to be talker-specific in specific contexts (e.g. Kraljic & Samuel, 2005), with speaker gender being particularly influential on a listener's perception of certain fricatives (Munson, 2011). Perhaps unsurprisingly, no transfer of phonetic retuning was seen from the female voice to the male voice in experiment two. Finally, listeners who heard a female voice and tested on the *male voice whose fricative tokens had been selected to be spectrally similar to the female fricatives* also showed a robust generalization of the effect. As previously shown in Kraljic and Samuel (2005), results of analyses in this experiment confirmed that phonemic category shifts transferred from the female speaker to the male speaker when their fricatives shared similar spectral properties. These findings lent support to the already established importance of the acoustic properties of speech, but what was new about Reinisch and Holt's study is that retuning of phonetic categories still occurred even in the context of accented speech. This new finding indicated that lexically-guided retuning may not just be a driving mechanism for shifts induced by ambiguous sounds, but by global accents as well.

At this point, it is important to note that shared phonology between languages could potentially play a crucial role in listener adaptation. A key to understanding how we cope with the variability in a native accent as opposed to a foreign accent may lie in the phonological and suprasegmental features that the two languages share. All of the studies in the field of perceptual learning in speech that have been summarized herein have recruited either American or Dutch listeners and exposed them to speech in their native accent (in the case of

Reinisch and Holt, it was Dutch-accented English). Although in this recent study Reinisch and Holt used a speaker whose accent was not easily identifiable to the listeners, English and Dutch both share many important features including phonology, stress-timed segmentation, and lexical roots (Roach, 1982; Cambier-Langeveld & Turk, 1999; Johnson & Babel, 2010). One study conducted by Sebastián Gallés, Dupoux, Costa & Mehler (2000), found the shared phonology between two languages was the critical factor in cross-linguistic adaptation. Using monolingual Spanish listeners, they showed that only exposure to Italian was as beneficial as exposure to Spanish in increasing the accuracy of syllable identification in time-compressed Spanish sentences. Furthermore, Greek (which shares many phonological and rhythmic elements with Spanish) also significantly increased Spanish monolinguals' syllable identification accuracy to the same degree as listening to Spanish and Italian (e.g. Fourakis, Botinis & Katsaiti, 1999). This finding led the authors to conclude that shared phonology and rhythmic patterns are the crucial factor most likely responsible for cross-linguistic adaptation.

Given these potentially large or small differences between a speaker's native phonology and their second language phonology, our present study aims to approach Reinisch and Holt's questions from a new perspective. We aim to investigate if a listener's expectation (given a speaker's accent and the effect of a differing L1-L2 phonology) deploys different mechanisms to cope with the variability in foreign-accented speech than for native-accented speech. If so, this will manifest as a measurable effect on a listener's phonemic category shifts; namely, foreign accents that result from little overlap in phonological features should lead to measurably larger phonetic retuning effects than those seen in native accent contexts. In short, we hypothesize that our perceptual system will treat deviant productions from native and foreign accents differently, which will present as a measurably larger shift in phonetic category boundary. We expect a larger retuning effect in listeners who are exposed to foreign accented speech as a result of the increased variation in their phonetics (as attributable to the differing phonological inventory of the speaker's native language).

The possibility exists that the phonetic category boundary shifts may look different if the added variation of the accent can be attributed to an “external” source. For instance, Kraljic and Samuel (2011) demonstrated that perceptual learning for a speaker’s deviant pronunciations effectively did not occur if the source of the variation could be attributed to a source other than the speaker’s inherent idiosyncracies (in that study, a pen in the speaker’s mouth was responsible for the deviant productions). Will the presence of a foreign accent be treated as an “external” source of variability, thereby blocking the ability to adjust to a speaker’s productions? Or rather, will the speaker’s accent be treated as an extension of her idiosyncratic attributes?

On the flipside, a listener might be more forgiving of an accented speaker and “loosen” their categorical representations to allow a wider array of tokens as acceptable category members. The possibility that a more phonologically variable accent may simply broaden the criteria for what is an acceptable member of a phonemic category also exists. In contrast to the typical category shifts seen in previous behavioral studies (e.g. Norris et al., 2003; Kraljic & Samuel, 2005), in which category boundaries are graphically represented as abrupt and steep after shifting, a generalized loosening of the target phonetic category might appear less clear and less steep. A listener who adapts by categorizing several tokens along the continuum without consistency (as a direct result of the inconsistency in the accented productions) effectively makes it difficult to pinpoint and identify a clear category boundary. This “loosened” categorization could arise as a general pattern across the accent in order to better accommodate the variability of the accented speakers. In this case, phonetic retuning may not be limited to the specific phonetic information one is exposed to, but may also be greatly influenced by the more global characteristics of the phonological and prosodical information in the accent.

Recall that perceptual adaptation in speech seems to consistently occur in specificity to a particular accent or deviation. For example, when listeners are exposed to multiple speakers of one accent, the intelligibility for that accent increases but does not transfer to a new accent

(Bradlow & Bent, 2008; Baese-Berk et al., 2013), and when listeners are exposed to an ambiguous phoneme, the shift of the feature is systematic and directional even if it transfers to new speakers and/or phonemes (Kraljic & Samuel, 2005; 2006; Eisner & McQueen, 2005). Given this knowledge, how might the combination of talker expectation (the expectation of a native-accented talker versus a foreign-accented talker) and ambiguous phonemes embedded in an already deviant phonological context affect phonemic category boundary shifts in accented speech?

In order to address these questions, we conducted a series of three experiments. In Experiment 1, we compared the phonetic retuning effects of a foreign-accented male speaker (Brazilian Portuguese) to a native-accented male speaker (American English) to test the hypothesis that foreign-accented speech would result in a larger degree of phonetic retuning (the ‘accent expectation’ hypothesis). Experiment 2 sought to test the same hypothesis as Experiment 1 using female voices instead of male voices and a different target phoneme pair. This also served as a way to better understand the generalizability of the results from Experiment 1. Finally, Experiment 3 addressed the question of whether low-level acoustic information is a more crucial factor in phonetic retuning (the ‘no expectation’ hypothesis) or whether listeners rely more on the expectation of a foreign-accented speaker.

Experiment 1

In Experiment 1 we exposed listeners to one of two male speakers: a native speaker of American English (AE) or a native speaker of Brazilian Portuguese (BP) with a clearly detectable accent. As explained above, the phonological inventories of languages used in previous studies of this type have been highly similar (i.e. English and Dutch), along with other linguistic dimensions (e.g. Roach, 1982; Johnson & Babel, 2010). Brazilian Portuguese, on the other hand, is a Romance language with fewer vowels (only seven; Frota & Vigarrio, 2001), and different prosody (mixed -- syllable and mora; Frota & Vigarrio; 2002) than English or Dutch.

In Experiment 1 we measured and compared the degree of phonetic retuning between the AE and BP exposure groups, using /s/ and /f/ as our target phonemes. We hypothesized that listeners in the BP groups would show a greater degree of retuning (i.e. larger boundary shift of /s/ or /f/) when directly compared against the AE groups. A second possibility may be that listeners shift their categories at the lower, more basic acoustic level rather than the suprasegmental level. If this is the case, then shifts should be consistent with their ambiguous phoneme exposure and show no difference between the native and the foreign accent conditions.

Method

Participants. A total of 190 participants (102 women, 88 men) were recruited from the University of Connecticut community and completed the study for course credit. All participants were 18 years of age or older, monolingual speakers of American English with no history of language, hearing or neurological disorders.

Stimulus selection. A total of forty critical words were selected, ranging from two to five syllables and containing one instance of either /s/ or /f/ at a syllable-initial, medial position within the word. We selected critical words with a word-medial phoneme in order to ensure that strong lexical information preceded the appearance of the target phoneme (e.g. Jesse & McQueen, 2011). Twenty critical words contained one instance of the phoneme /f/ (e.g. *amplifier*, *benefit*) and the remaining twenty critical words contained one instance of the phoneme /s/ (e.g. *assembly*, *capacity*). Appendix A lists all of the critical words, fillers and nonwords used in Experiment 1. None of the words contained the voiced counterparts of /s/ or /f/, respectively /z/ and /v/. All critical phonemes were preceded and followed by vowels so as to minimize coarticulation effects from consonants. Both the /s/ and /f/ sets were matched for written frequency (Francis & Kučera, 1982), as well as the number of letters, phonemes and syllables.

Neighborhood density (Washington University Speech and Hearing Lab Neighborhood Database; Sommers, 2002) and phonotactic probability (University of Kansas Phonotactic Probability Calculator; Vitevitch & Luce, 2004) were also matched across both lists using the calculated values provided by each online resource. Statistical differences between the lists on the aforementioned matched features were not significant as confirmed by two-sample t-tests which all resulted in p-values greater than 0.05.

