Journal of Experimental Psychology:
Human Perception and Performance
2013, Vol. 39, No. 1, 75–86

© 2012 American Psychological Association
0096-1523/13/$12.00 DOI: 10.1037/a0027979

# Listeners Retune Phoneme Categories Across Languages

Eva Reinisch
Max Planck Institute for Psycholinguistics

Andrea Weber
Max Planck Institute for Psycholinguistics and Donders Institute for Brain Language and Behaviour

Holger Mitterer
Max Planck Institute for Psycholinguistics

Native listeners adapt to noncanonically produced speech by retuning phoneme boundaries by means of lexical knowledge. We asked whether a second language lexicon can also guide category retuning and whether perceptual learning transfers from a second language (L2) to the native language (L1). During a Dutch lexical-decision task, German and Dutch listeners were exposed to unusual pronunciation variants in which word-final /f/ or /s/ was replaced by an ambiguous sound. At test, listeners categorized Dutch minimal word pairs ending in sounds along an /f/–/s/ continuum. Dutch L1 and German L2 listeners showed boundary shifts of a similar magnitude. Moreover, following exposure to Dutch-accented English, Dutch listeners also showed comparable effects of category retuning when they heard the same speaker speak her native language (Dutch) during the test. The former result suggests that lexical representations in a second language are specific enough to support lexically guided retuning, and the latter implies that production patterns in a second language are deemed a stable speaker characteristic likely to transfer to the native language; thus retuning of phoneme categories applies across languages.

*Keywords:* speech perception, perceptual learning, foreign accent, second language listening

Because everyone speaks a little differently, listeners have to flexibly adapt to a wide range of pronunciation variants. It is important to note that this adaptation occurs quickly, and listeners are able to generalize from only a few instances of a specific noncanonical pronunciation variant to new words spoken by the same speaker (see Samuel & Kraljic, 2009, for an overview). Such quick learning can occur with foreign-accented speech (e.g., Bradlow & Bent, 2008; Clarke & Garrett, 2004), as well as for phonemes that are "oddly" pronounced in native speech (e.g., Norris, McQueen, & Cutler, 2003). In particular, studies on unusual pronunciations in native speech have led to suggestions about the underlying mechanisms for adaptation: Listeners flexibly incorporate ambiguous sounds into existing phoneme categories if there is contextual evidence about the phonemic identity of this sound. The context most extensively explored is lexical context, even though, for example, visual speech (e.g., a lip closure) can also guide adaptation (e.g., Bertelson, Vroomen, & de Gelder, 2003). In the present study, we asked what role the context language plays during training in lexically guided category retuning. That is, we present the training context in the listeners' native versus nonnative language, and thereby, we explore consequences for the postulated mechanisms of lexically guided category retuning.

In the original study by Norris et al. (2003), an exposure-test paradigm was used in which exposure condition was varied between participants. One group of Dutch listeners heard an ambiguous sound halfway between /f/ and /s/ in a lexical context where it could only be interpreted as /f/ (e.g., *olijf,* "olive"), whereas the other group of listeners heard the same ambiguous sound in a lexical context biasing toward the interpretation of /s/ (e.g., *radijs,* "raddish"). Exposure consisted of a lexical-decision task with only 20 critical items. Immediately thereafter, all participants completed the same test phase. In the test phase, participants were asked to categorize stimuli from an [εf]–[εs] continuum as containing either /s/ or /f/. The listeners who had heard the ambiguous sound as a replacement for /f/ categorized more sounds along this continuum as /f/ than those who had heard the ambiguous sound as a replacement for /s/. That is, listeners learned to incorporate the ambiguous sounds into the category that reflected the lexical interpretation heard during exposure.

Lexically guided category retuning is a consistent finding in research on native language listening. It has not only been found with a lexical-decision task during exposure, but also when parti-

This article was published Online First April 30, 2012.

Eva Reinisch and Holger Mitterer, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; Andrea Weber, Max Planck Institute for Psycholinguistics and Donders Institute for Brain Language and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands.

Eva Reinisch is now at Department of Psychology, Carnegie Mellon University. We thank the student assistants of the Adaptive Listening Group at the Max Planck Institute for help with running the experiments, and Lies Cuijpers for suggesting the testing of German learners of Dutch for Experiment 2. We also thank J. M. McQueen and F. Eisner for providing their datasets (published in Norris et al., 2003, and Eisner & McQueen, 2006) for the comparisons of variance with our dataset in Experiment 1. We are grateful to three anonymous reviewers for their suggestions, which helped us to significantly improve the quality of the paper. Results of this paper were presented at the 162nd meeting of the Acoustical Society of America, San Diego, California, in November 2011.

Correspondence concerning this article should be addressed to Eva Reinisch, Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. E-mail: evarei@andrew.cmu.edu

cipants count the number of words (McQueen, Norris, & Cutler, 2006) or when they listen to a story (Eisner & McQueen, 2006). During the test, perceptual learning has been shown with categorization tasks, with cross-modal priming (McQueen, Cutler, & Norris, 2006; Sjerps & McQueen, 2010), and with eye tracking (Poellmann, Mitterer, & McQueen, 2011). Furthermore, adapted phoneme categories have been shown to generalize across the lexicon (e.g., McQueen, Cutler, et al., 2006) and to be stable over time (Eisner & McQueen, 2006; Kraljic & Samuel, 2005). However, perceptual learning is also restricted in multiple ways. For example, when the ambiguous sounds are presented in word-initial position, no perceptual learning is found (Jesse & McQueen, 2011). Also, the generalization across larger sets of sounds is restricted. For example, the adaptation of front vowel categories does not affect the perception of back vowels (Maye, Aslin, & Tanenhaus, 2008).

It is important to note that category retuning also depends on the set-up of the exposure phase. Kraljic and Samuel (2005; see also Kraljic, Samuel, & Brennan, 2008) showed that learning of ambiguous fricatives is blocked when listeners have heard the speaker produce good instances of the critical fricatives before. Moreover, learning does not occur when during exposure the speaker has a pen in her mouth while producing ambiguous sounds, (Kraljic et al., 2008). That is, learning does not occur when the circumstances in which the ambiguous sound is heard can be attributed to a transient cause. In a follow-up study, Kraljic and Samuel (2011) further investigated the underlying mechanism of this blocking effect. They found that learning is not blocked when during exposure the speaker first has a pen in her mouth while producing ambiguous sounds, and then produces ambiguous sounds without the pen, or when the pen in the mouth is paired with canonical pronunciations. They conclude from their results that with a transient cause for the unusual pronunciation listeners do not strip incidental ambiguity away and perceive canonical sounds. Rather, the combined event of the pen and the pronunciation maps onto a different representation than the pronunciation itself. Therefore, Kraljic and Samuel (2011) suggest that for perceptual learning to occur, listeners store episodes of circumstantial information alongside the pronunciation variants and use these mappings for the interpretation of new utterances by the speaker.

