Lexically-guided phonetic retuning of foreign-accented speech and its generalization

Eva Reinisch[a] & Lori L. Holt[a]

[a]Department of Psychology, and Center for the Neural Basis of Cognition, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, 15213, USA

Running head: phonetic retuning in a foreign accent and its generalization

IN PRESS: JOURNAL OF EXPERIMENTAL PSYCHOLOGY: HUMAN PERCEPTION AND PERFORMANCE

Corresponding author:

Eva Reinisch, PhD

now at: Dept. of Phonetics and Speech Processing

Ludwig Maximilian University Munich

Schellingstr. 3

80799 Munich

Germany

E-Mail: evarei@phonetik.uni-muenchen.de

Abstract

Listeners use lexical knowledge to retune phoneme categories. When hearing an ambiguous sound between /s/ and /f/ in lexically unambiguous contexts such as "gira[s/f]", listeners learn to interpret the sound as /f/ since "gira[f]" is a real word and "gira[s]" is not. Later they apply this learning even in lexically ambiguous contexts (perceiving “knife” rather than “nice”). Although such retuning could help listeners adapt to foreign-accented speech, research has focused on single phonetic contrasts artificially manipulated to create ambiguous sounds whereas accented speech varies along many dimensions. It is therefore unclear whether analogies to adaptation to accented speech are warranted. In the present studies, the to-be-adapted ambiguous sound was embedded in a global foreign accent. In addition, conditions of cross-speaker generalization were tested with focus on the extent to which perceptual similarity between two speakers’ fricatives is a condition for generalization to occur. Results showed that listeners retune phoneme categories manipulated within the context of a global foreign accent, and that they generalize this short-term learning to the perception of phonemes from previously unheard speakers. However, generalization was observed only when exposure and test speakers' fricatives were sampled across a similar perceptual space.

193

Listeners are surprisingly good at understanding what different speakers say despite the acoustic variability introduced by differences in vocal tract size (e.g., male vs. female speakers; Strand & Johnson, 1996), dialects (e.g., Clopper & Pisoni, 2004), foreign accents (e.g., Bradlow & Bent, 2008), and speech impairments (e.g., dysarthria; Liss, Spitzer, Caviness, & Adler, 2002), to name just a few sources of variability. One reason for this success is that speech perception capitalizes on context to resolve acoustically ambiguous speech. Lexical context (e.g., Norris, McQueen, & Cutler, 2003; Kraljic & Samuel, 2005), the co-variation between acoustic cues (e.g., Idemaru & Holt, 2011), visual information from the speaker's lip movements (e.g., Bertelson, Vroomen, & de Gelder, 2003) and written subtitles (Mitterer & McQueen, 2009) have all been demonstrated to impact speech perception. Norris et al. (2003), for example, created an ambiguous sound between /f/ and /s/ and tested how lexical information that biased perception to /f/ versus /s/ led listeners to later interpret the acoustically ambiguous sound. When the ambiguous sound was heard in contexts where it could only be interpreted as /f/ (as e.g., in "gira_" because "gira[f]" is a word but "gira[s]" is not), listeners later categorized stimuli along an [ɛf]-[ɛs] continuum more often as /ɛf/ than listeners who had heard the ambiguous sound replace /s/ (e.g., "notice"; note that the Norris et al. study was conducted in Dutch but English examples are given for illustration). That is, experiencing an ambiguous sound in lexically-disambiguating contexts led subsequent perception of the sound to be less ambiguous, and more consistent with the sound indicated by lexical context.

This finding has been replicated and extended multiple times using different tasks during exposure (e.g., counting the number of words: McQueen, Norris & Cutler, 2006; listening to a story: Eisner & McQueen, 2006) and test (e.g., cross modal priming: McQueen, Cutler, & Norris, 2006, Sjerps & McQueen, 2010; eyetracking: Poellmann, Mitterer & McQueen, 2011, Mitterer & Reinisch, 2012, in press). In general, these findings and their replications and extensions (see also, Samuel & Kraljic, 2009a, for an overview) have been thought to reflect processes that adjust speech perception to accommodate

acoustic variability arising from deviations from the norm of the native language, as in listening to foreign-accented speech. However, typically studies of such lexically-guided "perceptual retuning" have manipulated single pairs of phonemes in native speech (e.g., Norris, et al. 2003; see Kraljic & Samuel, 2009) and, by design, only detailed acoustic cues related to a single phonemic contrast have been manipulated to create the artificially "accented" speech (e.g., the /f/ and /s/ in the example above; Norris et al., 2003). Of course, real-world acoustic speech variability, such as arises in foreign-accented speech, deviates from native speech along many more dimensions. An open question therefore is whether the kind of lexically-guided phonetic retuning observed for carefully-controlled acoustic manipulations extends to more natural circumstances where multiple acoustic dimensions are impacted by a foreign accent.

Studies of word recognition for speech with interfering noise (McQueen & Huettig, 2012) and for casual speech (Brower, Mitterer, & Huettig, 2012), suggest the possibility of "global" retuning such that familiarity with an accent could lead to a general loosening or reevaluation of constraints for signal-to-word mappings that are independent of specific mispronounced segments. However, studies of adaptation to natural foreign accents suggest a mechanism of lexically-guided retuning to global foreign accents similar to that observed for artificially manipulated single segments in native speech (e.g., Bradlow & Bent, 2008). It appears that listeners use lexical knowledge to interpret accented words having acoustic variability that deviates from the native norm in a manner that facilitates later perception of foreign-accented speech. For example, when native English listeners were asked to transcribe Chinese-accented sentences in noise, over time they got better at performing this task (Bradlow & Bent, 2008). The rate of adaptation depended on baseline intelligibility of the accented speech, with faster adaptation to more intelligible speech. Bradlow and Bent suggested that this is because intelligibility relates to access to lexical information that can guide adaptation to accented speech. However, such studies have focused on tasks measuring changes in comprehension of foreign-

accented spoken words (see also, e.g., Mitterer & McQueen, 2009; Sidaras, Alexander, & Nygaard, 2009) and so it remains unclear whether the kind of lexically-guided retuning observed by Norris et al. (2003) for single segments is involved in adjustments in perception elicited by exposure to natural global foreign accents.

A study by Reinisch, Weber, and Mitterer (2013) suggests there may be a relationship. The study consisted of a series of lexically-guided category-retuning experiments similar to the original study by Norris et al. (2003). However during exposure, one group of Dutch native listeners performed a lexical decision task in Dutch whereas another group of listeners performed the lexical decision task in English (spoken by the same Dutch speaker; both groups of Dutch listeners were proficient in English). Within each of these "language groups" half of the listeners heard tokens of /s/ being replaced by an ambiguous sound whereas the other half heard the ambiguous sound replace /f/. At test all groups categorized five different Dutch minimal word pairs produced by the same speaker they heard during exposure. For listeners in the English language exposure group the test language changed from English to test Dutch. Importantly, category retuning could be shown for both language groups suggesting that retuning generalizes across language input and occurs even when exposure to the ambiguous sounds is provided in a second language (this latter finding was confirmed by testing native German learners of Dutch on the Dutch version of the experiment). Critically, as the speaker for the Dutch and English materials was the same, words in the English exposure condition were spoken with a perceptible Dutch accent. This indicates that specific non-canonically pronounced segments can be subject to phonetic retuning via lexical context, even when embedded in a foreign accent. However, the native Dutch listeners in Reinisch et al. (2013) were nonnative speakers of English highly familiar with Dutch-accented English. They may have adapted to many characteristics of the Dutch-accented English through long-term exposure (see Witteman, Weber, & McQueen, 2013). In this case the only new deviation that

would have required phonetic retuning would have been the single artificially-manipulated fricative contrast, weakening the case for phonetic retuning within the context of a global foreign accent.

In the present study we used a more rigorous test to assess whether single phoneme retuning via lexical context can be measured in the context of a global foreign accent. The same Dutch-accented English exposure materials used in Reinisch et al. (2013) were presented to native speakers of American English who were unfamiliar with the Dutch accent. If phonetic retuning is found within the global foreign accent, it would provide evidence that laboratory studies of adaptation to single artificially-manipulated sounds model a process available to listeners encountering natural foreign-accented speech. Moreover, such a finding would suggest that segment-by-segment retuning in the context of multiple sources of acoustic variability arising from foreign accent is plausible.

The demands for adaptation to a real-world foreign accent vs. single artificially-manipulated segments may differ on another level, however. In the literature on adaptation to natural foreign-accented speech it has been found repeatedly that exposure to multiple speakers of an accent leads to better comprehension of new speakers with the same accent (e.g., Bradlow & Bent, 2008; Sidaras et al. 2009). That is, when listeners were asked to transcribe multiple speakers' sentences during exposure, performance on a new speaker's sentences of the same accent was also improved as compared to no training or even training with only one accented speaker (Bradlow & Bent, 2008). Additionally, exposure to speakers of a variety of different accents during exposure appeared to improve comprehension of foreign-accented speech independently of the specific accents heard during exposure or test (Baese-Berk, Bradlow, & Wright, 2013). It has thus been suggested that listeners extract information about commonalities in the foreign accents in a specific target language and apply this information to the comprehension of a new talker. This finding goes well with the common (anecdotal) experience that understanding a given foreign accent gets easier over time.

In contrast, talker generalization of phonetic retuning elicited by speech manipulated to create acoustic ambiguity for a single phonetic contrast in the context of otherwise native speech (e.g., Norris et al. 2003) has been mixed (but note the task differences between accent and native speech studies discussed above). Whereas lexical retuning along the voiced-voiceless dimension of stop consonants appears to transfer across speakers (Kraljic & Samuel, 2006, 2007), retuning for fricatives has been argued to be speaker-specific (Eisner & McQueen, 2005; Kraljic & Samuel, 2005; 2007). Kraljic and Samuel (2007) suggested that the spectral cues in fricatives convey more information about the speaker than the durational cues in stops; hence, due to this "double role" for fricatives (providing information about the segment and speaker), they are retuned speaker-specifically whereas stops are not.

Eisner and McQueen (2005) suggested that adaptation is speaker-specific but operates at the phoneme level. They found that when the voice during exposure was speaker A, the voice during test was speaker B but all fricatives during exposure and test came from the same speaker (either A or B), category retuning could be observed (Eisner & McQueen, 2005). This was despite the fact that listeners clearly perceived the presence of different voices during exposure and test. Hence it appears that for the purpose of category retuning listeners don't track voices in general but the speakers' specific pronunciation variants. Kraljic and Samuel (2005) suggested that the decisive factor in cross-speaker generalization for fricatives is the acoustic similarity between the fricatives heard during exposure and those categorized at test. They found generalization from a female speaker heard during exposure to a male speaker at test but not the reciprocal generalization pattern. Acoustic measurements of the fricatives revealed that the female speaker's fricatives during exposure fell within the range of the male speaker's test continuum whereas the male speaker's fricatives during exposure were acoustically distinct from female speaker's test continuum. Listeners thus seem to track acoustic properties of each speaker's fricative productions and apply generalization whenever there is a sufficient match.

However, the fact that results on speaker generalization are mixed suggests that attention to detail is required to further refine the constraints under which lexically-guided phonetic retuning is observed. Thus the second purpose of this study was to test whether the same speaker-specificity for fricatives manipulated to create acoustic ambiguity would apply in the presence of a global foreign accent. Note that cross-speaker generalization of foreign accents has been shown to occur (see discussion above). Given that retuning of single category contrasts can be measured within a foreign accent (our first question) the presence of a foreign accent might facilitate cross-speaker generalization of the retuned categories. In addition we manipulated cross-speaker similarity in the fricative test continua putting Kraljic and Samuel's (2005) explanation to test that similarity between the speakers' productions of the critical segments is the decisive factor of whether cross-speaker generalization of category retuning can be found.