Sixty filler words were chosen, each containing no instance of the phonemes /s/, /f/, /z/ or /v/ (e.g. *ability*, *canyon*). The set of filler words was matched to both sets of critical words on all levels previously mentioned: Kucera-Francis written frequency (mean = 20.38), number of letters (mean = 7.6), phonemes (mean = 6.64), syllables (mean = 2.88); neighborhood density and phonotactic probability. Again, no statistically significant difference was found between the critical words and fillers on these dimensions as confirmed by two-sample t-tests (all p-values greater than 0.05).

The final set consisted of 100 nonwords to equal the total number of critical and filler words. Nonwords were generated using the MRC Psycholinguistic Database (http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm) by taking real words and replacing one or two phonemes (e.g. *aviation* became '*aliation*' and *blackboard* became '*drackboard*'). Like the filler words, the nonwords did not include the phonemes /s/, /f/, /z/ or /v/. This set of words was matched to the combined set of real lexical items (i.e. critical plus filler words) on number of letters (mean = 7.91), phonemes (mean = 7.03), and syllables (mean = 2.73). Finally, phonotactic probability and neighborhood density were matched for the nonword list and the real word list. No significant differences were found for any of these dimensions, as confirmed by independent sample t-tests (all p-values greater than 0.05).

Stimulus construction. Each of the 200 stimulus items (40 critical words, 60 filler words and 100 nonwords) was recorded by an American male (mean f_0 = 130Hz) speaking a standard

American accent, and a male native speaker of Brazilian Portuguese (mean $f_0 = 147$ Hz). Our American speaker originated from the northeastern area of the United States (specifically, Pennsylvania) while our Portuguese speaker originated from São Paulo, Brazil. Measurements of centroid frequency for both speakers' natural and modified /s/ and /f/ phonemes were taken from the middle half of each individual phoneme (i.e. the first quarter and last quarter of each phoneme were excluded) using Praat (Boersma & Weenink, 2009).

Acoustic Properties of Brazilian Portuguese male fricatives – mean (SD)				
	/s/ (centre)	/s/ (dispersion)	/f/ (centre)	/f/ (dispersion)
Natural	6203 (1390) Hz	2860 (642) Hz	7744 (2457) Hz	3812 (990) Hz
Modified	6835 (1829) Hz	3745 (621) Hz	6637 (1339) Hz	3458 (709) Hz

Table 1. Spectral mean and SD (dispersion) of BP male

Acoustic Properties of American English male fricatives – mean (SD)				
	/s/ (centre)	/s/ (dispersion)	/f/ (centre)	/f/ (dispersion)
Natural	7234 (536) Hz	2080 (250) Hz	3907 (2306) Hz	3520 (801) Hz
Modified	6137 (848) Hz	3177 (470) Hz	6310 (874) Hz	3292 (359) Hz

Table 2. Spectral mean and SD (dispersion) of AE male

Recordings were saved onto a PC and trimmed to make individual files for each stimulus item using Praat sound editing software. For each critical word, both speakers also recorded an alternate version which replaced the original critical phoneme (/s/ or /f/) with the other target phoneme. For example, both speakers recorded 'amplifier', followed by 'amplisier' and 'capacity' followed by 'capafity'. This was done in order to create ambiguous phoneme blends /?sf/ which would be unique to each critical word rather than using one standard ambiguous phoneme across the critical word set. Previous studies such as Kraljic & Samuel (2005) and Norris, McQueen & Cutler (2003) have used this same method of recording critical words in pairs to preserve the coarticulation effects specific to each word. These studies also established the

norm of using one consistent version of the items as the “frame” to use when inserting the ambiguous phonemes. Kraljic and Samuel (2005) for example, used the s-versions of all their items as their frame while Norris et al. (2003) used the f-versions. In Experiment 1, we used the /f/ frame.

The ambiguous /?sf/ phoneme was constructed by first cutting the /s/ and /f/ phoneme from each pair of critical words, and trimming the longer phoneme down to the length of its shorter counterpart to ensure both phonemes were of equal length. Mean intensity for each phoneme segment was measured, and the average of the intensity values for /s/ and /f/ segments was calculated for each pair. The phoneme pair and its corresponding /f/ frame word (e.g. *ampli/f/ier*, *capa/f/ity*, *democra/f/y*) were then scaled to match this mean intensity. The phoneme pairs were mixed with a weighting of 50% /s/ and 50% /f/. Once an ambiguous /?sf/ phoneme was created for each word, the /f/ phoneme in the original frame word was measured for mean amplitude before being cut. The blended phoneme was then scaled to match this intensity before being inserted in place of the original. After editing was completed, all words were scaled one last time to an average intensity of 70dB, for consistency.

For our test items, the nonwords “*asi*” and “*afi*” were recorded by each speaker. The same cutting and blending procedures described above were used for the test items but instead of creating a single 50/50 blend, blends of ambiguous phonemes were created at seven different levels and spliced into the “*afi*” frame. The seven phoneme blends began at 20% /f/ + 80% /s/ and continued at ten-point intervals through 80% /f/ + 20% /s/.

Procedure. Two stimulus lists were created for each of the two speakers so participants would be assigned to the /s/ bias condition or /f/ bias condition for either the American male or Brazilian male talker. Each list consisted of 20 /s/ words, 20 /f/ words, 60 filler words and 100 nonwords presented in random order. Those in the /s/ condition heard ambiguous phonemes inserted into the /s/ critical words while the /f/ words remained unaltered and those in the /f/

condition heard ambiguous phonemes inserted into /f/ critical words while the /s/ words remained unaltered. Participants were randomly assigned to one of four possible conditions: Brazilian Portuguese male /s/ ($n=30$); Brazilian Portuguese male /f/ ($n=38$); American English male /s/ ($n=37$); American English male /f/ ($n=36$).

Participants were given two tasks, a lexical decision task (LD) for exposure and a phonetic categorization task (PC) for testing. All instructions were presented in written form on the computer screen, stressing the importance of speed and accuracy. Participants were tested before a computer in a quiet room and heard the words over headphones at 70 dB. The LD task required participants to press one button when they heard a word and a different button when they heard a nonword. Each participant was given a different randomized order of word presentation. They were given three seconds post word onset to respond, after which the next word would automatically play if no response was given. When response was given, the next word would play following 1s of inter-stimulus interval. Reaction times were recorded along with number of correct and incorrect responses and saved on an individual file for each participant. Participants were not made aware that ambiguous sounds would be present in some of the words, neither were they given feedback on their responses.

Following the LD exposure task, the PC test portion again presented participants with written instructions on the computer screen. This time, the instructions prompted them to listen to each test token and decide whether each sounded more like 'asi' or 'afi' by pressing corresponding keys as quickly and accurately as possible. In total, seventy test tokens were presented (ten of each blend) in a unique randomized order. Responses and reaction times were recorded for each participant. Again, no response feedback was provided to the participants.

Results

From the original 190 participants, a total of seven listeners were excluded from the

analysis if their lexical decision accuracy scores fell below seventy-five percent on total lexical decision accuracy or below seventy-percent on critical word accuracy. Average percentages for lexical decision accuracy in each of the four exposure groups appear in Table 3 below. An additional 42 participants who categorized half or more endpoints on the test continuum incorrectly were also excluded from the final analysis. The final analysis included 141 total participants with n's as follows: BP /s/ = 30, BP /f/ = 38, AE /s/ = 37, and AE /f/ = 36.

Exp 1 - Average lexical decision accuracy – mean (SD)					
	critical	unaltered	filler	nonword	Total
BP /s/	85(7)%	98(9)%	94(4)%	83(9)%	88(5)%
BP /f/	88(7)%	92(6)%	94(3)%	82(9)%	87(5)%
AE /s/	95(6)%	99(2)%	96(3)%	90(8)%	93(4)%
AE /f/	96(5)%	97(5)%	97(3)%	83(11)%	90(6)%

Table 3. Mean lexical decision accuracy by condition and word type.

Lexical decision. Listeners' accuracy on lexical decision was assessed in order to ensure that listeners did, in fact, perceive the altered critical tokens as real words. Lexical decision accuracy data were submitted to a 2x4 mixed ANOVA with word type as a within-subjects factor (i.e. critical, unaltered, filler, nonword) and speaker as a between-subjects factors. Overall, accuracy scores differed significantly between speakers ($F[1,137]=69.39$, $p<0.001$) with the BP groups having a lower overall accuracy score (87%) than the AE groups (92%). Across word types (i.e. critical, filler, nonword and unaltered), accuracy was also significantly different ($F[3, 411]=123.42$, $p<0.001$) with nonwords having the lowest mean accuracy scores and fillers or unaltered words having the highest mean accuracy scores. A post-hoc Tukey test showed significant differences between the critical (altered) and unaltered word types, $p<0.001$. A significant word type by speaker interaction was also shown, indicating that listeners accuracy differed across word types depending on which speaker they heard (with

BP groups having lower mean accuracy as stated above), $F(3,411)=7.47$, $p<0.001$.