The locus of episodic storage, however, would have to be at a prelexical level. McQueen, Cutler, et al. (2006) showed that an episodic storage of word forms could not explain all generalization effects of category retuning. Category boundaries that have been shifted due to exposure to one set of words can be applied to the recognition of a new set of words. This generalization is best accounted for by assuming a prelexical locus of adaptation (see also, Mitterer, Chen, & Zhou, 2011). Therefore, the present study further explores the boundaries of flexibility in category retuning and possible implications on the mechanisms postulated to underlie this process. Kraljic and Samuel (2011) base their hypotheses about episodic storage on a single paradigm using the pen in the mouth as the transient circumstance (but note that the effects could be replicated). Here, we use another ecologically valid circumstance: exposure language. Specifically, we test whether retuning is possible in a second language (L2) and, moreover, whether shifted category boundaries transfer from the second to the native language (L1). Note that cross-language transfer should occur only if language is not a part of the assumed episodically stored information.

In order to test cross-language transfer of category retuning, it first has to be shown that category retuning can be guided by the second language lexicon. Listening in a second language is usually more demanding than listening in one's native language. Some nonnative sounds are stubbornly difficult to perceive (for a recent overview, see Bohn & Munro, 2007). This may impede the retuning of categories. However, not all L2 sounds are difficult to perceive. Even though non-native sounds almost always differ from native sounds, some can be perceived as good or acceptable exemplars of a native category. If both members of a non-native contrast are assimilated to different native categories, discrimination is expected to be good (see the two-category assimilation type in the L2-version of the Perceptual Assimilation Model; Best & Tyler, 2007; see also the Speech Learning Model; Flege, 1992, 1995). It is reasonable to expect that in such a case, listeners are able to retune phoneme categories in their L2. Therefore, by testing similar sound contrasts across languages, the present study provides a starting point for the exploration of second language and cross-language category retuning by maximizing the chances of detecting effects if they occur.

There are, however, additional reasons why retuning in a second language may not occur even for categories that are similar across languages. L2 listeners typically have a smaller vocabulary, and, as a consequence of suboptimal phonological processing, the lexical representations in an L2 may also be phonetically only coarsely specified, particularly for less-proficient L2 learners. It has been argued that children with smaller vocabularies have coarser phonological representations because they need to distinguish between a smaller number of words (Metsala & Walley, 1997). A similar argument has been made to explain the phonological problems in dyslexia (Bryant, 1997). L2 learners definitely have smaller vocabularies than native speakers, and they may hence not be "bothered" by the small deviations from the canonical form that the typical ambiguous stimuli in perceptual learning paradigms provide. Their coarsely specified representations may hence not be able to provide a training signal for lexically guided category retuning, as they are equally well activated by canonical and slightly ambiguous input. Somewhat in line with this assumption, Bundgaard-Nielsen, Best, and Tyler (2011) showed that L2 learners with larger vocabularies have more consistent L1–L2 category mappings, and they are better at the discrimination of L2 contrasts than learners with smaller vocabulary sizes. Another reason for difficulties with L2 category retuning is the lack of strong associations between the L2 lexicon and semantic representations, which are another prerequisite for perceptual learning to occur (see Leach & Samuel, 2007). Therefore, the question of whether retuning can occur in a second language is an interesting question to answer. Again, to maximize the chances of detecting an effect, the present study started by exploring late learners who are proficient users of their L2.

If the context of a second language is sufficiently strong to guide category retuning, the next question to ask is whether these retuned phoneme categories also transfer to the native language. The answer has consequences for the postulated mechanisms of category retuning. If identical circumstances of exposure and test are necessary for category retuning (see Kraljic & Samuel, 2011), a shift in phoneme perception should only be observed when expo-

sure and test are in the same language. A language switch between exposure and test may block perceptual adaptation effects. If, in contrast, language is not considered a "circumstance" that is stored during exposure, category retuning should also occur across languages.

Because retuning of fricatives in native speech has been shown to be speaker-specific (Eisner & McQueen, 2005; Kraljic & Samuel, 2007), it is essential that in the present study the same speaker produces both the exposure and test materials. This implies that when adaptation from L2 to L1 is being investigated, the speaker produces the exposure materials in her second language with a foreign accent. Note that a perceptible foreign accent could be another circumstance that blocks category retuning. The accent thus works, if anything, against our search for category retuning across languages. Listeners may consider unusual pronunciations of the critical fricatives as part of the accent and not as a stable speaker characteristic. If the accent is treated as circumstantial, category shifts should not transfer to the native language during the test.

The questions of category retuning in L2 and across languages were addressed in three experiments using a modified version of the paradigm established in Norris et al. (2003). Exposure consisted of a lexical-decision task where, depending on listener group, either word-final /f/ or /s/ was replaced by an ambiguous sound halfway between these fricatives. The other fricative was presented in its natural form. That is, a given listener who heard f-final words with an ambiguous fricative heard s-final words with good examples of /s/. At the test, listeners categorized five Dutch minimal word pairs ending in sounds along /f/–/s/ continua. In Experiments 1 and 2, the exposure items and the test items were in Dutch spoken by the same Dutch native speaker. Experiment 1 was set out to replicate previous findings with Dutch native listeners performing the experiment in their native language. Going beyond mere replication, we tested a new method for creating ambiguous stimuli by morphing the fricatives, plus a number of preceding segments, instead of selectively mixing frication noise as has been done in previous studies (see the Methods section for details). Experiment 2 tested whether category retuning occurs in a second language. To this end, German learners of Dutch performed the same experiment as native Dutch listeners did in Experiment 1. Note that German has a similar fricative system as Dutch, and both languages use an alveolar fricative /s/ and a labio-dental /f/. German learners of Dutch are, therefore, in a good position to show retuning of these categories in a nonnative language. Finally, Experiment 3 addressed the critical question of whether, following exposure to Dutch-accented English, Dutch listeners show effects of category retuning as when the same speaker uses her native language, Dutch. We asked whether the exposure language and/or the presence of a foreign accent on the exposure language are interpreted as a transient cause to which the noncanonical pronunciation of the sounds can be attributed. Using this combination of languages, we were able to test different groups of listeners with different exposure materials but eventually compare responses to the exact same test phase across experiments and conditions. By using phoneme contrasts that exist in all three languages, and by testing relatively proficient learners, we attempted to maximize the chances of finding effects of category retuning.