In three experiments listeners heard a female Dutch learner of English produce words and nonwords in a lexical decision task in which either word-final /f/ or /s/ was replaced by an ambiguous sound. At test listeners categorized English minimal word pairs spoken by the same speaker heard during exposure and one of two new "generalization" speakers. In Experiment 1 the generalization speaker was another female Dutch learner of English. In Experiments 2 and 3 the generalization speaker was a male Dutch learner of English. Accents were the same between all speakers but the presumed similarity of the voices (male vs. female) differed. Critically, for the male generalization speaker we introduced an additional within-speaker manipulation of perceptual similarity to the exposure speaker's fricatives. Whereas in Experiment 2 the generalization speaker's continuum spanned the range between his natural endpoint productions of /f/ and /s/, in Experiment 3 we selected a subset of these stimuli from the /f/-side of the continuum to better match the sampling of perceptual space between the continua of the female exposure speaker and the male generalization speaker (for details see the

Methods section below). This manipulation of *perceptual* similarity in the idiolects complements previous suggestions of the need for *acoustic* similarity between the two speaker's fricatives.

Evidence for the fact that perception of sounds or voices rather than acoustics may be the decisive factor in category retuning comes from studies investigating conditions under which retuning of a single phoneme contrast is blocked (e.g., Kraljic & Samuel 2005, 2011; Kraljic, Samuel, & Brennan, 2008). One such condition occurs when, in addition to ambiguous tokens, listeners also hear the same speaker produce good, perceptually unambiguous tokens of the critical fricatives (Kraljic & Samuel 2005). However, new evidence suggests that this effect depends upon who listeners believe they have heard more than the acoustic characteristics of the speaker (Samuel & Kraljic, 2009b, submitted). In their study listeners heard a voice produce good tokens of the critical fricatives and, subsequently heard ambiguous tokens from the same voice. Critically, during this exposure listeners saw videos of the same speaker or different speakers presumably articulating unambiguous vs. ambiguous sounds. Results showed that retuning was blocked when the ambiguous and unambiguous tokens were paired with a video of a single talker, but was evident when videos of different talkers suggested that the ambiguous and unambiguous sounds originated from different talkers.

In the present experiment we address whether studies of single artificially-manipulated phoneme contrasts in native speech can indeed figure as a model for studying the processes involved in adaptation to a global foreign accent, and examine the conditions under which listeners generalize phonetic retuning to a previously unheard speaker. By manipulating the perceptual similarity of the new speaker to the speaker heard during exposure we sought to gain insight into the mechanisms of lexically-guided phonetic category retuning.

EXPERIMENT 1

In Experiment 1 we first addressed whether listeners show phonetic category retuning of an artificially manipulated phonetic contrast even when the ambiguous sound is embedded in a global foreign accent during exposure. The goal was to examine how studies of phonetic category retuning in native speech relate to adaptation to natural foreign accented speech where more than one speech characteristic deviates from the native norm. More specifically, we asked whether a segment-by-segment retuning to foreign-accented speech, in this case Dutch-accented English, is plausible. Secondly, we addressed whether listeners generalize phonetic retuning to a previously unheard speaker. Since previous studies have suggested that the similarity of the speakers' voices and specifically the similarity of their phonetic segments is crucial for cross-speaker generalization – if generalization occurs at all - we tested a condition in which transfer was favorable: fricative categorization was tested for the same Dutch-accented female speaker heard during exposure and a new Dutch-accented female speaker. We predict no transfer of lexically-guided retuning if category retuning is speaker-specific, as has been suggested previously (e.g., Eisner & McQueen, 2005; Kraljic & Samuel, 2007). If, however, generalization is influenced by the similarity between the speakers' fricatives (as first suggested in Kraljic & Samuel, 2005) and/or the presence of a foreign accent, then category retuning should be observed for the exposure speaker and the female generalization speaker (note that Experiments 2 and 3 will address the issue with a male generalization speaker).

Methods

Participants.

Twenty-eight participants from the student population of Carnegie Mellon University and the University of Pittsburgh participated for pay. Ten additional participants from the same population took part in the pretest of the test continua. All participants reported they were native speakers of American English with no history of hearing problems.

Materials.

Following the methodology of Norris et al. (2003), stimulus materials included words and nonwords for exposing participants to the accented speech via a lexical decision task. At test we examined the impact of exposure on subsequent fricative categorization by presenting listeners with an /f/-to-/s/ continuum embedded in word-final position of minimal word pairs. Listeners had to categorize which sound/word they heard. The materials for the exposure phase were the same as in Reinisch et al. (2013; Experiment 3). They consisted of 40 critical words, 60 filler words, and 100 nonwords. Twenty of the critical words were /f/-final and did not create another existing English word when /f/ was replaced with /s/ (e.g., *belief*). The other 20 critical words were /s/-final and did not create another English word when /s/ was replaced with /f/ (e.g., *notice*). Nonwords were created to be phonotactically legal sequences in English. Filler words and nonwords did not contain /f/, /s/ or the acoustically/articulatorily similar sounds /v/, /z/, /θ/, or /ð/. Four English minimal word pairs ending in /f/ and /s/ were selected as test items for phonetic categorization (*elf-else, knife-nice, leaf-lease, graph-grass*). The use of multiple word pairs at test should discourage listeners from simply comparing the critical sounds on a trial-by-trial basis and rather focus their attention on the functional role of the critical phonemes in distinguishing existing words[1].

All words and nonwords were recorded by a female Dutch native speaker (age 28) in a sound-proof booth. The speaker had started to learn English at age 12 and at the time of recording she used it on a daily basis for her studies and work. Despite her high proficiency and fluency in English, colleagues who are native speakers of English typically characterize her speech as having a perceptible accent. This

[1] Note that differences in frequency between s-final and f-final members of the minimal pairs as counted in the CELEX lexical database (Baayen, Piepenbrock, & Glikers, 1995; with the s-final member being the more frequent one in three of the four pairs) are the same for both exposure groups. Possible interactions between word frequency and category retuning remain to be shown.

was confirmed by the accent ratings in the present study (see Results sections). Critical words from the exposure phase were recorded in pairs, once correctly and once with the word-final fricatives exchanged. That is, *belief* was also recorded as *belie[s]* and *notice* was also recorded as *noti[f]*. Exchanging the word-final fricatives always resulted in nonwords. The speaker was asked to produce both forms of the words (i.e., correct and with fricatives exchanged) with a comparable speech rate, speech style, and intonation contour.

The minimal word pairs for the test phase were recorded by the same speaker who recorded the exposure words as well as by another female and a male native speaker of Dutch (henceforth referred to as "female generalization speaker" and "male generalization speaker"; note that aside from the pretest described below the male generalization speaker's recordings were used only in Experiments 2 and 3). At the time of recording both generalization speakers were in their mid-twenties and used English on a daily basis for their work. Both speakers are characterized by native English speaking colleagues as being fluent in English but speaking with a perceptible Dutch accent. Speakers were asked to produce the members of each minimal pair as similarly as possible by only substituting the critical word-final fricatives.

All words were equalized in their overall RMS amplitude. The creation and selection of ambiguous sounds for exposure are extensively described in Reinisch et al. (2013) and are briefly summarized. To select ambiguous sounds for exposure approximately the last syllables of the two recordings per word were excised and morphed into an 11-step continuum (for the morphing procedure see below). Splicing was done at positive going zero crossings using Praat (Boersma & Weenink, 2009) at points in the signal where a significant acoustic change could easily be identified in both recordings (e.g., at the onset of a vowel as indicated by the start of voicing after a voiceless portion of the signal). Morphed syllables were then spliced back onto the word stems. The word stems were selected from the

word or nonword (i.e., with the fricatives were exchanged) recordings depending on the naturalness of the resulting tokens. The most ambiguous tokens of these continua were taken from the Reinisch et al. (2013) study. These tokens had been selected in a pretest (reported in Reinisch et al.) in which all forty continua were presented to Dutch native listeners for phonetic categorization. To reduce the number of trials, only seven of the eleven morphs were used (steps 2, 4, 5, 6, 7, 8, and 10). Participants' task was to indicate by button press whether the last sound of an item sounded like an /f/ or /s/ irrespective of whether the result would be a real word. In order to select the most ambiguous token of the fricative, we compensated for the expected bias towards the respective word endpoints (Ganong, 1980). Continuum steps of s-final words were selected as ambiguous when they received approximately 30 % f-responses, and steps of f-final words were selected when they received about 70 % f-responses. Note that due to this procedure the acoustics of the fricatives varied by critical word and were not acoustically identical across words. The fact that not only the word-final fricatives but rather larger portions of the words were morphed (mostly the last syllable) ensured that cues to the fricatives other than the frication noise were also ambiguous. Reinisch et al. (2013) argued that this procedure, in addition to the use of multiple minimal pairs at test, leads to reduced variability in the degree of retuning among participants as compared to previous studies of lexically-guided phonetic category retuning (i.e., in a direct comparison to participants' performance in Norris et al., 2003, and Eisner & McQueen, 2005).

The minimal pairs used at test were morphed in their entirety in an 11-step continuum using the STRAIGHT algorithm (Kawahara, Masuda-Katuse, & Cheveigné, 1999) in Matlab (The MathWorks Inc.). The same method had been used to morph the last syllables of the critical words for exposure. The morphing procedure consisted of a stepwise mixture of the respective sounds such that the resulting morphs contained an increasing amount of the /f/-final signal. The eleven continuum steps spanned the whole range form 0 % to 100% of the /f/-final recording. The morphing algorithm decomposes the

speech signal into a voice source, a noise source, and a dynamic spectral filter with time windows of 10 ms. Interpolation is achieved by first mixing the parameters, and then generating a new signal from these mixtures. We used a time-aligned version of the algorithm to encourage morphing across similar segments of the speech signals. Temporal anchors were set at points in the signal where there were significant acoustic changes, as informed by acoustic phonetics. For example, for the minimal pair knife-nice, an anchor was set at the end of the low-frequency noise characteristic of nasals, at the onset of voicing (indicated by fundamental frequency) of the vowel, and at the offset of voicing and onset of frication noise. In this way only segments of the same type were morphed (i.e., nasals with nasals, vocalic portions of the signal with other vocalic portions, fricative noise with fricative noise, etc.). Duration differences between two morphed segments were also interpolated. That is, if a vowel in one utterance was 80 ms long and 100 ms in another, then at the 50% morph the resulting resynthesized vowel would be 90 ms long.

Pretest.

Method and Procedure. The newly-created minimal pairs for the test phase were subjected to a pretest to establish that native English listeners perceive a continuum from /f/ to /s/ for all three Dutch speakers. We compared the categorization curves of the speakers' continua, and, additionally tested whether the three voices could be identified as belonging to three different speakers. Acoustic analyses of the minimal pairs suggested that the three speakers should be easy to distinguish. Among other possible differences, the speakers clearly differed in their fundamental frequencies (F0). The female exposure speaker's F0 was highest with an average of 212 Hz; the female generalization speaker's F0 was 166 Hz, and the male generalization speaker's F0 was 100 Hz.

The pretest consisted of two consecutive parts. First, participants were presented seven steps of the morphed eleven-step continua of all four minimal pairs from all three speakers. The selected

morphs contained 10%, 30%, 40%, 50%, 60%, 70%, and 90% of the original /f/-final stimuli (henceforth referred to as steps 1, 3, 4, 5, 6, 7, and 9). Participants were seated in a sound-attenuated booth and listened over headphones to the words. Their task was to indicate by button press whether the last sound of an item sounded like an /f/ or /s/. Each continuum step of each word and speaker was presented 4 times resulting in a total of 336 trials. Words and speakers were randomly intermixed with the restriction that all stimuli were presented once before a repetition occurred.