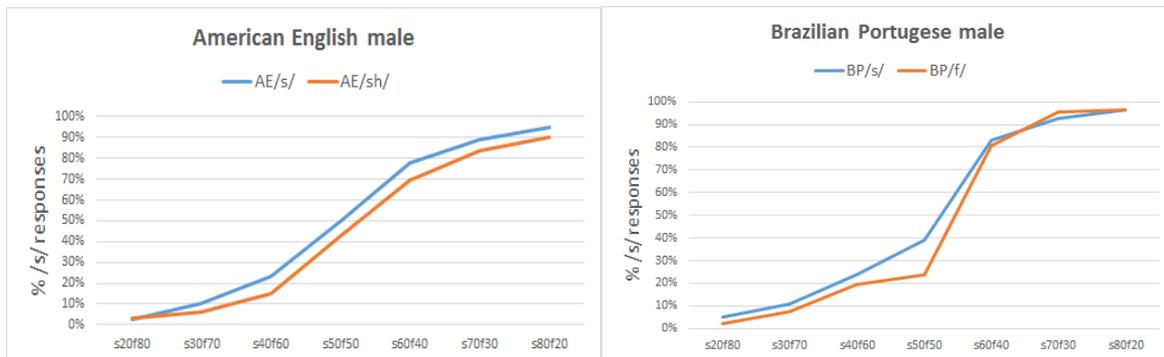
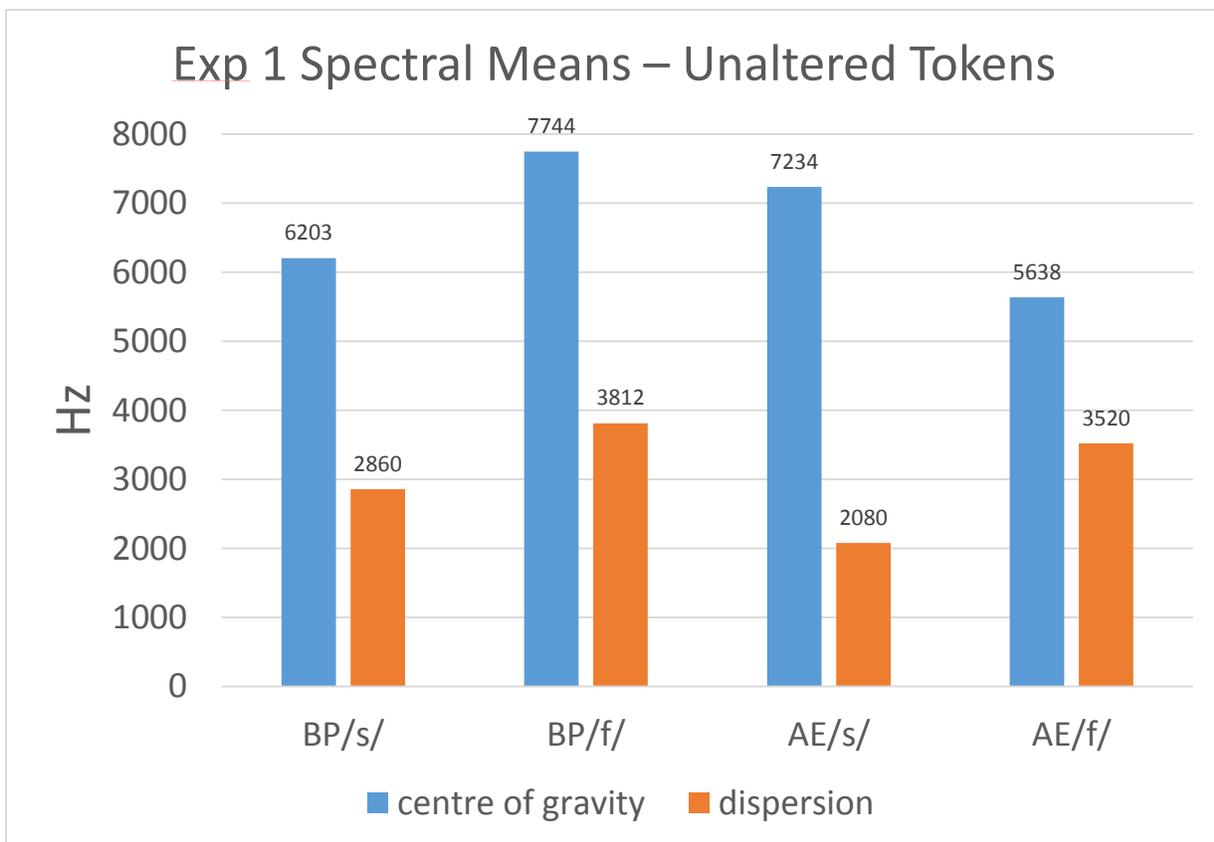


Figure 1. These graphs show the average percent /s/ responses (y-axis) per continuum token. The x-axis shows the continuum points from mostly /f/-sounding (left) to mostly /s/-sounding (right).

Categorization. Phonetic categorization data were submitted to a $2 \times 2 \times 7$ repeated measures ANOVA with all seven continuum test items as within-subjects factors and with *speaker* and *exposure phoneme* as between-subject factors. Phoneme bias condition had a significant effect on listeners' categorization, as shown by a main effect of exposure phoneme, $F(1,137)=4.30$, $p<0.05$, meaning more /s/ responses were given by those in the /s/-bias groups and more /f/ responses were given by those in the /f/-bias groups. Results also showed a main effect of continuum ($F[6,822]=742.08$, $p<0.001$), confirming the most basic assumption -- that listeners are categorizing tokens from one end of the continuum to the other as different (/s/ or /f/) speech sounds. Listeners also showed a difference in categorization depending on which speaker they were exposed to, irrespective of phoneme bias condition (significant continuum \times speaker interaction, $F[6,822]=7.51$, $p<0.001$). No other significant interactions or main effects were found. Notably, no significant speaker by phoneme interaction emerged, $F(1,137)=0.14$, $p=0.70$, nor any other interaction in which phoneme and speaker participated. This suggests that listeners adjust their categorization curves similarly, whether they are listening to a foreign- or native-accented talker.

Discussion

The present experiment sought to determine the role of accent expectation and the speaker's native phonology on the phonetic retuning effect. Building off the work of Reinisch and Holt (2014) (who used Dutch-accented English), Experiment 1 expanded the use of embedded ambiguous phonemes to a previously unresearched accent with greater phonological disparities to English -- namely, Brazilian Portuguese (BP).