## Experiment 1

Perceptual learning paradigms by necessity use a between-subjects design: A given participant can only be biased to perceive ambiguous fricatives as either /s/ or /f/. As the learning effect can be quite stable (Eisner & McQueen, 2006; Kraljic & Samuel, 2005), it is not possible to train listeners to perceive the ambiguous sound first as /f/ and later as /s/ without interference. It is, however, not only more difficult to find effects in a between-participants design but, in addition, previous studies have found extensive individual differences in responses during the test phase. Listeners in the f-biased and s-biased groups tended to cover nearly the complete range of possible values on the proportion of /f/-responses they gave; that is, in both groups there were participants who responded nearly always with /f/ or /s/ independent of training. As we set out to find differences in the amount of learning under different conditions (i.e., with native, nonnative, and nonnative-accented exposure languages), this state of affairs would make it difficult to find differences in the amount of learning.

Therefore, we introduced two methodological innovations. First, instead of using one test continuum with nonword endpoints (e.g., [ɛf]-[ɛs]), we used five different continua with word endpoints (e.g., Dutch *doof–doos* "deaf", "box"). Participants who mainly use one response option often indicate after the experiment that they found the stimuli "sounding all very similar." Using multiple test continua prevents this apparent strategy of mentally comparing the stimuli over trials, as the carrier words sound rather different. Second, we made use of a different stimulus-generation method. Previous studies manipulated only the fricative noises and spliced them onto words in which the vowel transitions were uninformative. The vowel transitions either consistently cued one fricative endpoint (e.g., all transitions indicated a following /f/; Norris et al., 2003), or they cued a third fricative, for example /x/ (e.g., Sjerps & McQueen, 2010). Because listeners may use different strategies when confronted with such stimuli (i.e., assigning less or more weight to cues in the frication noise), we aimed at using stimuli in which all relevant phonetic cues were ambiguous. This included cues that were present in preceding phonemes such as the transitions into the fricatives. To this end, we used an audio morphing technique (STRAIGHT; Kawahara, Masuda-Katsuse, & de Cheveigné, 1999; for details see the Methods section below). This allowed us to generate ambiguous phoneme sequences, for instance by morphing between [of] and [os], which could then be reappended to a natural [d], giving rise to a stimulus in which all cues were ambiguous between *doof* and *doos* ("deaf," "box").

### Methods

**Participants.** Thirty native speakers of Dutch (mean age 21, range 18–27) participated for a small payment. They were recruited from the student population in Nijmegen, The Netherlands. Twenty additional native Dutch participants from the same population took part in the pretest that was conducted to determine the most ambiguous sounds.

**Materials.** One hundred Dutch words and 100 nonwords that were phonologically legal in Dutch were used as exposure materials. The set of words consisted of 40 critical items and 60 filler words. Of the 40 critical items, half ended in /f/ (e.g., *locomotief* "locomotive"), and half ended in /s/ (e.g., *geitenkaas* "goat

cheese"). Importantly, these words are nonwords if the fricatives are exchanged. That is, *locomotie*[s] and *geitenkaa*[f] are nonwords in Dutch. Except for the word-final position of the critical items, none of the words or nonwords contained the sounds /f/, /s/, or their voiced counterparts /v/ and /z/ (this additional constraint was imposed because many Dutch speakers use only unvoiced fricatives). The word set was mostly identical to materials used in previous perceptual learning studies in Dutch (Norris et al., 2003; McQueen, Cutler, et al., 2006; Sjerps & McQueen, 2010). Based on informal pretesting, however, a number of words was replaced to ensure that German learners of Dutch would know them. Five Dutch minimal pairs ending in /f/ and /s/ were selected as test items for phonetic categorization (*doof–doos* "deaf", "box"; *les–lef* "lesson", "guts"; *roof–roos* "robbery", "rose"; *half–hals* "half", "neck"; *kuif–kuis*, "tuft of hair", "chaste").

All words and nonwords were recorded by a female Dutch native speaker (age 28) in a soundproof booth. Critical words from the exposure phase were recorded in pairs, once correctly and once with the word-final fricatives exchanged. That is, *locomotief* was also recorded as *locomotie*[s], and *geitenkaas* was also recorded as *geitenkaa*[f]. Exchanging the word-final fricatives always resulted in nonwords (i.e., *locomotie*[s] and *geitenkaa*[f] are nonwords in Dutch). The speaker was asked to produce the correct and incorrect form of the words with a comparable speech rate, speech style, and intonation contour (f0). This ensured that morphing the critical portions of the words and splicing them back onto the beginning of the words sounded natural.

**Creating ambiguous stimuli.** For each f-final and s-final recording of the critical words, as well as the minimal pairs, the fricatives plus one or two preceding phonemes were spliced out and morphed in an 11-step continuum (0%–100% of the f-final recording, in steps of 10%) using the STRAIGHT algorithm (Kawahara et al., 1999) in Matlab (The MathWorks Inc.). This algorithm decomposes the speech signal into three parameters: 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. The time-aligned version that was used here further takes into account time anchors. That is, if a vowel in one utterance was 80 ms long and 100 ms in another, then the resulting resynthesized vowel would be 90 ms long. The parameters used for the vowel midpoint would come from the fourth 10-ms time window of the shorter sound and the fifth time window from the longer sound.