During the second part of the pretest, listeners were presented with the re-synthesized endpoints of the minimal pair continua (i.e., the morphs containing 0% and 100% of the /f/-final stimulus; 8 different words per speaker). First listeners were given the chance to learn voice-name associations. Listeners first heard four of the words spoken by the exposure speaker while the name "Lisa" was displayed on the screen, then they heard four words by the female generalization speaker named "Anna", and then four words by the male generalization speaker (named "Peter"). In a second block the other four words of each speaker were presented. Then listeners were asked to perform a speaker categorization task. They were presented one word at a time and their task was to indicate by button press which of the three speakers they heard. Button-name associations were displayed on the screen throughout the experiment and participants used the number keys 1 to 3 to indicate their response. No feedback was given. Immediately after the response was registered, listeners were prompted to rate on a scale from 1-7 how confident they were in their decision (7 = *very confident*, 1 = *not confident at all*). Each word (continuum endpoint) was presented three times resulting in a total of 72 trials.

(insert Figure 1 about here)

Results. Figure 1 shows the categorization responses to the minimal pair continua for each of the three speakers. As evident from the figure, the response functions of the three speakers' continua

differ, especially at the /s/-side of the continuum. Whereas listeners had a strong /s/-bias for the male speaker's continua, a strong /f-/bias was observed for the female exposure speaker. The categorization function for the female generalization speaker's stimuli fell in between, patterning with the male speaker at the /s/-side of the continuum and patterning with the female exposure speaker at the /f/-side of the continuum. This was despite the fact that the endpoints of the continua matched the speakers' natural productions (i.e., the 0% and the 100 % steps of the morphs are a simple resynthesis of the speakers' productions). However, given that the voices of the speakers were intermixed, it is possible that listeners interpreted the different speakers' fricatives in relation to each other, as spanning a single perceptual space. This finding is informative as in the test phase of the category retuning experiments the speakers' voices will be intermixed as well (for details see below) and consequences on retuning will be discussed.

The observations from Figure 1 were confirmed in a statistical analysis. Listeners' proportion /s/-responses were analyzed using a linear mixed-effects model (Baayen, Davidson, & Bates, 2008) as provided in the lme4 package (Bates & Sarkar, 2007) in R (*version 2.15.1*; The R foundation for statistical computing). Linear mixed-effects models have been argued to be superior to traditional analyses using ANOVAs as they are less susceptible to Type-I errors (Quené & van den Bergh, 2008), especially with dichotomous dependent variables (Jaeger, 2008) as we are analyzing here (i.e., response is /f/ vs. /s/). To account for the dichotomous dependent variable we used a logit linking function that gives more weight to differences near the floor and the bottom of the probability scale. We report up to four terms per fixed factor. The regression weight (beta-score), a z-score (based on Wald's z-score) that indicates the coefficient's distance from zero in terms of its standard error (Jaeger, 2008), the standard error and the p-values associated with the factor.

Two fixed factors and their interaction were entered into the model: Speaker and Continuum Step. The factor Speaker had three levels (F1 = exposure speaker, F2 = female generalization speaker, M1 = male generalization speaker) of which the factor F1 would be mapped onto the intercept and regression weights for the other two levels would indicate differences between these levels and the level mapped onto the intercept. The factor Continuum Step was entered as a numeric factor centered on zero such that the intercept would indicate overall effect of Continuum Step for the level of the other factor mapped onto the intercept. Participant was entered as a random factor for which an intercept as well as slopes for the within-participant fixed factors and their interaction was estimated. This allowed the intercept of the regression model as well as slopes of the within-participant factors to vary by participant with the restriction that the mean of this random variation was zero (Baayen, Davidson, & Bates, 2008; see Barr, Levy, Scheepers, and Tily, 2013, for a discussion of the necessity for a complete random effects structure including random slopes for all within participant factors). This should minimize chances that fixed effects would be significant due to random by-participant variation.

As Figure 1 suggests, responses for the female speaker used to create exposure stimuli significantly differed from the male generalization speaker ($b_{SpeakerM1}$ = 4.36, *SE* = 0.44, z = 9.94, $p < .001$) but did not differ from the female generalization speaker ($b_{SpeakerF2}$ = 0.10, *SE* = 0.31, z = 0.33, $p = .77$). More /s/-responses were given for the male speaker than for the two female speakers. There was also an effect of Continuum Step (b_{Step} = -0.69, *SE* = 0.11, z = -6.07, $p < .001$) suggesting that listeners gave more /s/-responses the more /s/-like the fricatives of the exposure speaker were. The continuum manipulation for the exposure speaker was thus successful, despite the somewhat lower proportion /s/-responses at the /s/-endpoint of the continuum as compared to the other speakers. Continuum Step, however, interacted with the other two levels of the factor Speaker. On the more /s/-like side of the continua significantly more /s/-responses were given for the two generalization speakers than for the exposure speaker ($b_{Step*SpeakerF2}$ = -0.44, *SE* = 0.11, z = -4.15, $p < .001$; $b_{Step*SpeakerM1}$ = -0.79, *SE* = 0.15, z = -

5.03, $p < .001$). That is, the categorization function of the exposure speaker's continuum was less steep than those of the generalization speakers.

With regard to the speaker identification task, listeners were very good at labeling the three talkers. Overall correct identification of the female exposure speaker was at 93.4%[2] and received an average confidence rating of 6.2 out of 7 (7 = *very confident*). The female generalization speaker was identified correctly 95.9 % of the time (confidence 6.3), and the male voice was identified 100% of the time (confidence 6.97). Overall, the pretest established that the manipulation of the test continua was successful even though differences in the categorization functions were found (i.e., overall more /s/-responses for the male generalization speaker, and more /s/-responses for both generalization speakers at the /s/-like side of the continua). These differences will be taken into account when discussing results on category retuning. Importantly, all speakers' voices could be identified with close to ceiling performance; even the voices of the two female speakers were clearly identifiable.

Procedure.

Exposure. All participants heard the female exposure speaker produce the English words and nonwords in her Dutch-accented English. All participants heard the same 60 filler words and 100 nonwords. Half of the participants were randomly assigned to the /f/-ambiguous condition and were presented with the 20 /f/-final words in which the /f/ had been replaced by a perceptually ambiguous sound between /f/ and /s/ and 20 /s/-final words in which the /s/ was naturally produced. The other half was assigned the /s/-ambiguous condition whereby the 20 /s/-final words for which the /s/ was replaced by a perceptually ambiguous sound between /f/ and /s/ and 20 /f/-final words produced naturally.

[2] One participant responded incorrectly on all trials involving the two female voices suggesting confusion of the names. Therefore responses for this participant were re-labeled to 100% correct identification.

Participants were seated in a sound-attenuated booth and were informed that they would hear a non-native learner of English. On each trial, participants indicated whether they heard an existing English word or not by pressing Key 1 or Key 2, respectively. Response options ("word", "not a word") were displayed (to the left and right of the screen, respectively) 500 ms before the audio started. The response options remained onscreen until response, which was indicated by a shift of the response option approximately 1 cm upwards and outwards on the screen for 400 ms. After a 500 ms pause, the next trial began. If a participant did not respond within 4 seconds "No answer registered" was displayed on the screen in red letters and the experiment proceeded to the next trial. The instructions emphasized speed as well as accuracy of response.

Words and nonwords were presented in random order. Every 50 trials participants were allowed to take a self-paced break. At the end of the exposure phase participants rated the accent of the exposure speaker (1 = *very strong accent*, 7 = *like a native English speaker*) using the computer keypad.

Test. Immediately following exposure participants from both the /s/-ambiguous and the /f/-ambiguous exposure conditions completed a phonetic categorization task with four English minimal pairs intermixed across tokens from the exposure speaker and the female generalization speaker (stimuli of the male generalization speaker were used in Experiments 2 and 3). Intermixing the two speakers' fricatives at test allowed for a within-participant comparison of the basic retuning effect for the exposure speaker's continua as well as a test of generalization to the new speaker. Participants were informed that they would hear multiple speakers produce English words and should decide whether the words ended in an /f/ or /s/. Note that participants were not informed of the number of speakers or whether new speakers had a foreign accent. However, we expected that intermixing the speakers' productions might facilitate the perception of the common accent. Each trial started with the presentation of the response options "…s" and "…f" on the screen (to the left and right of the screen,

respectively) for 500 ms after which the audio was played. Participants pressed the '1' key (for “s”) or the '2' key (for “f”) on the computer keyboard to indicate whether they heard an /f/ or /s/ at the end of the word, causing the response alternative to shift on the screen to indicate the response had been logged. The next trial began after 500 ms or a 4 second time-out period.

The same 7 steps of the two female speakers’ continua presented in the pretest were used during test. Each stimulus was presented 5 times resulting in a total of 280 trials (5 repetitions x 2 speakers x 7 steps x 4 minimal pairs). All different types of stimuli were presented once before a repetition occurred. Every 56 trials participants were allowed to take a self-paced break. After completion of the categorization task, participants answered several written questions using the computer keyboard: (1) How many speakers did you hear? (2) Did the speakers have the same accent? (3) Please guess what the native language of the speakers was. (4) The speakers’ native language was Dutch. On a scale from 1 to 7, how familiar are you with Dutch accented English? 1 = *not familiar at all*, 7 = *very familiar*. The experiment was implemented in E-Prime (version 2.0, Psychology Software Tools, Inc.) and took approximately 20 minutes to complete.

Analyses.

Data from participants who responded faster than 200 ms or slower than 2500 ms across more than 5% of all exposure and test trials were excluded (N=4) to ensure a sample with similar speed-accuracy trade-offs. Further, single trials that fell outside this reaction-time window were excluded from all analyses (292 trials or 2.2%). In previous studies of lexically-guided category retuning (e.g., Norris et al. 2003; Sjerps & McQueen, 2010) acceptance of at least half of the critical words with ambiguous sounds as permissible English words in the exposure phase was used as an inclusion criterion; all participants met this criterion. Table 1 shows the acceptance rates for critical words with ambiguous and unambiguous fricatives in the lexical decision task in all three experiments.

(insert Table 1 about here)

Listeners' responses during the categorization test were analyzed using linear mixed-effects models for the same reasons described for the Pretest above. Again, a logistic linking function was used, to take into account the categorical nature of the dependent variable (an /s/-response was coded as 1 and an /f/-response was coded as zero). Fixed effects were Exposure Condition (/f/-ambiguous or /s/-ambiguous), Speaker (F1 = exposure speaker, F2 = female generalization speaker), and the interaction of these factors. Continuum Step was entered as a numerical factor centered on zero. Overall, factors were coded such that the logistically transformed proportion of /s/-responses given by participants in the /f/-ambiguous condition for the exposure speaker was mapped onto the intercept at the middle step of the continuum (step zero after centering). The regression weight for Exposure Condition would then indicate whether participants in the /f/-ambiguous condition performed differently from the participants in the /s/-ambiguous condition for the exposure speaker. That is, an effect of Exposure Condition would indicate that adaptation to mispronounced phonemes can be measured by a shift in /s/-/f/ categorization even when the critical phonemes occurred in a global foreign accent during exposure. A significant effect of Speaker would indicate that the categorization function (proportion /s/-responses) for the generalization speaker differs from the exposure speaker. Critically, a significant interaction between Exposure Condition and Speaker would indicate a difference in the magnitude of categorization shifts for the two speakers. In other words, a non-significant interaction would suggest transfer of what has been learned about the exposure speaker's pronunciation to the generalization speaker. Since however, we have to be cautious predicting a null result, follow-up analyses were carried out to test for an effect of Exposure separately for responses to the exposure speaker and the generalization speaker. If an effect of Exposure were found for both speakers, then transfer of category retuning would be confirmed. A significant effect of Continuum Step is expected in all analyses and would show that more /s/-responses are given the larger the proportion of /s/ in the morphed fricative.