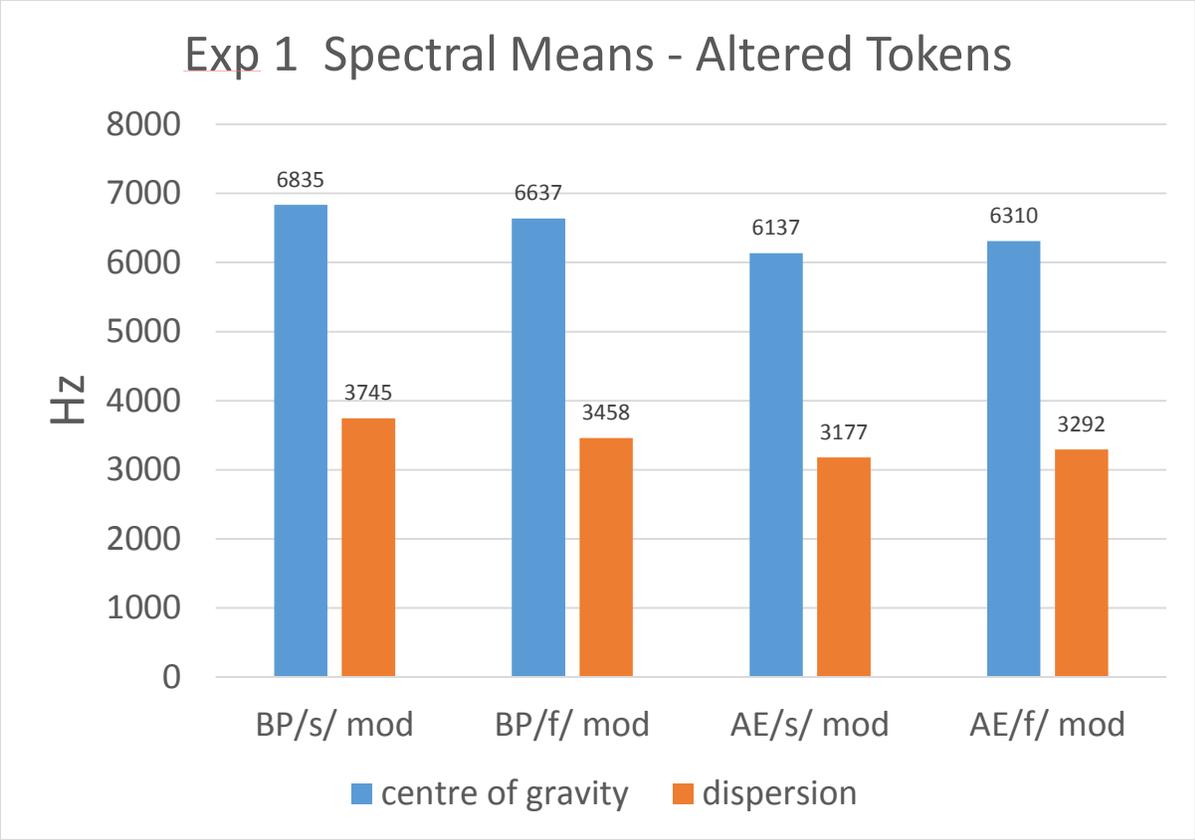


Figure 2. Average spectral means and standard deviations (dispersion) for Brazilian Portuguese and American English male speakers. Top graph shows acoustic measurements for the original fricatives while the bottom graph shows measurements for modified (ambiguous) phoneme blends.

We hypothesized that the widely different phonology of the BP accent would show a larger phonetic retuning effect than had been previously shown with either native-accented speakers (in this case, American English) or Dutch-accented English speakers (since English and Dutch share similar phonological inventories and stress patterns). The larger phonetic category shift, we reasoned, would appear as the result of the listener’s expectation that the accented speaker’s productions would be much more extreme and variable from their own native productions.

The results of Experiment 1 replicated previous findings, and ultimately did not support our initial ‘accent expectation’ hypothesis. While listeners in the BP conditions did show a robust category shift, it was not significantly different from the shift shown by the listeners in the AE

conditions in either magnitude or direction. These results suggest that the expectation created by a foreign accent may not play a role after all when it comes to phonetic retuning, but rather, it may be the acoustic information itself that drives the learning. Perhaps not surprisingly, this explanation is supported by previous work from Kraljic and Samuel (2005; 2006) which suggested that phonetic information was a crucial factor in the generalization of phonetic retuning to new speakers and phonemes.

In contrast to the Kraljic and Samuel (2011) study, in which deviant productions were caused by a pen in the speaker's mouth (i.e. the source of deviation was attributable to an external source not inherent to the speaker) and perceptual learning was effectively blocked, we might conclude based on the present results that a foreign accent is likely *not* treated as an "external" source of variation. And while it may be an integrated source of variation inherent to the speaker which will not be a roadblock to perceptual learning, it may also not help extrapolate it. One possibility is that the size of the shifts shown here reflect something about the acoustic characteristics of these voices in particular. As seen in Figure 2, acoustic properties for the modified (ambiguous) /s/ and /f/ phonemes were relatively similar across speakers (all within a range of about 700 Hz). The extent to which these particular acoustic properties influenced phonetic retuning cannot be determined by the current experiment, but it may be possible that the similarity of the modified phonemes might be one reason behind the lack of statistically significant difference in continuum categorization between speakers. Experiment 2 sought to replicate the findings from Experiment 1 with a new stimulus set, this time using female talkers and a new non-native accent (Greek-accented English).

Experiment 2

Experiment 2 followed the same design and procedure as Experiment 1, but this time featured female voices instead of male voices. We also used a new accent; the control voice was again, a native speaker of American English (AE) while the second, female voice was a

native female speaker of Greek (GR).

The objective of Experiment 2 was to replicate the results of Experiment 1. We sought to explore deeper into our own explanation of those results, which supported the hypothesis that low-level acoustic information is one of the primary cues on which phonetic retuning occurs, since results did not support a role of accent expectation (e.g., characteristics of a speaker's native phonology). If making changes to the speakers and the accent nonetheless replicated the results seen in the previous experiment, then we could add more supporting evidence to our previously stated explanation.

Method

Participants. A total of 157 participants (86 women, 71 men) were recruited from the University of Connecticut community and completed the study for course credit. All participants were 18 years of age or older, monolingual speakers of American English with no history of language, hearing or neurological disorders. None of the participants had previously taken part in Experiment 1.

Stimulus selection and construction. This experiment used a different set of lexical items, taken from Appendix A of Kraljic and Samuel (2005) and focused on a different pair of fricative phonemes -- /s/ and /ʃ/. Like the previous list, this set contained 40 critical words, 60 fillers and 100 nonwords. The recording and construction of Experiment 2 items followed the same procedure as described above for Experiment 1.

Acoustic properties of Greek female fricatives – mean(<i>SD</i>)				
	/s/ (centre)	/s/ (dispersion)	/ʃ/ (centre)	/ʃ/ (dispersion)
Natural	6770(1079) Hz	1479(425) Hz	3231(462) Hz	1310(224) Hz
Modified	5259(797) Hz	2301(389) Hz	4840(699) Hz	2168(434) Hz

Table 4. Spectral means and dispersion measures (in Hz) for altered and unaltered Greek female fricatives.

Acoustic properties of American English female fricatives – mean(SD)				
	/s/ (mean)	/s/ (dispersion)	/ʃ/ (mean)	/ʃ/ (dispersion)
Natural	6869(759) Hz	1886(623) Hz	3644(241) Hz	1367(303) Hz
Modified	4539(514) Hz	2062(392) Hz	4726(549) Hz	2209(229) Hz

Table 5. Spectral means and dispersion measures (in Hz) for altered and unaltered American English female fricatives.

Procedure. Experimental procedure for Experiment 2 followed that of Experiment 1 as described above.

Results

Out of the original 157 participants, 27 were excluded after failing to meet LD accuracy criteria (>75% total accuracy and >70% critical word accuracy) and/or categorization accuracy criteria (half or more correct responses on the /s/ and /ʃ/ ends). The final analysis included 130 participants, with n's as follows: Greek /s/ = 31, Greek /ʃ/ = 31, AE /s/ = 36, and AE /ʃ/ = 32.

Exp 2 - Average lexical decision accuracy – mean(SD)					
	critical	unaltered	filler	nonword	Total
GR/s/	96(5)%	86(8)%	86(6)%	87(8)%	87(4)%
GR /ʃ/	84(8)%	97(5)%	87(6)%	86(6)%	87(3)%
AE /s/	97(7)%	99(2)%	95(4)%	92(7)%	94(4)%
AE /ʃ/	96(5)%	95(2)%	99(2)%	89(12)%	93(6)%

Table 6. Mean lexical decision accuracy by condition and word type.

Lexical decision. Similar to Experiment 1, lexical decision accuracy was submitted to a 2x4 repeated measures ANOVA with word type as a within-subjects factor (i.e. critical, unaltered, filler, nonword) and speaker as the between-subjects factor. Results showed listeners differed significantly in their accuracy depending on the type of word (main effect of word type,

F[3,384]=22.6, $p<0.001$]). While speaker condition did show a significant difference in accuracy overall, $F(1,128)=158.84$, $p<0.001$), there was no statistically significant difference between speaker conditions across individual word types, (speaker x word type interaction; $F[3,384]=2.39$, $p=0.07$). In other words, average total accuracy for both GR groups (87%) was overall significantly lower compared to the AE groups (94%), but no significant difference was shown between speakers for each of the four word types.

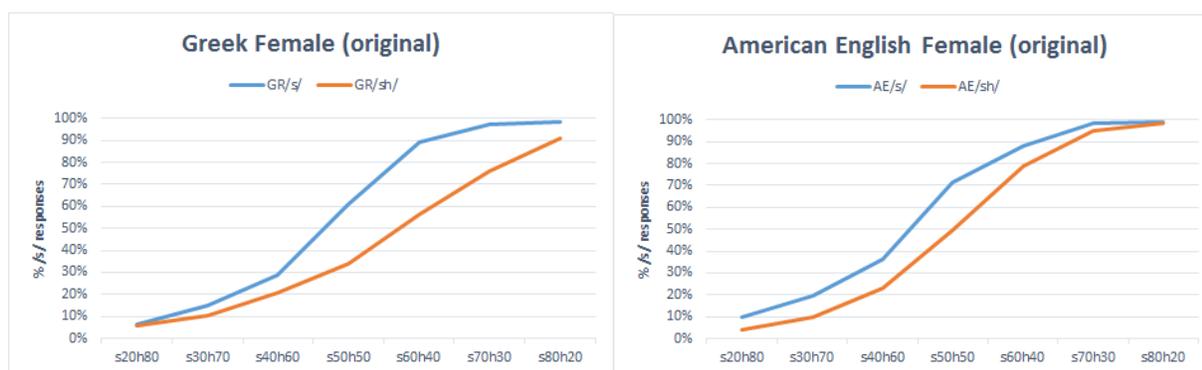


Figure 3. These graphs show the average percent /s/ responses (y-axis) per continuum token. The x-axis shows the continuum points from mostly /ʃ/-sounding (left) to mostly /s/-sounding (right).