The morphed portion of the critical words was adjusted in length such that the resulting word sounded maximally natural when it was spliced back onto the beginning of the word. For example, a larger portion of the word was morphed when the critical sounds were preceded by a nasal or the vowel was preceded by a liquid. That is, whereas for the minimal pair *doof–doos* the [os]-[of] part was morphed, for the minimal pair *roof–roos* the complete words were morphed because the /r/ was difficult to separate from the vowel. Approximate phoneme boundaries were used as temporal anchors for the morphing algorithm. In this way, different types of phonemes (i.e., fricatives vs. preceding sounds) were time-aligned, and only segments of the same type were morphed (i.e., vocalic portions of the signal with other vocalic portions and fricative noise with fricative noise). Morphed syllables were then spliced back onto the word stems. The word stems were selected from the word or nonword recordings depending on the naturalness of the resulting tokens. All splicing was done at positive-going zero-crossings using Praat (Boersma & Weenink, 2009).

## Pretest

To find the most ambiguous tokens of the word-final fricative for each word, a phonetic categorization experiment was administered. For each morphed word–nonword continuum (e.g., *locomotief–locomotie*[s] and *geitenkaaa*[f]–*geitenkaas*), and for the continua of the five minimal pairs (e.g., *doof - doos*), phonetic categorization responses were collected. Participants were presented seven-step continua of all words with morphs containing 10%, 30%, 40%, 50%, 60%, 70%, and 90% of the original f-final stimuli. Participants were seated in a sound-proof 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/ irrespective of whether the result would be a real word. Each participant was presented with each step of each continuum once, resulting in 315 trials per participant (40 exposure words plus five minimal pairs × seven steps). The experiment was implemented running Presentation software (Version 14.9, Neurobehavioural Systems Inc.), and it and took approximately 25 min to complete.

For analyses, the proportion of f-responses for each step of each continuum was calculated. Because we presented word and nonword endpoints of the continua, all pretest results were expected to include a bias toward the respective word endpoints (Ganong, 1980). Therefore, to compensate for this effect, 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. This compensation was introduced to keep the present stimuli comparable to previous studies where ambiguous tokens of the fricatives were determined in pretests with nonword syllables, hence not including a Ganong effect. Despite this compensation for lexical responses, however, the mean of the chosen steps of the continua was matched between f-final and s-final words. That is, the mean value for number of selected step (1–7; 7 = /f/) was 4.35 for f-final words and for s-final words. Moreover, the mean percentage of f-responses for the selected steps of f-final and s-final words was approximately equally distant from 50% (28% for s-words and 24% for f-words). For each of the five minimal pairs to be used in the test phase, the two steps that received just above or below 50% f-responses were selected for phonetic categorization during the test. The test continua thus contained four steps sampled from the ambiguous parts of the continua.

## Procedure

**Exposure.** Half of the participants were randomly assigned to what will be referred to as the f-ambiguous condition, and the other half was assigned the s-ambiguous condition. All participants heard the same 60 filler words and 100 nonwords. Participants in the f-ambiguous condition were further presented with the 20 s-final words in which the /s/ was unambiguous (i.e., the s-endpoint of the morphed continua). The 20 f-final words were presented with the ambiguous steps of the continua as established in the pretest (hence "f-ambiguous"). Participants in the s-ambiguous group heard all f-final words with the f-endpoints of the continua and the s-final words with the ambiguous sounds.

Participants were seated in a sound-proof booth. On every trial, participants listened to a word or nonword and had to indicate by button press whether they heard an existing Dutch word or not. Response options were displayed on the screen 500 ms before the audio started. The option *woord* ("word") was always displayed on the left side of the screen and corresponded to the left button. The option *geen woord* ("not a word") was displayed on the right and corresponded to the right button. The response options stayed on the screen until the participant responded. Participants were informed that their answer was registered by seeing the display of the chosen response option move approximately 1 cm upward on the screen where it stayed for 400 ms. Five hundred ms later (blank screen), the next trial started automatically. The instruction emphasized speed as well as accuracy of listeners' responses.

Words and nonwords were presented to participants in pseudo-random order. The experiment started with at least six filler word or nonword trials before an f- or s-final word occurred, and care was taken that critical trials (including f- or s-final words) did not directly follow one another. Every 50 trials participants were allowed to take a self-paced break. At the end of the exposure, phase participants were informed by means of written instructions to stay seated because the next part of the experiment was about to start.

**Test.** Immediately following exposure, all participants completed the same phonetic categorization task with the five Dutch minimal pairs. After participants pressed a button to confirm that they understood the written instructions about the phonetic categorization procedure, the screen went blank, and a prerecorded utterance of the speaker was played to the participants. This utterance announced that the second part of the experiment was about to start (*Het nieuwe gedeelte gaat nu beginnen*, "The new part will start now"). Note that the Dutch sentence does not contain any of the sounds /f/, /s/, /v/, or /z/. Two thousand, one hundred ms after audio offset, the first trial started with the presentation of two written words of a minimal pair on the screen. The word ending in /f/ was always displayed on the right. Five hundred ms later, the audio was played. Participants had to indicate which word they thought they heard by pressing a button on the respective side of a button box. Participants were informed that an answer had been registered by seeing the chosen word move 1 cm upward on the screen where it stayed for 400 ms. After 500 ms, the next trial started automatically.

Each of the four selected steps of each of the five continua was presented eight times, resulting in a total of 160 trials. Word pairs and steps were presented intermixed in a separate random order for each participant. Every 40 trials, participants were allowed a self-paced break. Exposure and test phase were implemented running Presentation software (Version 14.9, Neurobehavioural Systems Inc.). Completing the whole experiment took approximately 30 min.

## Results

### Exposure

All responses that were faster than 200 ms or slower than 2500 ms after target onset were counted as incorrect. Due to this procedure, 50 trials (0.83%) were relabeled. Overall, the accuracy in the lexical-decision task was high: 95.7% of the nonwords were correctly rejected, and 94.8% of the filler words were correctly accepted. Of the critical ambiguous f-final and s-final words, all participants accepted at least 17 of 20 words. Percentage correct responses and response latencies to the critical s-final and f-final words in their ambiguous and unambiguous forms are given in Table 1.