The random effect structure of the mixed-effects models included a random intercept for participants with random slopes for Speaker and Continuum Step over participants. A random slope for Exposure Condition over participants is not meaningful since it is a between participant factor (see Barr et al. 2013). Random effects over items were not included since we only tested four minimal pairs, a number small enough to render this effect negligible.

Results

Questionnaires.

The average perceived accent strength of the exposure speaker was 3.8 (range 3-6; 2 participants did not respond; 1 = *very strong accent*, 7 = *like a native English speaker*; this question was asked after the exposure phase). At test most listeners correctly identified that they heard two speakers (20 out of 28 participants). Six listeners indicated that they heard three speakers, one claimed to have heard four speakers, and one further participant did not respond to this question. Interestingly, all but 1 participant stated that the two speakers did *not* have the same accent. Informal interviews of participants indicated that they mostly considered the second speaker a native speaker of English. None of the participants guessed that the accent they heard was Dutch; 6 participants guessed German. Participants reported to be unfamiliar with Dutch accented English (mean 1.68, range 1-4; 1 = *not familiar at all*, 7 = *very familiar*).

Categorization.

(insert Figure 2 about here)

Figure 2 shows the proportion /s/-responses by listeners from the /f/-ambiguous and the /s/-ambiguous conditions for the exposure speaker (Panel A) and the female generalization speaker (Panel B). For both speakers the categorization functions appear to differ by Exposure group (difference in /s/-

responses between the /s/-ambiguous and /f/-ambiguous group for the exposure speaker is 17% and for the generalization speaker is 12%) suggesting that lexically-guided category retuning does occur even when the speaker has a perceptible foreign accent (i.e., the exposure speaker) and that this retuning generalizes to the female generalization speaker. This is confirmed by statistical analyses: Categorization of the English minimal pairs was influenced by exposure to perceptually ambiguous /s/-/f/ stimuli in lexically-biasing contexts ($b_{Condition}$ = 1.13, *SE* = 0.37, z = 3.08, $p < .005$). Participants in the /s/-ambiguous condition gave more /s/-responses than participants in the /f/-ambiguous condition. The effect of Speaker ($b_{Intercept}$ = -0.41, *SE* = 0.26, z = -1.55, $p = .12$, $b_{SpeakerF2}$ = -0.44, *SE* = 0.18, z = -2.41, $p < .01$) suggests that fewer /s/-responses were given for the generalization speaker than for the exposure speaker. Critically, despite the difference in overall /s/-responses the interaction between Exposure Condition and Speaker was not significant ($b_{Condition*Speaker}$ = -0.28, *SE* = 0.26, z = 1.10, $p = .27$). As expected, the effect of Continuum Step was significant, showing that more /s/-responses were given the more /s/-like the stimulus was (b_{Step} = -0.75, *SE* = 0.06, z = -12.35, $p < .001$).

To confirm that the effect of Exposure Condition was significant for both speakers, separate analyses for each of the speakers were run. These analyses confirm that the effect of Exposure Condition was significant for the exposure speaker ($b_{Intercept}$ = -0.41, *SE* = 0.24, z = -1.73, $p = .08$; $b_{Condition}$ = 1.01, *SE* = 0.32, z = 3.11, $p < .001$; b_{Step} = -0.6, *SE* = 0.06, z = -10.4, $p < .001$) as well as for the generalization speaker ($b_{Intercept}$ = -0.97, *SE* = 0.17, z = -5.74, $p < .001$; $b_{Condition}$ = 0.91, *SE* = 0.23, z = 3.94, $p < .001$; b_{Step} = -0.95, *SE* =0.07, z = -13.86, $p < .001$). Although listeners heard the perceptually ambiguous fricative in lexically-biasing exposure contexts only for the exposure talker, the influence on /f/-/s/ categorization appears to have generalized to another female talker.

Discussion

Experiment 1 demonstrated that listeners show lexically-guided phonetic category retuning of an artificially manipulated contrast even when the ambiguous segment is embedded in an unfamiliar global foreign accent possessing many dimensions of acoustic variability. This suggests that experiments investigating phonetic retuning to artificially manipulated segments in native speech can inform us about adaptation to a foreign accent where more than one speech characteristic deviates from the native norm. It also tentatively suggests that a segment-by-segment adaptation to foreign-accented speech is plausible.

The second finding of Experiment 1 is that, at least under the present conditions, listeners appear to generalize lexically-guided phonetic retuning to the foreign-accented speaker heard during exposure to a second, previously unheard, speaker. Listeners in the f-ambiguous exposure group gave more /f/ responses than listeners in the s-ambiguous group not only for the exposure speaker but also for the female generalization speaker. Since the generalization speaker has not been heard before, the between-group differences in the categorization functions for the generalization speaker suggest that retuning for the exposure speaker has been transferred from the exposure speaker to the generalization speaker. In Experiment 1 conditions for cross-speaker transfer were favorable because the second speaker was the same gender as the exposure speaker and the speakers shared the same accent. Unexpectedly, however, all but one listener perceived the speakers as having *different* accents, suggesting that explicit knowledge of the accents originating from the same native language is not a prerequisite for cross-speaker generalization. Also, an overall difference in the percentage /s/-responses for the two speakers (cf. the effect of Speaker) appeared not to influence generalization. However, perceptual similarity between the two speakers' voices (and potentially their fricatives) is suggested by the patterning of the categorization functions and speaker confusions observed in the Pretest. Whereas the overall proportion /s/-responses differed between the male and the female speakers in the Pretest, it did not differ between two female speakers (though the shapes of the functions differed at the /s/-

endpoints of the continua). Moreover, some mistakes were made in the identification of the female speakers but (unsurprisingly) not for the male speaker.

These findings relate to previous studies by suggesting that the fricatives heard between exposure and test need not be identical for cross-speaker generalization to occur. Eisner and McQueen (2005) had suggested that listeners would generalize category retuning only if the fricatives between exposure and test came from the same speaker. As long as the fricatives during exposure and test came from the same speaker, retuning was observed regardless of whether the fricatives were spliced onto word stems spoken by the same or a different speaker. In contrast, in the present study the fricatives of the generalization speaker did not consist of tokens spliced in from the exposure speaker but were morphs of the natural recordings of the generalization speaker. Still, generalization could be found.

This leads us to note a number of differences between this previous study and ours. First, Eisner and McQueen presented only one voice at test, here we intermixed tokens of both speakers. Second, Eisner and McQueen only presented one test continuum, namely [ɛf]-[ɛs] where speaker information outside the fricative was limited to the production of the vowel. Here we presented continua between four minimal word pairs mixed across speakers in presentation. One might expect the present conditions to lead to more prominent differences in retuning for exposure and generalization speakers [3]. We therefore suggest that given a certain degree of *perceptual* similarity in the speaker's voices (e.g., both being female) and fricatives (sampling a similar range in perceptual space), cross-speaker generalization can be found. This is in line with the suggestions by Kraljic and Samuel (2005) who found that given *acoustic* similarity between the two speakers' fricatives cross-speaker generalization does

[3] As discussed for the Pretest we are aware of the possibility that the two speaker's continua could have influenced each other's perception. We will return to this in the discussion of Experiment 2.

occur. Across studies, we are thus accumulating converging evidence that the similarity of the fricatives may be a crucial factor for cross-speaker generalization of phonetic category retuning.

Therefore in Experiment 2 we set out to further specify the notion of "similarity of the fricatives" as a condition of cross-speaker generalization. Following exposure to the same Dutch female talker as in Experiment 1, in Experiment 2 listeners were presented with a Dutch male generalization speaker at test alongside the female exposure speaker.

EXPERIMENT 2

Experiment 2 had two purposes. First we set out to replicate the Experiment 1 finding that lexically-guided retuning of single phonetic contrasts can occur within a global foreign accent. Second, we aimed to further test the conditions across which such retuning generalizes across speakers. By replacing the Dutch female generalization speaker with a Dutch male speaker we increased the differences between the voices. Note that previous research has shown that listeners are sensitive to speaker gender in categorizing fricatives (Munson, 2011; Strand & Johnson, 1996). If the perceived similarity of the speakers' voices or of the specific fricative realization is crucial for generalization (Kraljic & Samuel, 2005) then decreased similarity between speakers' voices may decrease the likelihood of speaker generalization. Note that the pretest categorization functions hint at differences in the perceptual space sampled by the Dutch female and male speakers' fricative continua (see Figure 1). More stimuli along the male generalization speaker's continuum were perceived as /s/ than along the female speakers' continua. Compared to Experiment 1, we would expect less favorable conditions for speaker generalization in Experiment 2 owing to reduced perceptual similarity between the exposure and generalization talkers. Moreover, since the Dutch female generalization speaker in Experiment 1 was not perceived to have the same accent as the Dutch female exposure speaker, the Dutch male

generalization speaker of Experiment 2 allowed us to re-examine the issue of accent similarity in generalization.

Method

Participants.

Twenty-eight participants fulfilling the same criteria as participants in Experiment 1 participated for pay. Three additional participants' data were collected, but not analyzed; one interrupted the experiment and two failed to respond within 2500 ms on more than 5 % of all trials.

Materials and Procedure.

The materials and procedure of the exposure phase were identical to those of Experiment 1. The test phase was the same with the exception that in place of the female generalization speaker, stimuli from the male generalization speaker were presented. The seven steps of the male speaker's continua used in the pretest (i.e., 10, 30, 40, 50, 60, 70, and 90% of the stimulus containing /f/) were used. Immediately following the experiment participants answered the same questionnaire as in Experiment 1.

Results

Questionnaires.

As in Experiment 1 the majority of listeners reported hearing two different speakers (25 of 28; one did not answer this question, one guessed 3 speakers, and one guessed 6). The average perceived accent strength of the exposure speaker was 3.5 (range 2-6; 1 participant did not respond; 1 = *very strong accent*, 7 = *like a native English speaker*). Interestingly, again all but 1 participant stated that the two speakers did *not* have the same accent, and again informal interviews indicated that participants

considered the male generalization speaker to be a native speaker of English. None of the participants guessed that the accent they heard was Dutch although 5 participants indicated the speaker's native language to be another Germanic language such as German or Swedish. Participants reported to be unfamiliar with Dutch accented English (mean 1.75, range 1-4; 1 = *not familiar at all*, 7 = *very familiar*).

Categorization.

All participants met the criterion of accepting at least half of the critical words with ambiguous fricatives presented during exposure as real words. Across subjects, 166 trials (1.2%) were excluded for not meeting the reaction time criterion of response falling between 200 and 2500 ms after word onset. Table 1 shows the acceptance rates for critical words with ambiguous and unambiguous fricatives.

(insert Figure 3 about here)

Figure 3 shows the proportion /s/-responses by listeners in the /f/-ambiguous condition and the /s/-ambiguous condition for the exposure speaker (Panel A) and the male generalization speaker (Panel B) in Experiment 2. Whereas for the exposure speaker the categorization functions between the two exposure groups are clearly different, they pattern closely together for the male generalization speaker. The figure thus suggests that category retuning is found for the exposure speaker but was not generalized to the male speaker's test continua. This was confirmed by statistical analyses. Statistical analyses showed that, as expected, the effect of Continuum Step was significant such that more /s/-responses were given the more /s/-like the stimulus was (b_{Step} = -0.85, *SE* = 0.06, z = -15.37, $p < .001$). The effect of Exposure Condition confirms that listeners' categorization of the female exposure speaker's fricatives was influenced by the lexically-biasing conditions experienced during exposure ($b_{\text{Condition}}$ = 1.15, *SE* =0.34, z =3.41, $p < .001$). Participants who heard the ambiguous fricatives in /s/-biasing contexts later categorized /s/-/f/ in minimal pair continua more often as /s/ than did the listeners who heard /f/-biasing contexts during exposure. The effect of Speaker ($b_{\text{SpeakerM1}}$ = 3.91, *SE*

=0.31, z = 12.78, p < .001) shows that more /s/-responses were given for the male generalization speaker than for the female exposure speaker for whom listeners showed an overall /f/-bias ($b_{Intercept}$ = -2.53, SE = 0.26, z = -9.87, p < .001). In contrast to Experiment 1, however, here the interaction between Exposure Condition and Speaker was significant ($b_{Condition*Speaker}$ = -0.92, SE = 0.36, z = -2.58, p < .01). That is, the effect of category retuning differed for the female exposure speaker and male generalization speaker.