Categorization. As with Experiment 1, phonetic categorization data were submitted to a 2x2x7 repeated measures ANOVA with all seven continuum test items as within-subjects factors and with speaker and exposure phoneme as between-subject factors. Again, the seven points along the continuum were heard as distinct by the listeners regardless of phoneme exposure conditions, as shown by a main effect of continuum, $F(6,756)=653.92$, $p<0.001$. As seen in Experiment 1, continuum tokens were also categorized differently by the listeners depending on the speaker they heard (continuum x speaker interaction, $F[6,756]=2.62$, $p<0.05$) and depending on their phoneme bias condition (continuum x phoneme interaction, $F[6,756]=7.87$, $p<0.001$), which indicates an effect of perceptual learning resulting from the ambiguous phoneme embedded in a lexically-guided context. Finally, the speaker and phoneme bias had

their own individual effects on listeners' categorization along the continuum overall (main effect of speaker, $F[1,126]=7.52$, $p<0.01$; main effect of phoneme bias, $F[1,126]=24.11$, $p<0.01$).

Results also showed that when taking both speaker and phoneme bias into account at once, listeners categorized the continuum tokens differently depending on which of the four condition they heard (continuum x speaker x phoneme interaction, $F[6,756]=4.08$, $p<0.001$).

Subsequent, individual 2x7 ANOVA's (one for each speaker) showed a significant phonetic retuning for the phoneme bias groups in both the Greek speaker (continuum x phoneme bias interaction, $F[6,360]=8.78$, $p<0.001$; main effect of phoneme bias, $F[1,60]=14.71$, $p<0.001$) and the American English speaker (continuum x phoneme bias interaction, $F[6,396]=3.02$, $p<0.01$; main effect of phoneme bias, $F[1,66]=9.00$, $p<0.01$).

Discussion

The results of Experiment 2 mostly followed those of Experiment 1. Both the speaker and phoneme conditions had individual influences over the listeners' categorization along the *asi-ashi* test continuum. This is perhaps unsurprising given that we expect listeners to show a shift in the phonetic categorization of their target phoneme after exposure to phoneme bias, regardless of the speaker. Our question here was how the shift might look *different* between the speakers. The fact that speaker condition had an individual effect on listeners, may be a reflection of the natural acoustic difference in the speech signal of any two individual talkers.

The significant finding in Experiment 2, which stands in contrast to Experiment 1, is the effect that speaker and phoneme bias condition *together* had on listeners' categorization. That is to say, listeners were affected not just individually by the speaker or ambiguous phoneme bias they were exposed to, but also by the influence of those combined. As seen in Figure 3, listeners in both Greek speaker conditions show a larger shift on average than those in the AE speaker conditions. Subsequent, individual ANOVA's for each speaker confirmed the phoneme

bias had a significant effect of learning (i.e. significantly different shifts for the /s/ and /ʃ/ groups) for both speakers, not just for one speaker.

While this result may seem, at first, to lend support to the hypothesis that some foreign-accented speech may induce larger phonetic retuning effects as a result of increased and unreliable variability in the non-native accent as a whole, it also contradicts the results from Experiment 1. However, if we delve deeper and look into the acoustics of the continuum tokens in each experiment, we see that the phonetics may be the underlying reason behind these seemingly contradictory results.

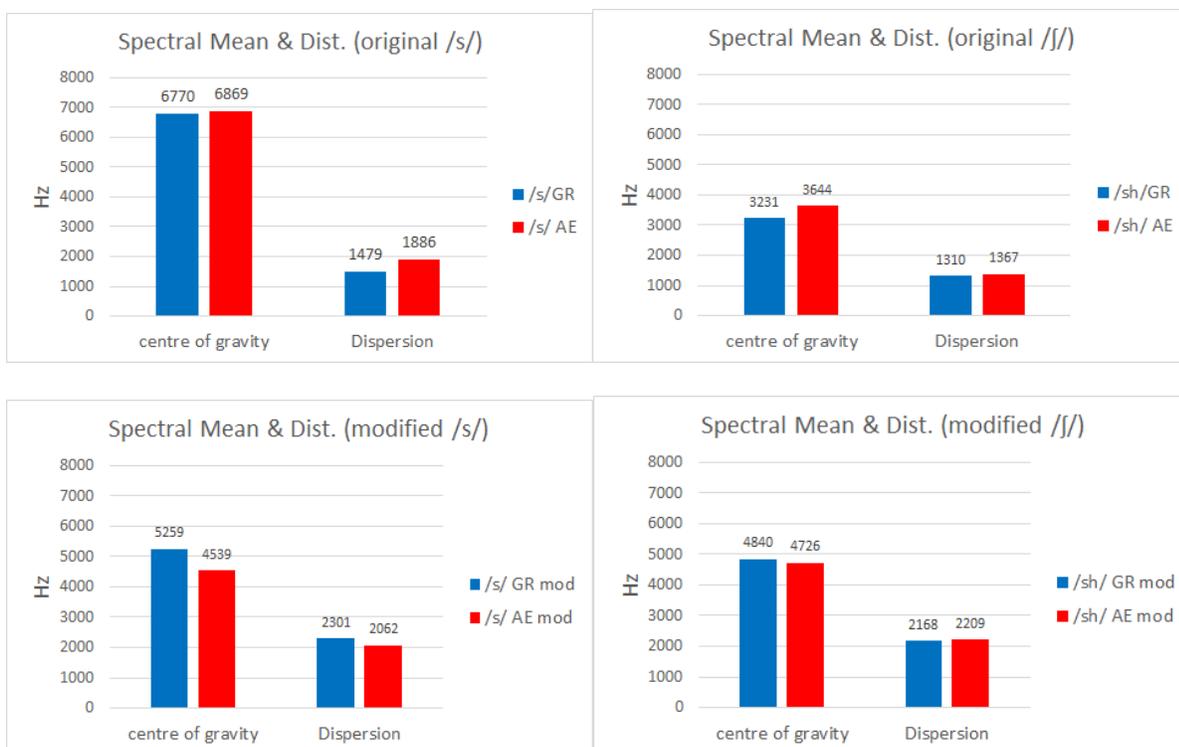


Figure 4. Average spectral means and standard deviations (dispersion) for Greek and American English female speakers. Top two graphs show acoustic measurements for the original fricatives while the bottom two graphs show measurements for modified (ambiguous) phoneme blends.

In Figure 4, we see the average spectral means and standard deviation measures for each speaker's original and modified fricative tokens. Notably, average spectral means for modified (ambiguous) Greek and American English /ʃ/ tokens fall relatively close (4840 Hz for

Greek /f/ vs. 4726 Hz for AE /f/), in contrast to modified /s/ tokens, which show a greater difference between speakers (5259 Hz for Greek /s/ vs. 4539 Hz for AE /s/). While these acoustic measurements may be insightful, we must take caution in interpreting the potential impact of these *particular* acoustic features on the magnitude of phonetic retuning effects. Although past research has shown that spectral similarity between tokens can correlate with transfer of learning to new speakers (Kraljic & Samuel, 2005), it is unclear just how close two fricative tokens must be to ensure this transfer. Again, however, we cannot definitively conclude whether this particular acoustic information is directly correlated with the size or magnitude of phonetic retuning, or whether it can be a proxy measure. Given the results of Experiment 2, it may be possible that the larger retuning effect seen in the Greek speaker conditions is affected by this difference in spectral means and dispersion, but this cannot be definitively determined in the current study.

Experiment 3

Experiment 3 used the exact same stimuli as Experiment 2, with one modification: the Greek speaker's fricatives were spliced into the American English speaker's words, and the American English speaker's fricatives were spliced into the Greek speaker's words. Fricatives were swapped in both the exposure and test phases. The objective of this experiment was to determine more precisely whether phonetic information or accent expectation would reproduce the phonetic shifts seen in Experiment 2. In other words, using the results from Experiment 2 as a baseline, we wanted to determine whether the listeners in this experiment would show phonetic retuning that would match the speaker they heard or the fricatives they heard (now that that information was switched). Results from this experiment would give us a better idea of which kind of information listeners mostly rely on when making perceptual changes.