Linear mixed-effects models with response (1 = word, 0 = not a word; logit link function) and response latency (RT) as dependent variables tested potential differences in the acceptance of words with ambiguous and unambiguous fricatives in the s-final and f-final words. Trial Type (f-final words = 0.5, s-final words = −0.5) and Condition (f-ambiguous = −0.5, s-ambiguous = 0.5) were entered as fixed factors, and participant and item were entered as random factors. Model comparisons with log-likelihood ratio tests, in which nonsignificant interactions and factors were successively removed, showed that in the analyses of correct versus incorrect responses all factors dropped out of the model. That is, no differences between trial types and conditions were found. For the analyses of RTs an interaction between trial Type and Condition was found ($b_{\text{trialType}} = 11.57$, $SE = 46.3$, $t = 0.25$, $p = .76$; $b_{\text{Condition}} = -55.29$, $SE = 42.08$, $t = -1.31$, $p = .15$; $b_{\text{trialType*Condition}} = -51.00$, $SE = 22.49$, $t = -2.27$, $p < .05$). This interaction can be attributed to the fact that only f-final words with ambiguous sounds were responded to slower than their unambiguous counterparts.

### Test

Figure 1 shows the proportion of f-responses along the four step test continuum for the Dutch participants in Experiment 1. Dutch

Table 1
*Mean Percentage of Correct Responses and Mean RTs From Word Onset for Ambiguous and Natural f-Final and s-Final Words During the Auditory Lexical-Decision Exposure Phase in L1 and L2 Dutch (Experiments 1 and 2) and L2 English (Experiment 3)*

| | Natural | | | | Ambiguous | | | |
|---|---|---|---|---|---|---|---|---|
| | f-words | | s-words | | f-words | | s-words | |
| Experiment | % correct | RT (ms) | % correct | RT (ms) | % correct | RT (ms) | % correct | RT (ms) |
| 1 | 97.3 | 989 | 97.3 | 1031 | 96.7 | 1069 | 96 | 1004 |
| 2 | 91 | 1107 | 89 | 1162 | 88.3 | 1144 | 84.3 | 1136 |
| 3 | 85.7 | 1073 | 96.3 | 1001 | 82.7 | 1162 | 90.7 | 1003 |

*Note.* RT = response latency.

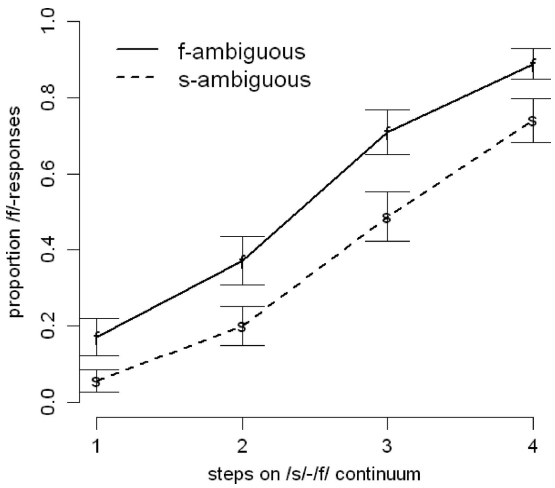


*Figure 1.* Proportion f-responses along the 4-step /s/–/f/ test continuum in Experiment 1. The solid line indicates responses for the f-ambiguous group; the dashed line indicates responses for the s-ambiguous group. The error bars represent one standard error.

listeners who had heard ambiguous sounds on f-final words during exposure chose the f-final word of the minimal pair more often (i.e., 53%) than Dutch listeners who had heard ambiguous sounds on s-final words (37%). As a first test of this difference, we compared the overall proportion of f-responses between the groups with a *t* test, which was statistically significant ($t = 3.4$, $p < .005$, $df = 28$). Note that the proportions going into this analysis are mostly between .3 and .7, the range in which the use of proportions in parametric tests is unproblematic. In a second step, linear mixed-effects models were fitted with Response (1 = f, 0 = s) as a dichotomous dependent variable, and Step (1–4 with the subtraction of 2.5 to center it on zero), Condition (f-ambiguous, s-ambiguous), and their interaction as fixed factors. Participant was entered as random factor such that the intercept, as well as the slope for the factor Condition, could vary randomly across participants with the restriction that the overall mean of this variation was zero. Nonsignificant interactions and factors were removed in a stepwise fashion given the simpler model fitted the data better than the model including the additional factor. Model comparisons were conducted using log-likelihood ratio tests. The results in Table 2 show that more f-responses were given the more f-like the final sound of the continuum and confirm that more f-responses were given by participants who, during exposure, had heard an ambiguous sound on f-final words. The effect of Condition is slightly larger than one logit, which translates into a group difference of 25% in the percentages of f-responses around the 50% point, and a group difference of 12% around the 80% point. This reflects the purpose of the logistic transformation, which is to give more weight to differences near the floor and the ceiling of the probability scale.

## Discussion

The purpose of Experiment 1 was to establish a perceptual learning effect with slightly modified stimuli and a slightly modified test phase compared to previous studies. With these changes, the learning effect was clearly observed. More importantly, the variance in responses between participants was reduced. Figure 2 shows comparisons between the range of individual means of proportion f-responses in three experiments: Experiment 1 of the present study (panel A), the study by Norris et al. (2003, Experiment 1; panel B), and the study by Eisner and McQueen (2006; panel C). The latter two studies only mixed the frication noises for the creation of the ambiguous sounds. The individual means of proportions f-responses are plotted in histograms with an overlaid density function. The distributions are narrower in the current data set than in the previous two experiments. The mean variance of these distributions in Figure 2 (in logistic space) is 0.95 for the earlier experiments, and it is 0.35 in our current data set. An *F* test comparing these variances indicates that this difference is significant, $F(30, 118) = 2.73$, $p < .01$. Although we cannot be sure which of our changes was critical, using multiple test continua or using the morphing of more than the fricative portion of the words, it is important for future research to ensure that the optimal type of sound manipulation is being used.