Follow-up analyses showed that an effect of Exposure Condition (i.e., whether during exposure listeners heard the ambiguous sound in lexical contexts biasing the interpretation towards /s/ vs. /f/) was found for the female exposure speaker ($b_{Intercept}$ = -2.07, SE = 0.21, z = -9.84, p < .001, b_{Step} = -0.60, SE = 0.05, z = -11.95, p < .001, $b_{Condition}$ = 0.94, SE =0.28, z =3.37, p < .001) but not for the male generalization speaker ($b_{Intercept}$ = 1.9, SE = 0.23, z = 8.19, p < .001, b_{Step} = -1.33, SE = 0.09, z = -14.83, p < .001, $b_{Condition}$ = 0.31, SE =0.27, z =1.13, p = .26). Note, however, that the male speaker's continuum steps 1-4 were identified as /s/ with almost ceiling performance by listeners in both exposure conditions. This at-ceiling performance over the larger part of the continuum could have minimized generalization of lexically-guided phonetic retuning because such effects are typically most evident for perceptually ambiguous mid-continuum sounds (compare categorization functions, e.g., in Kraljic and Samuel 2005 where the acoustically unambiguous continuum endpoints mostly do not differ between conditions). To test whether an effect of generalization could be found for the perceptually ambiguous part of the male speaker's continuum, analyses were run on steps 5-9 from the male generalization speaker. However, even for this subset of continuum steps, categorization was not influenced by the Exposure Condition ($b_{Intercept}$ = 0.63, SE = 0.22, z = 2.91, p < .005, b_{Step} = -1.44, SE =0.09, z =-16.7, p < .001, $b_{Condition}$ = 0.36, SE = 0.3, z = 1.2, p = .231). There was no evidence of generalization to the male speaker.

Discussion

In Experiment 2 we further tested lexically-guided phonetic category retuning in the context of a global accent and investigated the conditions under which category retuning generalizes across talkers. Replicating Experiment 1, we observed an effect of category retuning for the Dutch female exposure speaker indicating that phoneme-level category retuning occurs even within a global foreign accent that possesses many deviations from the native English norm. Contrary to the generalization across Dutch female speakers we observed in Experiment 1, listeners in Experiment 2 did not generalize from the female exposure speaker to a Dutch male talker. Curiously, despite the shared native language of the exposure and generalization speaker and the possibility for a direct comparison of the speaker's pronunciation of the minimal pairs, again listeners did not perceive the speakers as sharing the same accent, making it impossible to test whether accent similarity may play a role in the generalization of retuned phoneme categories.

Previous research has suggested that speakers or speakers' productions of the critical segments must be "sufficiently similar" for generalization to be observed (Kraljic & Samuel, 2005, pg. 166), but this construct is not well understood. The present data provide an opportunity to further investigate conditions of cross-speaker generalization of category retuning. A comparison of the categorization functions of the female exposure speaker and the male generalization speaker (see Figures 1 and 3; see also the effect of Speaker) indicates that their fricatives were perceived to span different ranges of the perceptual space between /s/ and /f/. Whereas most of the exposure female's stimuli were categorized as /f/, those of the male generalization talker were reported more often to be /s/. Since both continua were created from the natural endpoints of each of the speakers it can be assumed that the female and the male continua influenced one another. This would also explain the differences in the shapes of the categorization functions for the female exposure speaker between Experiment 1 and 2. Critically, for the exposure speaker the exact shape of the categorization functions appears not to have mattered and effects of exposure were found in both experiments.

However, by definition the exposure speaker has been heard speaking during exposure. In contrast, for the generalization speaker listeners have little information about his pronunciation habits, especially early in the post-exposure categorization test. That is, even though overall the male speaker's stimuli were perceived to differ from the female speaker's stimuli, upon first encountering this new speaker, listeners' "best guess" about his pronunciation might be to assume similarity to the exposure speaker. If differential sampling of the perceptual space for the female and male speakers played a role in the lack of generalization observed in Experiment 2, we may expect it to have a reduced effect early in testing when experience with the male voice is minimal. Said another way, it is possible that listeners initially generalized retuning to the male voice but, upon experience with the sampling of his productions along an /f/ to /s/ continuum in perceptual space, they failed to generalize later in testing.

To test this hypothesis we analyzed blocks of the test data collected with the talkers' fricatives intermixed in presentation and with all test stimuli presented once before they were repeated. Figure 4 shows the categorization functions for the male generalization speaker (black) and female exposure speaker (grey), by block. Table 2 reports the statistics. For the male speaker we focused on steps 5-9, steps that spanned most of the perceptual range from /s/ to /f/ but did not contain multiple tokens that were identified as /s/ by listeners in both exposure conditions with close to ceiling performance (i.e., steps 1-4). Indeed, categorization in the first block revealed an effect of exposure indicative of generalization from the female exposure voice. Listeners who had heard the ambiguous female fricative replace /s/ in words during exposure gave more /s/-responses than listeners who had heard the ambiguous sound replacing /f/. By the second repetition of the male test stimuli, generalization was no longer apparent and did not re-appear in any subsequent block (all p's >.16; see Table 2). For the female exposure speaker the influence of exposure was apparent across all blocks (all p's < .05; see Table 2).

(insert Figure 4 about here)

(insert Table 2 about here)

Viewed as snapshots across time, these data suggest an initially non-selective application of category retuning followed by increased speaker-specificity triggered by additional experience with the new talker's fricatives. This leaves us to explain why with increased experience the male speaker's fricatives were perceived as decidedly different from the exposure speaker's fricatives while the fricatives of the female generalization speaker in Experiment 1 were not. One possibility is the fact that multiple tokens along the male speaker's continuum were perceived as good instances of /s/. Kraljic and Samuel (2005) showed that once listeners heard a speaker produce multiple tokens of good instances of the critical sounds then retuning of this category is blocked. Thus once listeners encountered the whole range of the male speaker's continua including multiple tokes of a good /s/, generalization was no longer found. Critically, this possibility leads to the prediction that if the range of the male generalization speaker's fricatives presented at test were equated in ambiguity with the female exposure speaker's continuum (by removing the perceptually unambiguous /s/-like tokens), then generalization should be observed for the male voice throughout the categorization test. Therefore, in Experiment 3 the same male generalization speaker's most /s/-like continuum steps were eliminated. In this way the male talker's sampling of fricatives across perceptual space was more closely aligned with that of the exposure speaker.

EXPERIMENT 3

Experiment 3 explicitly tested the prediction that similarity of sampling of fricatives across perceptual space encourages generalization of lexically-guided perceptual retuning to new talkers. The experiment was identical to Experiment 2, except that the range of the male generalization speaker's test continuum was restricted such that the stimuli identified as /s/ nearly all of the time (stimuli 1-4) were not presented at test. In addition, the female exposure speaker's test continuum was reduced to

equate the overall number of stimuli to the new male continuum while the endpoints were kept the same. The female's fricative continuum thus spanned the same perceptual range as in Experiment 2. Combined, these adjustments better aligned the sampling of perceptual space and overall ambiguity of the continua for the two talkers. If generalization is indeed sensitive to the range of speakers' fricatives, then contrary to the results of Experiment 2 generalization should be observed for the male voice in Experiment 3.

Methods

Participants.

Twenty-eight new participants from the same population as the previous experiments participated for a small payment or for partial course credit. Four additional participants' data were not included in the analysis for failure to meet participation criteria.

Materials.

The stimulus materials were identical to those of Experiment 2 except that the test continua were a subset of the Experiment 2 continua, selected to compensate for an /s/-bias in sampling of perceptual space by the male generalization speaker's test stimuli (see Figures 1 and 3). That is, we defined new endpoints for the male speaker's continua by eliminating multiple tokens of unambiguous stimuli on the /s/-side of the continuum. While the /f/-endpoint stayed the same as before, the new /s/-endpoint was selected such that the continuum still spanned the whole range from /f/ to /s/ but no more than one step of each minimal pair continuum was identified as /s/ more than 95% of the time. For the minimal pair *leaf-lease* stimuli 4-8 of the original 11-step continua were selected, for the three other minimal pairs, stimuli 5-9 were chosen. To simplify description, the selected steps will be referred to as steps 5 through 9. In order to match the number of continuum steps between the new continua

for male generalization speaker and the female exposure speaker (i.e., reduce it to 5), steps 4 and 6 of the exposure speaker's continua were cut. That is, the endpoints of the exposure speaker's continua were identical to Experiment 2 but the number of stimuli was reduced to five to match the new continuum for the male talker (stimuli 1, 3, 5, 7, 9). The goal of these changes was to better align the sampling of perceptual space between /s/-/f/ for the two talkers by better equating the expected response patterns along the continua. This allowed for a test of whether this factor influences generalization of lexically-guided perceptual retuning across talkers.

Procedure.

The instructions, procedure, and analyses of Experiment 3 were identical to the previous experiments. Given that in the previous experiments' participants reported hearing the generalization speakers as native English speakers, we added additional questions about the female and male speakers' accents to the questionnaire.

Results

Questionnaires.

Although both voices were native-Dutch speakers, 26 of 28 participants indicated that the female exposure and male generalization speakers had different accents; 18 of them indicated that the male generalization speaker was a native speaker of English (though four participants specifically mentioned "British" English). The female exposure speaker was again assigned to a variety of possible accents: 8 participants guessed that her native language was a Germanic language naming German, Danish, or Swedish (1 participant did not answer this question). The average accent rating for the female speaker was 3.6 (range: 1-5, one participant did not respond; 1 = *very strong accent*, 7 = *like a native*

English speaker). Overall listeners were unfamiliar with the Dutch accent (average rating 2.0; range 1-6; 1 = *not familiar at all*, 7 = *very familiar*)

Categorization.

(insert Figure 5 about here)

No participants were excluded based on the response-time criterion of failing to respond within 200-2500 ms after word onset for more than 5% of trials. Across participants, 357 trials (3.2%) did not meet this criterion and therefore were excluded from analyses. Figure 5 shows the proportion /s/-responses by listeners from the two exposure conditions for the female exposure speaker (Panel A) and the male generalization speaker (Panel B). As expected, the effect of Continuum Step was significant (b_{Step} = -1.18, *SE* = 0.09, z = -13.69, $p < .001$). More /s/-responses were given the more /s/-like the fricatives were. A significant effect of Exposure Condition indicates that post-exposure categorization of the exposure speaker's continua was influenced by the lexical context in which the ambiguous fricative was presented during exposure ($b_{Condition}$ = 1.05, *SE* = 0.28, z = 3.72, $p < .001$). Listeners who had heard the ambiguous sound on /s/-final words gave more /s/-responses during test than listeners who had heard the ambiguous sound on /f/-final words. Overall listeners again had an /f/-bias for the exposure speaker's continua ($b_{Intercept}$ = -1.92, *SE* = 0.22, z = -8.90, $p < .001$). The effect of Speaker indicates that, as in Experiment 2, listeners gave more /s/-responses for the male generalization speaker than for the female exposure speaker ($b_{SpeakerM1}$ = 1.85, *SE* = 0.26, z = 7.24, $p < .001$). Despite adjustments in sampling the male speaker's continua listeners perceived his fricatives as more /s/-like than the female exposure speaker's fricatives. In contrast to Experiment 2, however, the critical interaction between Exposure Condition and Speaker was not significant ($b_{Condition*SpeakerM1}$ = -0.25, *SE* = 0.36, z = -0.70, $p = .48$). This suggests that in Experiment 3 listeners may have generalized category retuning to the male speaker's fricatives. This was confirmed in additional analyses.