Method

Participants. A total of 142 participants (74 women, 68 men) were recruited from the University of Connecticut community and completed the study for course credit. All participants were 18 years of age or older, monolingual speakers of American English with no history of language, hearing or neurological disorders. None of the participants had previously participated in Experiments 1 or 2.

Stimulus selection and construction. Stimuli for Experiment 3 were identical to those described for Experiment 2 above, with the difference of swapping each speaker’s fricatives into the other’s voice. Recording and construction of the stimuli followed the process described in Experiment 1.

Procedure. Experimental procedure for Experiment 3 followed that of Experiment 1 and 2 as described above.

Results

Of the original 161 participants, 51 were excluded after failing to meet LD accuracy criteria (>75% total accuracy and >70% critical word accuracy) and/or categorization accuracy criteria (half or more correct responses on the /s/ and /ʃ/ ends). The final analysis included 110 participants, with n’s as follows: Greek /s/ = 27, Greek /ʃ/ = 28, AE /s/ = 28, and AE /ʃ/ = 27.

Exp 3 - Average lexical decision accuracy					
	critical	unaltered	filler	nonword	Total
GR /s/	92(7)%	86(6)%	86(4)%	88(8)%	87(4)%
GR /ʃ/	86(8)%	97(3)%	86(6)%	88(6)%	88(4)%
AE /s/	99(3)%	99(2)%	95(4)%	93(6)%	95(4)%
AE /ʃ/	91(8)%	99(3)%	95(4)%	91(7)%	93(4)%

Table 7. Mean lexical decision accuracy by condition and word type.

Lexical decision - Accuracy for lexical decision was submitted to a 2x2x4 repeated measures ANOVA with word type as a within-subjects factor (i.e. critical, unaltered, filler, nonword) and speaker as the between-subjects factor. Similar to Experiments 1 and 2, listeners showed significantly different rates of lexical decision accuracy across the four word types, $F(3,324)=19.34$, $p<0.001$. Accuracy across word types also differed depending on the speaker, $F(3,324)=3.13$, $p<0.05$, evidenced by listeners' tendency to show lower scores across some word types in the GR conditions than in the AE conditions (see Table 7). Finally, overall total accuracy was also significantly different by speaker, $F(1,108)=112.324$, $p<0.001$, with the GR conditions having lower overall average accuracy rates than the AE conditions.



Figure 5. These graphs show the average percent /s/ responses (y-axis) per continuum token. The x-axis shows the continuum points from mostly /j/-sounding (left) to mostly /s/-sounding (right).

Categorization - Listeners' phonetic categorization data were submitted to a 2x2x7 repeated measures ANOVA with all seven continuum test items as within-subjects factors and with speaker and exposure phoneme as between-subject factors. As seen with both Experiments 1 and 2, all seven tokens along the continuum were heard and categorized distinctly by the listeners regardless of training, $F(6,636)=594.37$, $p<0.001$. Speaker condition was individually influential on token categorization (continuum x speaker interaction,

$F[6,636]=8.96, p<0.001$) as well as the phoneme bias condition, which points to an effect of retuning as a result of exposure to the ambiguous phoneme in the lexical-guided context (continuum x phoneme interaction, $F[6,636]=5.93, p<0.001$). Phoneme bias condition had an overall effect on phonetic learning, regardless of speaker condition, $F(1,106)=15.76, p<0.001$. Finally, a significant interaction between phoneme exposure and speaker revealed a significant impact of speaker *and* phoneme bias condition together on overall categorization, (speaker x phoneme bias interaction, $F[1,106]=5.36, p<0.05$). Subsequent ANOVA's for each individual speaker showed a significant effect of phoneme bias for the American English speaker, $F(1,53)=14.84, p<0.001$, but not for the Greek speaker, $F(1,53)=2.09, p>0.05$.

Discussion

Experiment 3 results mostly replicated the results seen from Experiments 1 and 2. In contrast to the previous experiments, however, no main effect of speaker was found. This may seem unexpected, given that the previous experiment (2) using the exact same female speakers did show a main effect of speaker. However, speaker condition had a significant individual interactions with both continuum categorization and phoneme bias condition. Individually, these interactions indicate that each speaker strongly influenced listeners' overall token categorization in conjunction with the phoneme bias condition. These results directly reflect the results of Experiment 2. A speaker by phoneme interaction suggested that overall, listener's categorization shifts were directly influenced by the combination of their speaker and phoneme bias condition.

Once broken down, however, results of two individual ANOVA's (by speaker) showed a significant effect of phoneme bias condition for the American English speaker, but not the Greek speaker. While Figure 3 shows a noticeable difference in categorization curves for each phoneme bias group under the Greek speaker, this was not statistically significant. The reason for this is not clear from the current results.

Overall, the results from Experiment 3 strongly pointed to a primary role of phonetic information in the process of phonetic retuning. Had the listeners in Experiment 3 directly replicated the retuning effects seen in Experiment 2 for each speaker, we might conclude that the collective results could support the accent expectation hypothesis. The accent expectation hypothesis predicted a stronger effect of the speaker's accent on the overall retuning effect, meaning that regardless of the specific phonetic information in the target phoneme, listeners should show larger shifts for fricatives in the foreign-accented voice. What we saw instead was the opposite -- a replication of the effect which followed the *phonetic information* in the target fricative, not the accent of the speaker. Comparing the categorization curves in Figure 3 (Experiment 2) to the curves in Figure 5 (Experiment 3), it is apparent that larger shifts occurred for the Greek speaker's fricatives, regardless of whether they were embedded in her original voice, or the American English speaker's voice. A similar effect is seen for the American English speaker's fricatives in which a smaller, but still robust shift emerged for both the original and swapped conditions. These results support the no-accent expectation hypothesis in which we expect phonetic retuning to occur primarily on the basis of the acoustic-phonetic information present in the specific target phoneme, rather than an effect of overall phonological expectation.

Experiments 2 and 3 - Omnibus ANOVA's

Motivated by the individual results from Experiments 2 and 3 in which the acoustic-phonetic information in the target fricatives seemed to drive the retuning effects regardless of the speaker's voice, we conducted two additional ANOVA's to determine which effects were significant as a result of the fricative swaps. First, a 2x2x2x7 ANOVA was conducted using the same between- and within-subjects factors as the previous three experiments (i.e. speaker, phoneme bias, and test continuum) with the addition of 'experiment condition' as a between-subjects factor (i.e. whether the listeners were in the original fricative experiment [2] or the swapped fricative experiment [3]). The second 2x2x2x7 ANOVA was identical, except the

'experiment' factor was replaced by a 'fricative source' between-subjects factor which indicated the talker from which the fricatives came from in both the original and swapped experiments.

Results

The first ANOVA (which added 'experimental condition' as an added between-subjects factor) confirmed many of the findings seen for Experiments 1 - 3 above. Tokens across the continuum were heard as distinct phonemes (/s/ to /ʃ/), $F(6,1392)=1236.25$, $p<0.001$, as they were previously. Phoneme bias condition showed a main effect on overall continuum responses, $F[1,232]=38.83$, $p>0.001$, as did speaker condition, $F[1,232]=6.60$, $p<0.05$. Continuum tokens were also categorized differently depending on the speaker (continuum x speaker interaction, $F[6,1392]=2.78$, $p<0.05$) or phoneme bias condition (continuum x phoneme bias interaction, $F[6,1392]=12.28$, $p<0.001$) a listener heard. Experiment condition (original fricatives vs. swapped fricatives) also had a significant effect on how the speaker affected categorization along the continuum (experiment x speaker x continuum interaction, $F[6,1392]=8.93$, $p<0.001$). Continuum categorization was also affected by the speaker and phoneme conditions together (speaker x phoneme x continuum, $F[6,1392]=2.39$, $p<0.05$). Finally, experiment condition had a significant effect in conjunction with speaker, phoneme bias and continuum tokens (experiment x speaker x phoneme bias x continuum interaction, $F[6,1392]=2.79$, $p<0.05$), as did the three between-subjects factors on overall categorization, (experiment x speaker x phoneme bias, $F[1,232]=5.80$, $p<0.05$).

The purpose of the second ANOVA (which replaced 'experimental condition' with 'fricative source' as a between-subjects factor) was to determine the source of the significant effects between Experiments 2 and 3. Since the same information was used in both analyses, all significant effects described above in the previous ANOVA results were identical in the present ANOVA (except those involving 'experiment' condition, since that factor was not present here). Fricative source (i.e. Greek fricatives in AE speaker vs. AE fricatives in Greek speaker)

had a significant effect of categorization along the test continuum, $F(6,1392)=8.94$, $p<0.001$. Continuum categorization was interactively affected by phoneme bias condition and the source of the fricatives (continuum x phoneme bias x fricative source interaction, $F[6,1392]=0.19$, $p<0.05$). Importantly, an interaction for all four factors was *not* observed in this analysis, which will be discussed shortly, $F[6,1392]=1.31$, $p>0.05$, nor was an interaction observed for the three between-subjects factors, $F[1,232]=0.52$, $p>0.05$.