With regard to the present results, we also observed that during exposure participants were more inclined to accept words with ambiguous fricatives as real words than in previous studies. None of our participants had to be excluded from further analyses for rejecting more than half of the critical words as nonwords. Previous studies (e.g., Norris et al., 2003; Sjerps & McQueen, 2010) had applied such a criterion and rejected overall about 5% of the participants. Moreover, none of the participants in the present study responded with only one response option during the test phase (i.e., no-one gave only f-responses or only s-responses). We assume that this was due to the fact that in our stimuli all cues to

Table 2
*Results of the Mixed-Effects Models for the Test Phases of Experiments 1, 2, and 3*

| | Experiment 1 | | Experiment 2 | | Experiment 3 | |
|---|---|---|---|---|---|---|
| Factor | *b* (*SE*) | *p* | *b* (*SE*) | *p* | *b* (*SE*) | *p* |
| Intercept | 0.22 (0.26) | .40 | −0.94 (0.29) | <.01 | 0.91 (0.29) | <.01 |
| Condition | −1.07 (0.31) | <.001 | −1.63 (0.38) | <.001 | −1.38 (0.39) | <.001 |
| Step | 1.42 (0.04) | <.001 | 1.64 (0.07) | <.001 | 1.41 (0.04) | <.001 |
| Condition ∗ Step | — | — | −0.34 (0.09) | <.001 | — | — |

*Note.* The intercept term maps onto the middle of the continuum (which has been centered on zero for participants in the /f/-ambiguous groups. Beta weights of the fixed factors indicate the adjustment needed to map from the intercept to the other levels of the factor. Standard errors of the beta weights are given in brackets. Note that the respective z-values can be calculated by dividing the beta by the standard error.

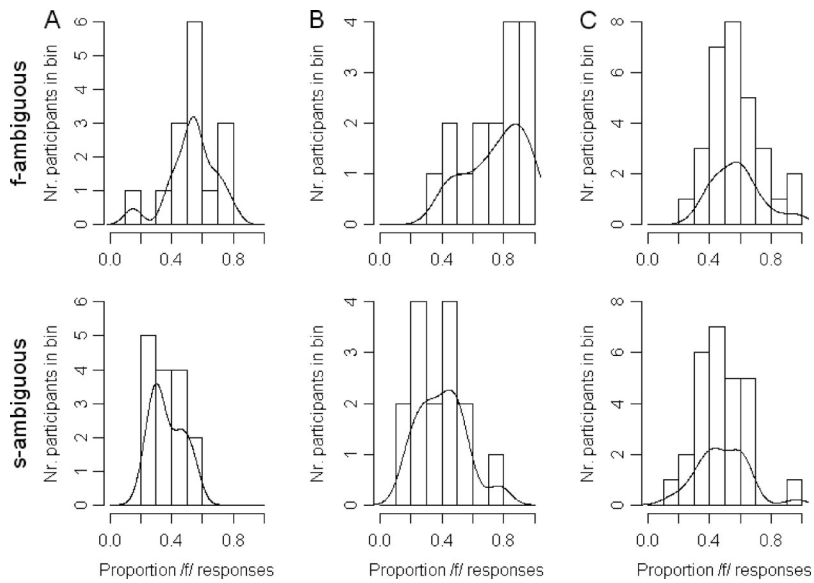


*Figure 2.* Histograms and density functions of the range of individual means of proportion f-responses per participant for Experiment 1 (panel A), Norris et al. (2003; panel B) and Eisner and McQueen (2006; panel C). The top row shows data for participants in the f-ambiguous groups. The bottom row shows data for participants in the s-ambiguous groups. Note that the different scales of the *y*-axis are a byproduct of the different sample sizes in the different experiments.

the fricatives were ambiguous, and participants were not misled by mismatching transitions.

It is also noteworthy that the use of morphed stimuli allows us to run perceptual learning studies on any kind of contrast. Previous studies had been limited to contrasts with one major cue (VOT in stops, fricative-noise spectrum in fricatives, and f0 contour in tones). These contrasts were to some extent chosen because the major cue is reasonably simple to manipulate by waveform editing or PSOLA resynthesis. None of these methods, however, is able to manipulate spectral properties in voiced signals. This severely limited the range of contrasts that could be investigated. By using morphed stimuli, it now becomes possible to generate ambiguous stimuli for a larger variety of contrasts, including those that differ spectrally in voiced portions of the signal.

## Experiment 2

### Methods

**Participants.** Thirty native speakers of German (mean age 24, range 20–31) participated for a small payment. They were recruited from the student population in Nijmegen, The Netherlands. They had started to learn Dutch when they came to the Netherlands to study. At the time of testing, they had been living in a Dutch-speaking environment for an average of 3.25 years (range 1–8). Their average self-rated proficiency in Dutch on a scale from 1 (*very poorly*) to 9 (*like a native speaker*) was 5.9 (range 2–8). At the time of testing, most of them reported using Dutch, German, and to a lesser extent English on a regular basis.

**Materials and design.** The materials and design were identical to Experiment 1. After the experiment, participants were required to fill in a questionnaire about their self-estimated proficiency in Dutch and their daily language use, as well as a multiple-choice test about the critical words used in the experiment. In that test, participants were presented with the critical Dutch words from the exposure phase and the words of the minimal pairs from the test phase in written form. Their task was to choose from a correct German translation (e.g., Dutch: *locomotief*, German: *Lokomotive* "locomotive"), a German translation of a phonological competitor of the Dutch word (Dutch: *locatie*, German: *Ort* "place"), and a German translation of a semantic competitor of the Dutch word (Dutch: *station*, German: *Bahnhof* "train station"). Participants were also allowed to indicate that they did not know the word.

### Results

**Language familiarity.** Of the 50 tested words (40 critical items plus five minimal pairs), only three words were unknown to more than half of the participants. Words of the minimal pair *kuif–kuis* ("head of hair," "chaste") were known to only two and seven participants, respectively. Also, the word *middenrif* ("diaphragm") from the f-final word set was known to only five of the 30 participants. On average, each word was known by 27 participants. Thus overall, participants were very accurate in selecting the correct answers. The average number of mistakes or unknown words was five per participant. The largest number of errors was made by only one participant who failed to select the correct answer in 14 of the 50 words.

**Exposure.** Again, all responses that were faster than 200 ms or slower than 2500 ms after target onset were counted as incorrect. Due to this procedure, 158 trials (2.63%) were relabeled. Participants correctly rejected 81.5% of the nonwords, and they accepted 88.5% of the filler words. Percentage correct responses and response latencies to the critical s-final and f-final words in

their ambiguous and unambiguous forms are given in Table 1. Linear mixed-effects models on response and RTs indicated no differences between trial Type (f-final vs. s-final words) and Condition (f-ambiguous, s-ambiguous) because all fixed factors dropped out of the models.