Separate analyses for the female exposure speaker and the male generalization speaker confirm that the effect of Exposure Condition was significant for both speakers. Listeners who had heard the ambiguous sound replace tokens of /s/ with the ambiguous sound during exposure gave more /s/ responses than the group who had heard the ambiguous sound replace tokens of /f/ when presented with the test continua of the female exposure speaker ($b_{\text{Intercept}}$ = -1.84, *SE* = 0.21, z = -8.63, $p < .001$; $b_{\text{Condition}}$ = 0.99, *SE* = 0.27, z = 3.75, $p < .001$; b_{Step} = -1.07, *SE* = 0.1, z = -11.24, $p < .001$) and of the male generalization speaker ($b_{\text{Intercept}}$ = -0.04, *SE* = 0.35, z = -0.11, p =.91; $b_{\text{Condition}}$ = 0.93, *SE* = 0.49, z = 1.89, p = .05; b_{Step} = -1.35, *SE* =0.09, z = -14.18, $p < .001$).

In order to show that the manipulation of Experiment 3 was effective, we compared the male speaker's continuum steps that overlapped between Experiment 2 and 3. This analysis[4] confirmed that the effect of Exposure Condition (and thus the generalization) was significantly stronger in Experiment 3 than in Experiment 2. This is shown by the significant interaction between Exposure Condition and Experiment ($b_{\text{Intercept}}$ = 8.68 , *SE* = 0.48, z = 18.07, $p < .001$; b_{Step} = -1.27, *SE* = 0.07, z = -17.96, $p < .001$; $b_{\text{Condition}}$ = 0.22, *SE* =0.14, z = 1.62, $p = .10$; $b_{\text{Experiment}}$ = -0.02, *SE* = 0.13, z = -0.17, p = .88 ; $b_{\text{Condition*Experiment}}$ = 1.68, *SE* = 0.21, z = 7.85 , $p < .001$).

A similar analysis comparing the effects on overlapping steps between experiments for the female exposure speaker showed that the basic retuning effect was also larger in Experiment 3 ($b_{\text{Intercept}}$ = 0.88, *SE* = 0.15, z = 5.81, $p < .001$; b_{Step} = -0.53, *SE* = 0.04, z = -14.54, $p < .001$; $b_{\text{Condition}}$ = 0.71, *SE* = 0.11, z = 6.47, $p < .001$; $b_{\text{Experiment}}$ = 0.03, *SE* = 0.12, z = 0.28, p = .78; $b_{\text{Condition*Experiment}}$ = 0.41, *SE* = 0.17, z = 2.38, $p < .05$). This suggests that the two speaker's test continua are indeed perceived in relation to each other. However, the retuning effect for the female exposure speaker was shown in all three

[4] The coding was the same as before; Experiment was contrast coded (Experiment 2 = -0.5, Experiment 3 = 0.5). Step was not centered since the overlapping steps of the male speaker's continua were steps 5, 6, 7, and 9, hence not symmetrical. The same coding was used for the female speaker.

Experiments. Additionally, no evidence for differences between the magnitude of the exposure speaker's retuning effect with comparison to Experiment 1 could be found ($b_{Intercept}$ = 1.68, *SE* = 0.17, z = 9.79, $p < .001$; b_{Step} = -0.52, *SE* = 0.03, z = -15.37, $p < .001$; $b_{Condition}$ = 0.49, *SE* = 0.13, z = 3.77, $p < .001$; $b_{Experiment}$ = -1.24, *SE* = 0.12, z = -10.3, $p < .001$; $b_{Condition*Experiment}$ = 0.03, *SE* = 0.18, z = 0.14, $p = .89$). That is, the magnitude of the retuning effect in Experiment 3 does not appear unusually large. It is hence unlikely that the magnitude of the effect is the sole trigger for generalization (because it is sufficiently strong to overcome adversities such as a change in the gender between speakers). Rather we suggest that the similarity in the perception of two speaker's continua determines whether cross-speaker generalization of lexically-guided category retuning can be found.

Discussion

Previous work has suggested that speakers must be "sufficiently similar" for generalization to be observed in lexically-guided perceptual retuning of phonetic categorization (Kraljic & Samuel, 2005, pg. 166). Experiment 3 explicitly tested how sampling across perceptual space influences generalization by presenting a subset of the male generalization speaker's stimuli presented in Experiment 2 to better align sampling of perceptual space with the female exposure speaker's fricatives. By presenting only a subset of the male speaker's continua we eliminated a number of tokens that were perceived as good instances of /s/. These good instances of /s/ could have blocked generalization either because the perception of good instances of a category blocks retuning in general (Kraljic & Samuel, 2005) or because the presence of good instances of /s/ simply did not allow for room for shifting perception further towards /s/ over a large part of the continuum. While the present study cannot decide between these two alternative explanations (or a potential combination thereof) what we could show is that no generalization was observed for the male talker in Experiment 2, but in Experiment 3 we found robust

generalization to exactly this speaker's voice. Thus, the range of responses to the test stimuli presented has an important influence on whether generalization is observed, independent of voice identity.

GENERAL DISCUSSION

The present study demonstrated two main findings providing insight into the mechanisms of lexically-guided category retuning. We first demonstrated that listeners show phonetic category retuning of an artificially manipulated segment even when the ambiguous segment was embedded in an unfamiliar foreign accent during exposure. This provides the first evidence that laboratory studies of the adaptation to single artificially manipulated sounds are an acceptable model for studying adaptation to naturally occurring accents. Second, we showed that the retuning of fricatives is not speaker specific, but generalization depends on how two speakers' test continua are sampled across perceptual space.

The first finding links two lines of research that have been mostly independent: lexically-guided phonetic retuning to single artificially-manipulated phoneme contrasts in native speech (starting with Norris et al., 2003) which is thought to mimic a perceptual challenge introduced by non-native accents, and adaptation to natural foreign-accented speech where more than one phoneme contrast differs from native norms (e.g., Bradlow & Bent, 2008). These mostly independent literatures have referenced one another and suggested parallels in listeners' use of lexical information to tune perception to speakers' pronunciation peculiarities to later better understand new utterances. However, different methods (phoneme categorization vs. word transcription) and the focus on different levels of processing (segment vs. word) have not allowed for direct comparisons. By presenting an artificially manipulated phoneme contrast in an unfamiliar global foreign accent we have demonstrated across three experiments that lexically-guided phonetic category retuning is observable in the context of a global foreign accent and, thus, could play a role in adaptation to naturally-occurring foreign accents. The next step in advancing such an account of adaptation to foreign-accented speech including a segment-by-

segment retuning of phonetic categories would be to demonstrate that similar effects of category tuning can be found when listeners are tested on more than a single phonetic contrast. After all, most commonly in naturally-occurring foreign accents the pronunciation of multiple segments deviates from the native norm.

Support for this suggestion comes from the literature on perception of foreign-accented speech. Sidaras et al. (2009) showed that listeners who had been exposed to a foreign accent showed specific improvement for the recognition of words that contained vowels that were highly confusable in this accent. They argue that exposure allowed listeners to tune into the specific non-native cues used by the learners to produce the vowel contrast. Therefore listeners in the exposure condition were better at understanding accented words than listeners who did not receive exposure to the accent and hence continued using native cues to interpret the non-native vowels. The present study demonstrated the retuning of a category contrast within a global foreign accent in greater detail. We used the classical paradigm to study phonetic category retuning (Norris et al. 2003) to which adaptation to foreign accents has frequently been compared (e.g., Bradlow & Bent, 2008; Sidaras et al. 2009) and demonstrated that listeners show shifts in phonetic categorization of a specific category contrast even when multiple characteristics of the speaker's speech deviate from the native norm. Hence evidence accumulates that the mechanisms found for lexically-guided adaptation to an artificially manipulated segment could be *part* of what happens during adaptation to a global foreign accent.

The present experiments further specified the circumstances under which listeners apply lexically-guided phonetic retuning evoked by one speaker's pronunciation variant to a new, previously unheard, speaker. One of our hypotheses -- that the presence of a common foreign accent facilitates generalization -- could not be addressed by the present studies because listeners failed to hear the

native Dutch talkers as speakers of the same native language. This was likely influenced by our native-English listeners' low familiarity with Dutch-accented English.

Nonetheless, the studies provided an opportunity to explicate constraints upon generalization. Previous studies (e.g., Eisner & McQueen, 2005; Kraljic & Samuel, 2005, 2007) have suggested that fricatives, which were the critical categories here, are treated by listeners in a speaker-specific fashion unless the two speakers' fricatives are "sufficiently similar" to one another for generalization to occur (Kraljic & Samuel, 2005). Here we set out to specify this notion of similarity by systematically manipulating the generalization-speaker's perceived match to the exposure speaker. Cross-speaker generalization was found in Experiment 1 where both speakers were female and their voices as well as fricatives appeared to be perceptually similar. Note also that there was some confusion between the identity of the speakers at Pretest. Their categorization functions were similar. The male generalization speaker's voice in Experiments 2 and 3 was perceived as more distinctive. Critically, generalization to the male speaker depended on sampling of his fricatives to match or mismatch the perceptual space of the female exposure speaker's fricatives. We showed that overall generalization could be "switched on" and "off" for the same male generalization speaker by sampling his fricatives from different ranges of perceptual space.

An issue that remains to be resolved in this regard is the issue of *perceptual* vs. *acoustic* similarity with respect to predicting speaker generalization. In the present studies, we have relied on listeners' perceptual judgments of the stimuli as a measure of *perceptual* similarity (see Pretest). Although acoustic similarity is likely to be correlated with this measure, the manner by which we created our stimulus continua (acoustic mixing via STRAIGHT; Kawahara, et al. 1999) precludes a straightforward link to acoustics. Given that the male speaker's new /s/-endpoint in Experiment 3 was a 50 % mix of his natural /s/ and /f/ productions, acoustic measurements would have to be interpreted with caution.

Nonetheless, the present results are definitive in that the precise sampling of speech stimuli presented at test has a major influence on whether generalization is observed (Experiment 3) or not (Experiment 2), *even for the same voice*. In this way, what has been investigated as speaker generalization may be less about *speaker* or voice, per se, than about the similarity of the generalization stimuli to those experienced during exposure. This observation is in line with Kraljic and Samuel (2005) and helps to interpret the diversity of speaker generalization findings for lexically-guided phonetic retuning: Eisner and McQueen (2005) found generalization only when they used the exactly same fricative tokens spliced onto the different speakers' voices. Kraljic and Samuel (2005) explained asymmetric generalization by using acoustic measurements, and finally we relied on listeners' perception of the fricative continua to demonstrate how generalization can be "switched on" and "off".