Discussion

Most of the effects seen in the omnibus ANOVA's replicated and further confirmed many of the effects seen for the individual experiments in the present study (e.g. main effects of continuum, speaker and phoneme; interactions of continuum by speaker, by phoneme and speaker *with* phoneme). Details about these effects can be found in the individual experiment discussions above.

The effects to focus on are the interactions involving the 'experimental condition' and 'fricative source' factors. Namely, significant three- and four-way interactions were observed in the first ANOVA (experiment x speaker x phoneme bias x continuum; experiment x speaker x phoneme bias) which indicate that the effect of speaker and phoneme bias conditions on continuum categorization is differentiated by the experimental condition (original vs. swapped fricatives) listeners heard. Specifically, as we see in comparing the retuning effects in Figures 3 and 4, categorization curves appeared to stay consistent with the phonetic information (i.e. fricative speaker source) and *not* with the speaker overall. Categorization curves clearly change within each speaker depending on whether they contained the speaker's original fricatives or swapped fricatives. The interactions seen here show the statistical significance of this effect.

Further confirmation comes from the lack of significant interactions in the second ANOVA – that is, no significant three- or four-way interactions involving fricative source were shown. This lack of interaction demonstrates that results of continuum categorization followed,

and were consistent with, the source voice of the *fricative*, not the voice of the speaker in which they were embedded. In other words, because no statistically significant change in categorization was observed as long as the fricative source remained constant, there is a strong indication that phonetic information is the primary driving factor in determining the size of the phonetic retuning shifts, not the speaker's phonology or accent expectation.

General Discussion

The mechanisms by which our perceptual system copes with the variability in the speech signal are not yet fully understood. A growing body of literature on the phonetic retuning effect in speech has helped advance our knowledge about how changes in the acoustic signal of speech are accommodated and learned by our perceptual system. Previous research has demonstrated the existence of a robust, lexically-guided phonetic retuning effect which allows listeners to adjust to some degree of acoustic deviation at the phoneme level (e.g. Norris et al., 2009; Kraljic & Samuel, 2009). However, more research is needed to expand our understanding of what types of real-word influences may play a role in this process.

To this end, the current set of studies aimed to explore the phonetic retuning effect and how expectations about the phonetic atypicality likely to be seen in accented speech might have a different effect on phonetic learning than what is typically seen in native accent contexts. We wanted to better understand the mechanisms that underlie phonetic retuning and whether the cues our perceptual system attends to during the retuning process related more strongly to the low-level acoustic information of the target phonemes, or rather a more global information related to the phonological characteristics of the speaker's accent.

Using monolingual American English listeners, a series of three experiments compared the phonetic category shifts induced by both native-accented (American English) speakers to those induced by foreign-accented speakers (Brazilian Portuguese and Greek). Our results largely support the idea that the process of lexically-guided phonetic retuning is heavily

influenced and limited by low-level acoustic (phonetic) information, rather than speaker's identity or the general phonological characteristics of their accent. As evidenced by our results from Experiments 1 and 2, this effect may be somewhat mediated by the natural variation in the acoustics of any two given speakers. That is, a speaker who produces a larger shift in listeners due to more natural variability in the acoustics of his/her tokens can also produce a much wider scale upon which to measure that shift than a speaker with less natural variability. In addition, Experiment 3 demonstrated that one speaker's tokens can produce similar effects in another speaker's voice when her critical tokens are spliced in (unbeknownst to the listener). Similar transfer of learning effects based on similarity of phonetic information between tokens have been observed in a number of previous studies (see Kraljic & Samuel, 2005; 2006, Reinisch & Holt, 2014).

We had expected to see some influence from the expectation created by the introduction of a foreign accent in speech. A speaker's native language may share many phonological and prosodic features with their second learned language, or it may share very few features. As we interact with an accented speaker, we as listeners can form expectations about that speaker's productions and the degree to which they may deviate in their productions of our native phonemes. This observation is not simply an anecdotal one, but has in fact been shown to have a demonstrable effect.

A number of studies have shown expectation to have an important and measurable influence on the way we understand the productions of accented speakers in other linguistic domains. Hanulikova et al., (2012), for instance, found that listeners do not process syntactic errors from L2 speakers in the same way as for L1 speakers. Using EEG to measure responses to grammatical errors in the sentences of native Turkish speakers of Dutch, the researchers found that native Dutch listeners did not exhibit the standard P600 effect for the accented L2 speakers that would normally be seen for syntactic errors in native speakers. Here, it seemed that expectation did play a role in the processing of foreign-accented speech, as listeners were

much more lax in their expectations of correct grammatical structure from the accented speakers. Similarly, Schmid & Yeni-Komshian (1999) found that the mispronunciations from L2 speakers were much more likely to be overlooked than mispronunciations from native talkers. Their results showed that native English listeners were more accurate and faster at detecting the mispronounced words of native speakers than those of the accented speakers. This effect was graded by the degree of accentedness -- with accuracy for detecting mispronunciations decreasing for heavier accents and increasing for milder accents. Studies like these confirm that to some degree, a listener's expectation can have an immediate, general impact on the perception of accented speech (at least at the syntactic and lexical levels). The results of the present studies, however, do not add support to the role of accent expectation in the particular domain of phonetic retuning for fricative phonemes.

Instead, our studies lend support to the ever-increasing evidence of phonetic, low-level acoustic information as a primary driver of phonetic retuning and its generalization. Very recently, more evidence has come to light about the importance of phonetic retuning and its connection (or lack thereof) to a given phonological representation. Recent studies in Korean and German have shown that phonetic retuning for an allophonic variant of a phoneme will not generalize to the original representation of that phoneme (Reinisch & Mitterer, 2015; Mitterer, Kim & Cho, 2015). For instance, category shifts for voiced stop-consonants that are realized as voiceless in word-final positions do not generalize back to the original voiced phoneme, but *will* generalize to originally voiceless stops (Reinisch & Mitterer, 2015). This led the authors to conclude that phonetic retuning is primarily based on acoustic information rather than overall phonological representations. In the context of our current study, we could similarly say that phonetic learning is not connected to the phonological features of the speaker's L1, but rather on the spectral and temporal features in the acoustics.

In an overarching sense, the current study can be contextualized as part of the greater conversation regarding the role of top-down information (including expectations) on the

phonological and phonetic processing of speech. The present study presents behavioral results which indicate that repeated presentations of deviant speech sounds do ultimately have an effect on the way those speech sounds are later perceived (whether that involves a change in the *mental representation* of a particular phoneme is an entirely separate conversation). Some might account for these findings by arguing that this kind of phonetic online processing is not the result of perceptual feedback and thus does not permeate down to the pre-lexical level (e.g. phonemes) (see Norris, McQueen & Cutler, 2000; 2015). Others would argue that feedback from the lexical level is *essential* for online perceptual changes to occur at the phonetic, pre-lexical level (see Elman & McClelland, 1986; 1988). Our current study cannot speak to these particular mechanisms (i.e. how exactly we arrive at these perceptual changes), but it can speak to the factors that do and do not affect changes in the perceptual *learning* of individual phonemes. At present, the evidence available does not suggest that expectation -- from a speaker's accent at least -- influences this particular process.

In order to more precisely determine the role of the acoustics, future research might look into using a closely controlled or manipulated speech signal. As Kraljic and Samuel (2005; 2006) as well as our present study demonstrated the generalizability of retuning as a result of similarity in acoustic information, future studies should tightly control for the acoustic properties of their speaker(s) in order to observe the direct effect on the retuning outcome. Studies should also provide an ample scale by which to measure categorical shift in order to capture the shift as accurately as possible. These kinds of control measures might help us better deduce what kind of information can permeate down to the phonetic processing system and which information is not particularly necessary to the retuning process.

In understanding how we as listeners adjust to variability at the phonemic level, an ever-growing body of evidence is emerging to support the crucial and perhaps primary role of low-level acoustic information in the phonetic retuning process. While other domains of language may be subject to the influence of our conscious expectations of a real-world variant (e.g. a

foreign accent), it is beginning to seem as though phoneme-level retuning may not be one of them. Uncovering the mechanisms that make this type of learning possible is a challenging endeavor given the natural variability found between any two given speakers. Nonetheless, knowing how we accommodate, learn and attribute perceptual adjustments to a particular speaker helps us better understand how the flexibility in our perception gives rise to the stability of categorical percepts and ultimately, meaning.