**Test.** Figure 3 shows the proportion of f-responses along the four-step test continuum for participants in Experiment 2. Participants who heard ambiguous sounds on f-final words during exposure chose the f-final word of the minimal pairs more often (63%) than participants who had heard ambiguous sounds on s-final words (39%). This overall difference in percentage of f-responses was statistically significant ($t = 4.3$, $p < .001$, $df = 28$). Linear mixed-effects models showed an effect of Step of the continuum such that the more f-like the final sound of the continuum was, the more f-responses were given (see Table 2). The factor Condition was also significant. More f-responses were given by participants who had heard an ambiguous sound on f-final words during exposure than participants who had heard an ambiguous sound on s-final words (see Table 2). These effects were mediated by a significant interaction. The effect of Condition was stronger the higher the steps of the continuum (i.e., the more f-like the sound was). Inspection of Figure 3 might suggest that the effect gets smaller again at step four, but note that this difference is larger in logistic space, in which the interaction has been calculated.

To assess whether the effect of phonemic retuning in a second language was the same size as for the native Dutch listeners in Experiment 1, linear mixed-effects models were fitted with participant Language ("du," "ge") as an additional fixed factor. Participant was entered as a random factor for which the intercept as well as the slopes for Condition and participant Language were allowed to vary (with the mean of the overall variation being zero). Nonsignificant interactions were successively removed from the model starting with the highest interaction and the interactions with the highest $p$ values. Results of the simplest model show effects of Condition ($b_{\text{condition}} = -1.33$, $SE = 0.25$, $z = -5.36$, $p < .001$; more f-responses in the f-ambiguous group), Step ($b_{\text{step}} = 1.55$, $SE = 0.05$, $z = 34.21$, $p < .001$; more f-responses the more f-like the final sound of the minimal pair continuum), and an interaction between Condition and Step ($b_{\text{condition*step}} = -0.21$, $SE = 0.06$, $z = -3.51$, $p < .001$; the effect of condition was stronger the higher the step of the continuum). Participant Language was neither significant as a main effect nor in interaction with any of the other factors. Therefore, these interactions with participant Language, as well as eventually the factor itself, were removed from the model ($\chi^2 = 1.64$, $p = .2$, $df = 1$). Thus, overall there was no difference in category retuning between first and second language listeners.

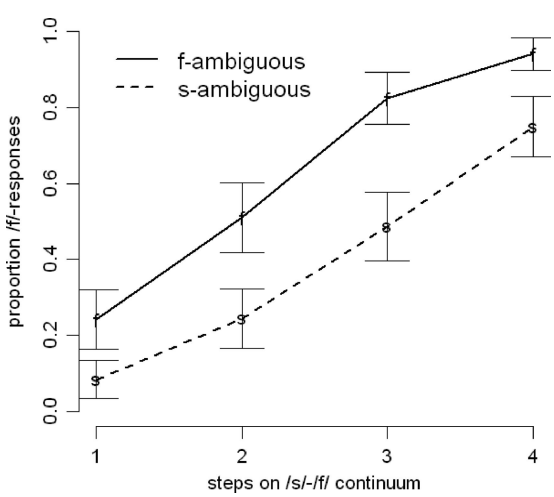


*Figure 3.* Proportion f-responses along the 4-step /s/–/f/ test continuum in Experiment 2. The solid line indicates responses for the f-ambiguous group; the dashed line indicates responses for the s-ambiguous group. The error bars represent one standard error.

## Discussion

Experiment 2 demonstrated for the first time that retuning of phoneme categories is possible in a second language. The phonological processing problems usually associated with processing a second language do not prevent the retuning of phoneme categories. It also seems that the putatively phonologically coarsely specified lexical representations in a second language are strong enough to provide guidance for category retuning. This is the case at least for advanced second language learners adapting categories that are similar in their first and second language. We could show that the effect of category retuning for German L2 listeners was as strong as the effect for Dutch L1 listeners in Experiment 1. Having established that under "optimal" conditions (i.e., proficient learners and similar phoneme categories in L1 and L2) a second language provides a sufficient training signal to guide category retuning, we set out to address the question about category retuning across languages.

# Experiment 3

## Methods

**Participants.** Thirty Dutch native speakers (mean age 20.3, range 18–29), who had not participated in Experiment 1 or in the pretests, took part for a small payment. They were recruited from the same student population and were randomly assigned to the English (Experiment 3) rather than the Dutch (Experiment 1) experiment. According to a language questionnaire that participants filled in after the experiment, they started learning English at school at an average age of 10.4 years (range 7–12). Their self-rated proficiency in English on a scale from 1 (*very poor*) to 9 (*like an English native speaker*) averaged 6.5 (range 5–8). Two thirds of the participants reported using English on at least a weekly basis. Many participants had knowledge of additional languages such as German, French, or Spanish.

**Materials.** A set of 100 English words and 100 nonwords was created in analogy to the exposure material used in Experiment 1. The same critical sounds /s/ and /f/ were used. Twenty words of the exposure set were f-final and did not create another existing English word when /f/ was replaced with /s/ (e.g., *autograph*). Twenty s-final words were selected that did not create another existing 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 any of the sounds /f/, /s/, /v/, or /z/. Additionally, words containing the dental fricatives /θ/, or /ð/ were excluded, because these two sounds are

regularly replaced with [s] or [z] by Dutch learners of English (e.g., Hanulikova & Weber, 2010; Wester, Gilbers, & Lowie, 2007). Critical words did not contain any of these sounds except for the critical sounds in word-final position.

Recordings were made in the same session as the Dutch recordings used in Experiments 1 and 2 by the same Dutch native speaker. The speaker had started to learn English at the age of 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, however, she is typically characterized by English native speakers as having a perceptible Dutch accent. For the critical words, the speaker was again instructed to pronounce them in pairs, correctly as well as with the last phoneme substituted by the other critical fricative. For example, *autograph* was also produced as *autogra*[s], and *notice* was also pronounced as *noti*[f]. The speaker was again reminded that the beginning of these pairs should sound as similar as possible to facilitate the manipulation procedure.

**Creating and pretesting ambiguous stimuli.** The procedure to create continua and find ambiguous tokens of the fricatives was identical to Experiment 1. The same 20 participants who took part in the pretest for the Dutch words in Experiment 1 also completed the English pretest. The two pretests had to be completed on different days. The order of language in the pretests (i.e., Dutch or English first) was counterbalanced across participants.