Samuel and Kraljic (2009b, submitted) hypothesize that listeners build person-specific representations of pronunciation variants. These representations are stored as episodes and include information about the phonetic category, the speaker, and the circumstances in which the information was encountered (e.g., whether there was some potential external explanation for the presence of ambiguous sounds, such as the speaker having a pen in her mouth while articulating the words; see Kraljic, et al., 2006; Samuel & Kraljic, 2011). If episodes match across encounters with critical tokens in the experiment then retuning is observed; if not, retuning is not observed. This account can explain findings (first reported in Kraljic & Samuel, 2005) that if episodes for a speaker are first associated with unambiguous fricatives under "normal" listening conditions, then for this speaker later retuning may fail to be observed even if the next episode involves an ambiguous sound. This is because the listener already has a representation of this speaker's fricatives from the previous encounter that now (temporarily) "blocks" the retuning (e.g., Kralic & Samuel, 2005, 2011). In contrast, category retuning is observed if either the speaker has been heard producing only ambiguous sounds (i.e., what is shown, for example, for our exposure speaker), or if, following the exposure to unambiguous fricatives, either the

speaker or the situation has changed. For example, if the speaker happens to have a pen in her mouth while producing unambiguous tokens and later does not have a pen in her mouth while producing ambiguous tokens, then retuning is found (Kraljic & Samuel, 2011). The same holds for a change of visual speaker-identity. If listeners see and hear one speaker produce unambiguous fricatives and then hear the same voice produce ambiguous fricatives but paired with the video of a different speaker then retuning is found for the "second" speaker (Samuel & Kraljic, submitted). If no speaker change is perceived either because listeners see the video of the same speaker (Samuel & Kraljic, submitted) or they don't see a video at all and infer from the voice that the speaker has not changed (Kraljic & Samuel, 2005) then retuning is not observed. This confirms that listeners' *perception* of the situation plays a role for retuning and not the acoustics of the voice alone.

Note that the account of matching episodes between encounters is based on an experimental paradigm in which (failure of) retuning after hearing unambiguous sounds is tested. In our case listeners heard one speaker produce ambiguous sounds during exposure and the same speaker plus a new speaker at test. In terms of an account of episodic speaker information this would mean that listeners built up a speaker model for the exposure speaker that includes information leading to retuning for this speaker. For the second speaker, a perceptual space and speaker-specific episodes must be established at test. If the second speaker's perceptual space is sufficiently similar to the exposure speaker's space then the retuning is generalized. A question that remains to be answered in future research is the scope of episodic information that is stored for each speaker. The language heard during exposure or presence of an accent, for example, do not appear to be part of the episodes (Reinisch et al. 2013). If language or accent were part of the episodes, Reinisch et al. (2013) should not have found category retuning for a speaker speaking Dutch-accented English during exposure and native Dutch at test. Additionally, it remains to be shown whether the episodes refer to the speaker as an abstract entity, the voice (which is unlikely given Eisner & McQueen, 2005; Samuel & Kraljic, submitted) or the critically-affected phonetic

categories (fricatives in this case). Nevertheless, taken together, a picture emerges that attention to the details of both the perceptual/acoustic space sampled by the exposure speaker as well as that of the test speaker will be essential in determining the cases under which to expect generalization.

Studies of adaptation to naturally occurring foreign accents (e.g., Bradlow & Bent, 2008; Sidaras, et al. 2009) have found that when listeners are trained on only one speaker they robustly adapt to this speaker's accent but transfer of adaptation to new speakers - even of the same native language - is limited. Speaker-independent adaptation to a foreign accent is achieved best if listeners are trained on multiple speakers of the same accent (Bradlow & Bent, 2008; Sidaras, et al. 2009) or alternatively on a variety of different accents (Baese-Berk, et al. 2013). Authors of these studies cite inter-speaker variability as preventing generalization from one to another speaker and argue that exposure to multiple speakers allows listeners to abstract across this variability to distill the most relevant features of the accent, thus aiding generalization to new speakers of the same accent. To restate this in terms of our hypothesis about sampling perceptual space, including multiple talkers at exposure samples a wider acoustic/perceptual space. In light of the present results, this may be beneficial in promoting generalization of lexically-guided phonetic retuning to natural foreign accents.

Finally, the timecourse analyses of listeners' failure to generalize to the male voice in Experiment 2 lend some insight into possible mechanisms in cross-speaker generalization. Even though in the overall analysis no generalization to the male speaker’s continua could be found, splitting the data into blocks revealed that listeners first applied what they had learned during exposure regardless of the speaker. Similar findings on cross-speaker usage of acoustic context information such as speaking rate and spectral context have been used to argue for speaker-independent, pre-categorical application of these processes in speech perception (speaking rate: Green, Stevens, & Kuhl, 1994; Green, Tomiak, & Kuhl, 1997; Newman & Sawusch, 2009; Sawusch & Newman, 2000; spectral context: Lotto & Kluender,

1998; Watkins, 1991). By analogy, the initial generalization of retuned categories can inform us about their early application during speech processing. That is, the learning appears to directly affect the categorization of fricatives (see also Mitterer & Reinisch, 2012, in press). Critically in our study, if over time during the test, listeners accumulated evidence that the sampling of perceptual space differed between speakers they refrained from interpreting the new speakers' fricatives relative to what they had learned about the speaker heard during exposure. This suggests that the perceptual system flexibly tracks a speaker's pronunciation characteristics. Studies in which two speakers are presented during exposure confirm that listeners are able to track two speakers' pronunciation variants independently (e.g., Kraljic & Samuel, 2007; Trude & Brown-Schmidt, 2012). Here we addressed generalization to a speaker who had not been encountered before but showed that listeners start tracking the second speaker as soon as they can.

In summary, the present findings bring together related lines of research on the perception and adaptation to nonnative pronunciation variants. Listeners track artificially manipulated segments within a global foreign accent. Moreover, listeners flexibly track perceptual space spanned by different speakers. We found that sampling of perceptual space across speakers, not speaker identity, predicted generalization of lexically-guided phonetic retuning.

Acknowledgments

We would like to thank Christi Gomez and the research assistants of the Holt lab for help with running the experiments, and Matthias Sjerps, Frank Eisner, and Holger Mitterer for helpful discussions of the results. Thanks also go to Arthur Samuel, Dennis Norris, and two anonymous reviewers for helpful comments on a previous version of the manuscript. This research was supported by NIH (R01 DC004674) and NSF (0746067). Parts of the results were presented at the 164th Meeting of the Acoustical Society of America, Kansas City, Missouri, October 2012.

References

Baayen, H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

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

Baese-Berk, M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *Journal of the Acoustical Society of America*, online first publication (4 February, 2013).

Barr D. J., Levy R., Scheepers C. & Tily, H. (2013) Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255-278.

Bates, D. M., & Sarkar, D. (2007). *lme4: Linear mixed-effects models using S4 classes* (version 0.999375-27) [software application].

Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk after effect. *Psychological Science, 14,* 592-597.

Boersma, P., & Weenink, D. (2009). *PRAAT: doing phonetics by computer* (version 5.1) [software application] retrieved from http://www.praat.org *(date last viewed 06/06/2011)*

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

Brouwer, S., Mitterer, H., & Huettig, F. (2012). Speech reductions change the dynamics of competition during spoken word recognition. *Language and Cognitive Processes, 27*), 539-571. doi:10.1080/01690965.2011.555268.

Clopper, C. G., & Pisoni, D. B. (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech, 47*, 207-239.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67*, 224-238.

Eisner, F., & McQueen, J.M. (2006). Perceptual learning in speech: Stability over time. *Journal of the Acoustical Society of America, 119,* 1950-1953.

Green, K. P., Stevens, E. B., & Kuhl, P. K. (1994). Talker continuity and the use of rate information during phonetic perception. *Perception & Psychophysics, 55*, 249-260.

Green, K. P., Tomiak, G. R., & Kuhl, P. K. (1997). The encoding of rate and talker information during phonetic perception. *Perception & Psychophysics, 59*, 675-692.

Idemaru, K., & Holt, L. L. (2011). Word Recognition Reflects Dimension-based Statistical Learning. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 1939-1956.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59,* 434-446.

Kawahara, H., Masuda-Katsuse, I., & de Cheveigné, A (1999) Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction. *Speech Communication, 27*, 187-207.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology, 51*, 141-178.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review, 13*, 262-268.

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

Kraljic, T., & Samuel, A. G. (2011). Perceptual learning evidence for contextually-specific representations. *Cognition, 121*, 459-465.

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science, 19,* 332-338.

Liss, J. M., Spitzer, S. M., Caviness, J. N., & Adler, C. (2002). The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *Journal of the Acoustical Society of America, 112*, 3022-3030.

Lotto, A. J. & Kluender, K. R. (1998). General contrast effects of speech perception: Effect of preceding liquid on stop consonant identification. Perception & Psychophysics, 60, 602-619.

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

McQueen, J. M., Norris, D., & Cutler, A. (2006). The dynamic nature of speech perception. *Language & Speech*, *49*, 101-112.

McQueen, J. M., & Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *Journal of the Acoustical Society of America, 131*, 509-517.

Mitterer, H., & McQueen, J.M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS One, 4*, A146-A150.

Mitterer, H., & Reinisch, E. (2012). Category Retuning Affects Early Stages of Speech Processing. *Abstracts of the Psychonomic Society, 17*, 208.

Mitterer, H., & Reinisch, E. (in press). No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language*.

Munson, B. (2011). The influence of actual and imputed talker gender on fricative perception, revisited. *Journal of the Acoustical Society of America, 130*, 2631–2634.

Newman, R. S., & Sawusch, J. R. (2009). Perceptual normalization for speaking rate III: Effects of the rate of one voice on perception of another. *Journal of Phonetics, 37*, 46-65.

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

Poellmann, K., McQueen, J. M., & Mitterer, H. (2011). The time course of perceptual learning. In W.-S. Lee, & E. Zee (Eds.). *Proceedings of the 17th International Congress of Phonetic Sciences 2011 [ICPhS XVII]* (pp. 1618-1621). Hong Kong: Department of Chinese, Translation and Linguistics, City University of Hong Kong.

Quené, H., & Van Den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*, 413-425.

Reinisch, E., Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 75-86.

Samuel, A. G., & Kraljic, T. (2009a). Perceptual learning for speech. *Attention, Perception, & Psychophysics, 71*, 1207-1218.

Samuel, A. G., & Kraljic, T. (2009b). Identical speech acoustics, different perceptual learning: Faces matter. *Abstracts of the Psychonomic Society*, *14*, 12.

Samuel, A. G., & Kraljic, T. (submitted). Visually-specified speaker identity can dominate processing of spoken words.

Sawusch, J. R., & Newman, R. S. (2000). Perceptual normalization for speaking rate II: Effects of signal discontinuities. *Perception & Psychophysics, 62*, 285-300.

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

Sjerps, M. J., & McQueen, J. M. (2010). The bounds of flexibility in Speech Perception. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 195-211.

Strand, E., and Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. in Gibbon, D. (ed.). *Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference Bielefeld*. Berlin, Germany: Mouton de Gruyter (pp. 14–26).

Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during on-line speech perception. *Language and Cognitive Processes*, *27*, 979-1001.

Watkins, A. J. (1991). Central, auditory mechanisms of perceptual compensation for spectral envelope distortion. *Journal of the Acoustical Society of America, 90*, 2942–2955.

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

Tables

Table 1. Mean percentage of correct responses for ambiguous and natural /f/-final and /s/-final words during the auditory lexical-decision exposure phase in Experiments 1, 2, and 3.