Appendix A – List of critical words used in Experiment 1

Critical /s/ words

acid
capacity
castle
classical
consider
democracy
dissent
episode
implicit
impossible
lesson
listen
medicine
municipal
parasite
participate
peninsula
reconcile
rehearsal
wholesome

Critical /f/ words

amplifier
awful
barefoot
benefit
biography
buffalo
cafeteria
coffee
definite
diaphragm
difficulty
elephant
identify
informal
lifetime
manufacture
nephew
perfect
peripheral
preferable

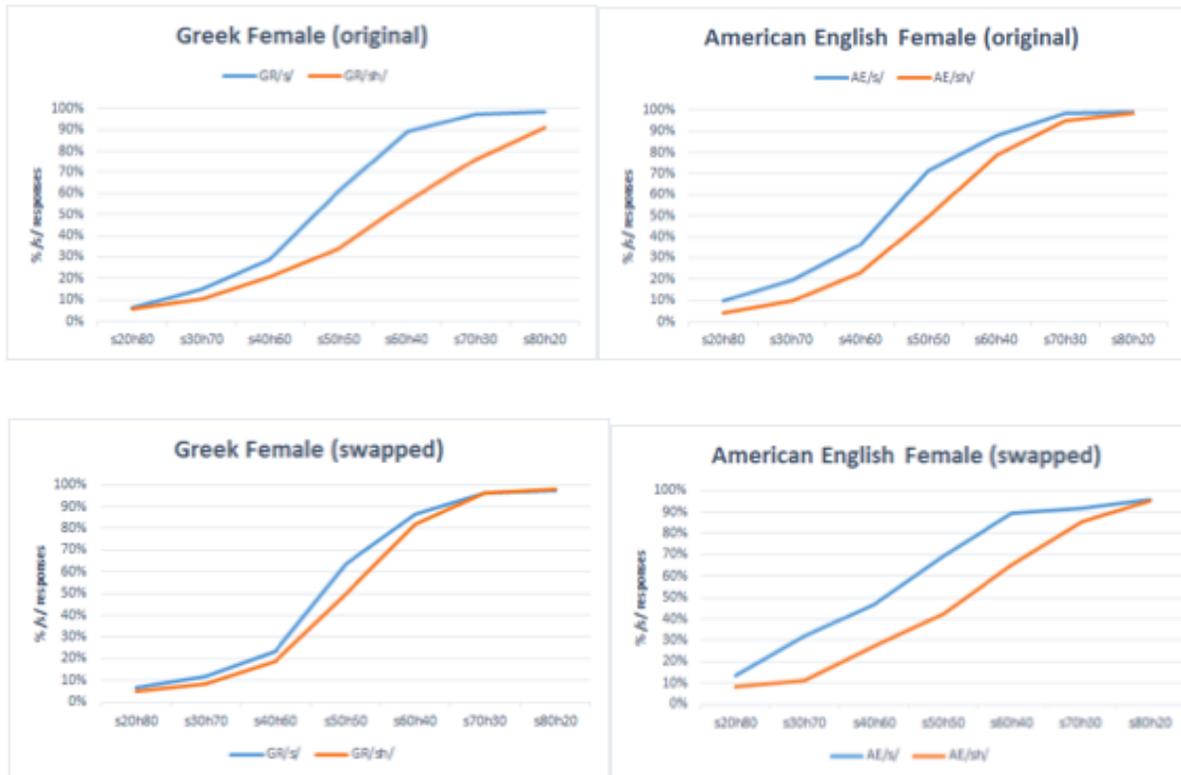
References

- Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America*. Retrieved from <http://psycnet.apa.org/psycinfo/2003-04364-014>
- 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*(3), EL174–EL180.
- Boersma, P., & Weenink, D. (2009). *Praat: doing phonetics by computer [Software package]*.
- Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729.
- Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America*, *106*(4), 2074–2085.
- Cambier-Langeveld, T., & Turk, A. E. (1999). A cross-linguistic study of accentual lengthening: Dutch vs. English. *Journal of Phonetics*, *27*(3), 255–280.
- Cohen, J., Kamm, T., & Andreou, A. G. (1995). Vocal tract normalization in speech recognition: Compensating for systematic speaker variability. *The Journal of the Acoustical Society of America*, *97*(5), 3246–3247.
- Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, *67*(2), 224–238.
- Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language*, *27*(2), 143–165.
- Fourakis, M., Botinis, A., & Katsaiti, M. (1999). Acoustic characteristics of Greek vowels. *Phonetica*, *56*(1-2), 28–43.

- Francis, W., & Kucera, H. (1982). Frequency analysis of English usage. Retrieved from <http://www.citeulike.org/group/1778/article/930447>
- Frota, S., & Vigário, M. (2001). On the correlates of rhythmic distinctions: The European/Brazilian Portuguese case. *Probus*, 13(2), 247–275.
- Frota, S., Vigário, M., & Martins, F. (2002). Language discrimination and rhythm classes: evidence from portuguese. In *Speech Prosody 2002, International Conference*. Retrieved from http://www.isca-speech.org/archive_open/sp2002/sp02_319.html
- Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1), 110.
- Hanulíková, A., Van Alphen, P. M., Van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience*, 24(4), 878–887.
- Jesse, A., & McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic Bulletin & Review*, 18(5), 943–950.
- Johnson, K., & Babel, M. (2010). On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. *Journal of Phonetics*, 38(1), 127–136.
- Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51(2), 141–178.
- Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review*, 13(2), 262–268.
- Kraljic, T., & Samuel, A. G. (2011). Perceptual learning evidence for contextually-specific representations. *Cognition*, 121(3), 459–465.
- Maddieson, I., & Ladefoged, P. (1996). *The sounds of the world's languages*. Oxford Publishers.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86.

- Mitterer, H., Kim, S., & Cho, T. (2015). Patterns of generalization of perceptual learning on phonetic representations. In The Scottish Consortium for ICPHS 2015, *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow.
- Munro, M. J. (2008). Foreign accent and speech intelligibility. *Phonology and Second Language Acquisition*, 193–218.
- Munson, B. (2011). The influence of actual and imputed talker gender on fricative perception, revisited (L). *The Journal of the Acoustical Society of America*, 130(5), 2631–2634.
- Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23(03), 299–325.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47(2), 204–238.
- Norris, D., McQueen, J. M., & Cutler, A. (2015). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience*, 1–15.
- 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.
- Reinisch, E., & Mitterer, H. (2015). Perceptual learning in speech is phonetic, not phonological: Evidence from final consonant devoicing. In The Scottish Consortium for ICPHS 2015, *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow.
- Roach, P. (1982). On the distinction between “stress-timed” and “syllable-timed” languages. *Linguistic Controversies*, 73–79.
- Samuel, A. G. (1981). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology: General*, 110(4), 474.

- Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, 71(6), 1207–1218.
- Schmid, P. M., & Yeni-Komshian, G. H. (1999). The effects of speaker accent and target predictability on perception of mispronunciations. *Journal of Speech, Language, and Hearing Research*, 42(1), 56–64.
- Sebastián-Gallés, N., Dupoux, E., Costa, A., & Mehler, J. (2000). Adaptation to time-compressed speech: Phonological determinants. *Perception & Psychophysics*, 62(4), 834–842.
- Sidas, S. K., Alexander, J. E., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, 125(5), 3306–3316.
- Sommers, M. (2002). Washington University Speech and Hearing Lab Neighborhood Database. Retrieved from *Neighborhoodsearch. Wustl. edu/Home. Asp.*
- Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers*, 36(3), 481–487.
- Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *The Journal of the Acoustical Society of America*, 52(4B), 1238–1250.
- Xue, A. (1999). Age-related changes in human vocal tract configurations and the effects on speakers' vowel formant frequencies: a pilot study. *Logopedics Phoniatrics Vocology*, 24(3), 132–137.



Questions from defense (November 2, 2015)

With respect to “dumb learning” – why the assumption that listeners have no expectations about how the voice will sound? Based on everything that we’ve heard before? Expect speech to sound like it always had? Adapt to whatever deviation can be attributed to a speaker? Why is this dumb learning?

It isn't clear why the dumb vs. smart learning is different? They aren't orthogonal to each other?

Why would one even think that an accent would be “external” ---

Difference in prosody for the Greek and English females?

Duration of fricatives?

EXP 2 – Significant 2-way interaction between phoneme bias and speaker?

Data showing acoustics of test continua for EXPs 2 and 3? Where are the error bars coming from? Wasn't there only one test continuum?

Following the fricative information – what information do you mean?

You're testing in different acoustic-phonetic space --- doesn't that constrain this learning (swapped fricative) account?