	Ambiguous fricative		Natural fricative	
	/f/	/s/	/f/	/s/
Experiment 1	75 %	95 %	80 %	96 %
Experiment 2	80 %	94 %	84 %	97 %
Experiment 3	78 %	95 %	83 %	94 %

Table 2. Analyses per block for the female exposure speaker and the male generalization speaker in Experiment 2.

		female exposure speaker			male generalization speaker		
	beta *p-value*	intercept	Exposure Condition	Continuum Step	intercept	Exposure Condition	Continuum Step
block 1	*b*	-2.51	1.64	-0.68	0.44	1.17	-1.46
	p	<.001	<.001	<.001	.15	<.01	<.001
block 2	*b*	-2.37	0.99	-0.69	0.86	0.42	-1.56
	p	<.001	<.005	<.001	<.01	.31	<.001
block 3	*b*	-1.77	0.77	-0.46	.089	-0.08	-1.49
	p	<.001	<.05	<.001	<.005	.84	<.001
block 4	*b*	-2.18	0.84	-0.69	0.63	-0.14	-1.26
	p	<.001	<.01	<.001	<.05	.67	<.001
block 5	*b*	-0.53	0.66	-0.53	0.39	0.52	-1.74
	p	<.001	<.05	<.001	.13	.16	<.001

Figure captions

Figure 1. Categorization curves of the minimal pair words for the test phase. Proportion of /s/-responses is plotted for each tested step of the continuum for each of the three Dutch speakers. The solid line shows responses to the continuum of the female exposure speaker, the dashed lines show responses to the continua of the generalization speakers. The dashed line with the point character 'F' shows responses to the female generalization speaker (labeled "2nd female speaker" in the legend), and the dashed lines with the point character 'M' show responses to the male generalization speaker.

Figure 2. Proportion /s/-responses along the morphed /s/-/f/ continua for listeners in the /s/-ambiguous exposure condition (dashed lines) and the /f/-ambiguous exposure condition (solid lines). Panel A shows responses for the exposure speaker; panel B for the female generalization speaker.

Figure 3. Proportion /s/-responses along the morphed /s/-/f/ continua for Experiment 2 listeners in the /s/-ambiguous condition (dashed lines) and the /f/-ambiguous condition (solid lines). Panel A shows responses for the female exposure speaker, panel B for the male generalization speaker.

Figure 4. Learning effect by block for the male generalization speaker (black) and the female exposure speaker (grey) for comparison. Dotted lines represent the s-ambiguous condition, solid lines the f-ambiguous condition.

Figure 5. Proportion /s/-responses along the morphed /s/-/f/ continua for listeners in the /s/-ambiguous condition (dashed lines) and the /f/-ambiguous condition (solid lines) in Experiment 3. Panel A shows responses for the female exposure speaker, panel B for the male generalization speaker.

Figures

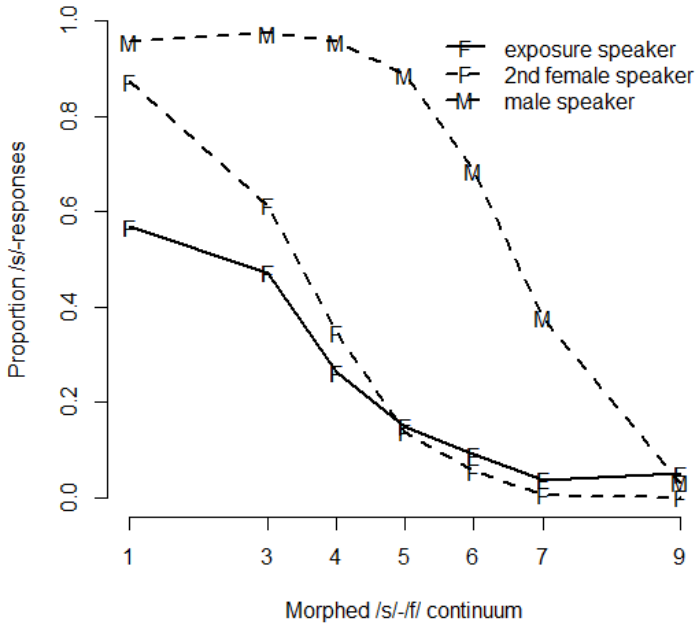

Figure 1

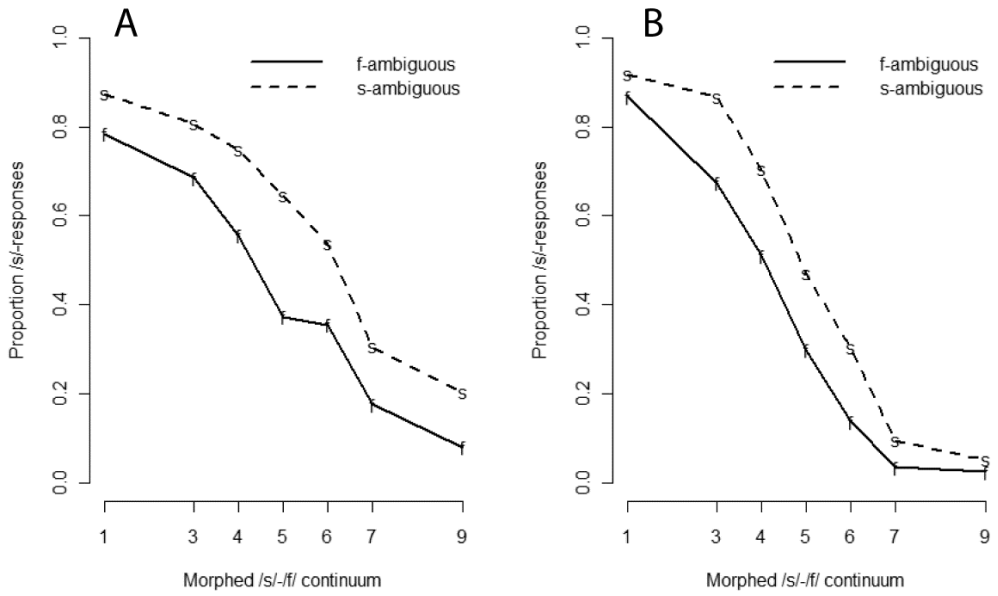
A
B
f-ambiguous
s-ambiguous
Proportion /s/-responses
Morphed /s/-/f/ continuum
0.0
0.2
0.4
0.6
0.8
1.0
1
3
4
5
6
7
9

Figure 2

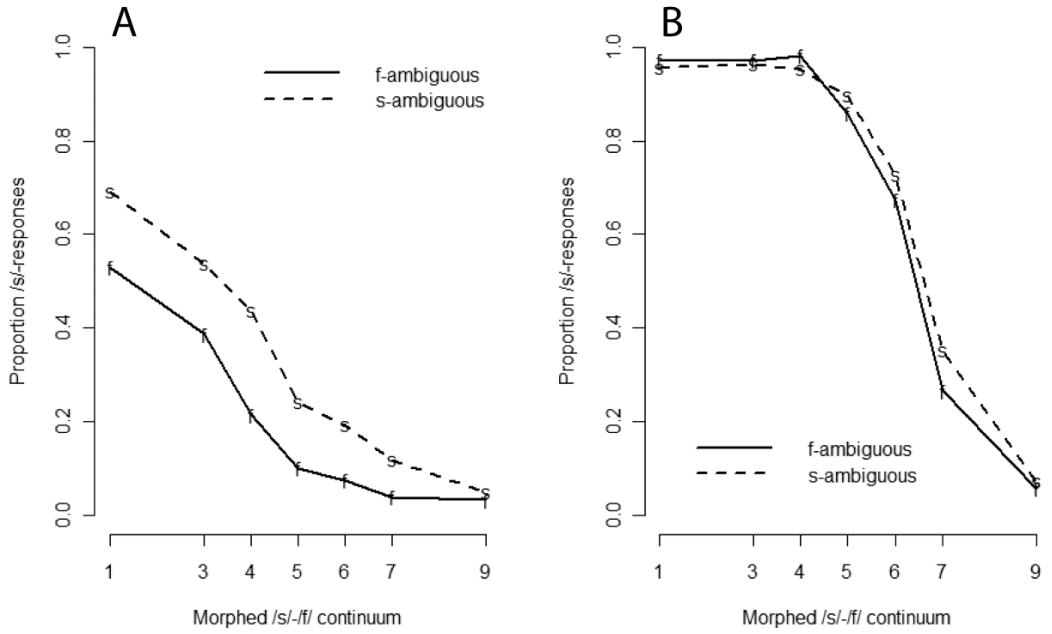
A
B
f-ambiguous
s-ambiguous
Proportion /s/-responses
Morphed /s/-/f/ continuum
0.0
0.2
0.4
0.6
0.8
1.0
1
3
4
5
6
7
9

Figure 3

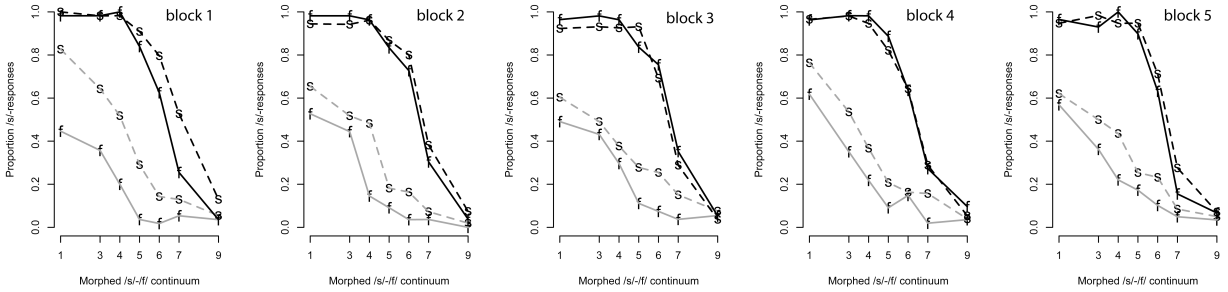
block 1
block 2
block 3
block 4
block 5
Proportion /s/-responses
Morphed /s/-/f/ continuum

Figure 4

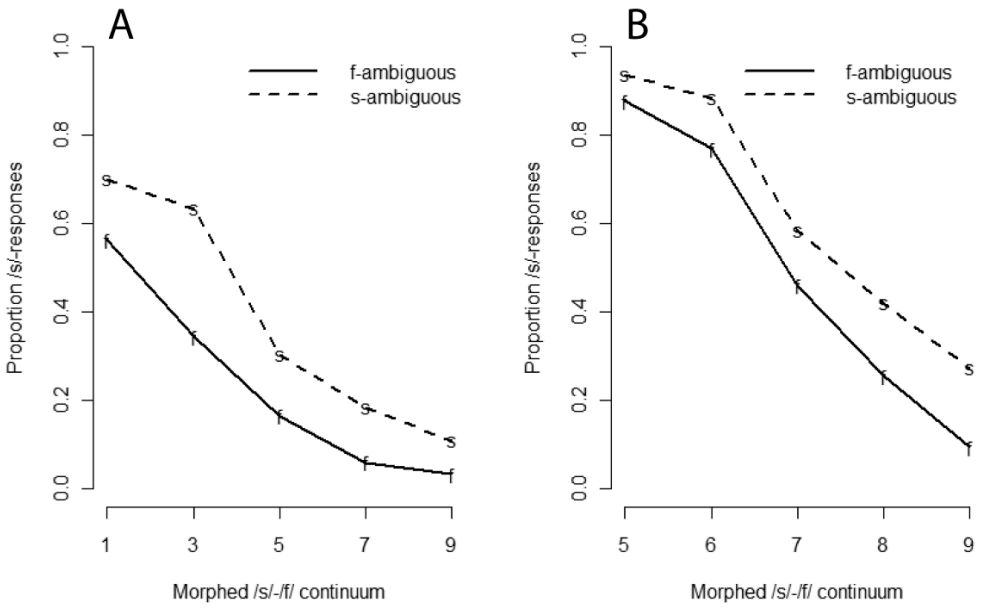
A
f-ambiguous
s-ambiguous
Proportion /s/-responses
0.0
0.2
0.4
0.6
0.8
1.0
1
3
5
7
9
Morphed /s/-/f/ continuum
B
f-ambiguous
s-ambiguous
Proportion /s/-responses
0.0
0.2
0.4
0.6
0.8
1.0
5
6
7
8
9
Morphed /s/-/f/ continuum

Figure 5