The procedure of selecting the most ambiguous steps of the continua was also identical to Experiment 1. Again, we tried to compensate for the lexicality effect by selecting f-final words that received approximately 70% f-responses and s-final words that received approximately 30% f-responses. As before, the mean step on the continuum that was selected to represent the most ambiguous sounds was matched between f- and s-final words (i.e., mean value for number of step [1–7; 7 = /f/] was 4.8 for both types of continua). Further, the mean percentage of f-responses for f- and s-final words was of similar distance from 50% (f-final words 24%, s-final words 29%). In comparison to the Dutch continua, the most ambiguous sounds for English were shifted toward the f-endpoints (step 4.35 for Dutch and step 4.8 for English). For the test phase, the same five four-step continua between Dutch minimal word pairs were used as in Experiment 1.

**Procedure.** The design and procedure of the exposure phase were identical to Experiment 1 except that the lexical-decision task was in English. Participants also received English written instructions to put them into an English language mode. They were informed that they were going to hear real English words and nonexistent English words spoken by a native speaker of Dutch. Their task was to decide as quickly, and as accurately as possible, whether the heard word was a real English word or not. Response options on the screen were given in English ("is a word," "is not a word"), as was the instruction to stay seated at the end of the exposure phase.

The test phase was identical to Experiment 1. Written instructions were given in Dutch, and before the first categorization trial started, the Dutch sentence *Het nieuwe gedeelte gaat nu beginnen* ("The new part will start now") was presented to switch listeners to a Dutch language mode. Immediately following the test phase, participants were asked to fill in a language background questionnaire and a vocabulary test. The vocabulary test was similar to the one described for the German participants in Experiment 2. On average, participants chose the correct translation for 44 of 50 words (range 34–49; the set consisted of the 40 critical words from the exposure phase plus 10 words from a set of English minimal pairs that were not used in the present study), with an average score of 87% correct.

## Results

**Exposure.** Again, all responses that were faster than 200 ms or slower than 2500 ms after target onset were counted as incorrect. Due to this procedure, 117 trials (1.95%) were relabeled. Overall, 86.6% of the nonwords were correctly rejected, and 95% of the filler words were correctly accepted. Percentage correct responses and response latencies to the critical s-final and f-final words in their ambiguous and unambiguous forms are given in Table 1. Linear mixed-effects models on response accuracy showed that listeners responded overall more accurately on s-final words than on f-final words ($b_{\text{trialType}} = -1.29$, $SE = 0.47$, $z = -2.75$, $p < .01$). Although participants were not overall less accurate when responding to ambiguous items ($b_{\text{condition}} = -0.41$, $SE = 0.33$, $z = -1.22$, $p = .22$), they were so within the set of s-final words ($b_{\text{trialType*condition}} = 1.53$, $SE = 0.50$, $z = 3.04$, $p < .01$). The analysis of RTs showed that listeners responded more slowly to f-final than s-final words ($b_{\text{trialType}} = 118.28$, $SE = 39.75$, $t = 2.98$, $p < .01$) and that, in addition, they were especially slow for f-final words in which the fricative was ambiguous between /f/ and /s/ ($b_{\text{trialType*condition}} = -89.73$, $SE = 29.95$, $t = -3.00$, $p < .01$). There was, however, no general slowing down due to ambiguous items ($b_{\text{condition}} = -44.09$, $SE = 49.50$, $t = -0.98$, $p = .37$).

**Test.** Figure 4 shows the proportion of f-responses along the four-step test continuum for participants in Experiment 3. Participants who had heard ambiguous sounds on the f-final words during exposure chose the f-final word of the minimal pair more often (i.e., 63%) than participants who had heard ambiguous sounds on s-final words (43%). This overall difference in percent-

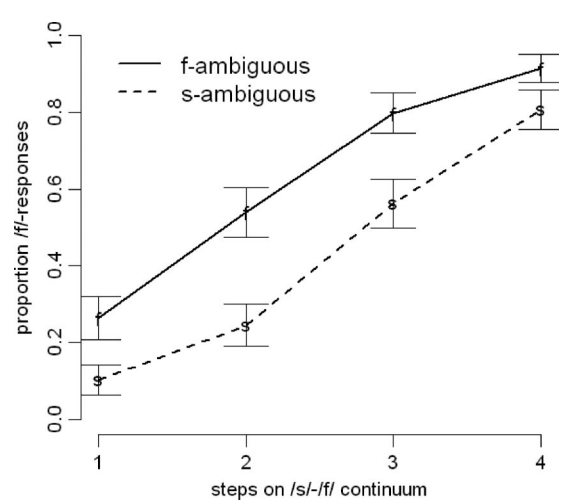


*Figure 4.* Proportion f-responses along the 4-step /s/–/f/ test continuum in Experiment 3. The solid line indicates responses for the f-ambiguous group; the dashed line indicates responses for the s-ambiguous group. The error bars represent one standard error.

age of f-responses was statistically significant ($t = 3.58$, $p < .001$, $df = 28$). Linear mixed-effects models showed that the more f-like the final sound of the continuum was, the more f-responses were given and furthermore confirmed that more f-responses were given by participants who had heard an ambiguous sound on f-final words during exposure (see Table 2). Phonemic retuning thus occurred across a language switch between English exposure and Dutch test phase.

To assess whether the effect of phonemic recalibration across languages was the same size as for the within-language native Dutch group in Experiment 1, a linear mixed-effects model with exposure Language ("du," "en") as an additional fixed factor was used. Model comparisons were carried out in the same fashion as before. Results show effects of Condition ($b_{\text{condition}} = -1.19$, $SE = 0.24$, $z = -4.88$, $p < .001$; more f-responses in the f-ambiguous group), Step ($b_{\text{step}} = 1.41$, $SE = 0.03$, $z = 47.55$, $p < .001$; more f-responses the more f-like the final sound of the minimal pair continuum), and exposure Language ($b_{\text{exposureLanguage}} = 0.49$, $SE = 0.24$, $z = 2.05$, $p < .05$; more f-responses for the English exposure group). The effect of exposure Language indicates a slightly different location of the retuned category boundary between participants from the two language groups. Note, however, that overall the effects for the English exposure group were as large as for the Dutch control group because no interaction was found between exposure Language and Condition.

### Discussion

Experiment 3 demonstrated that phonemic recalibration occurs across languages. That is, retuning of phoneme categories is not only possible in a second language, but the retuned category boundaries also transfer to the listeners' native language. The slightly shifted category boundary between the English exposure group and the Dutch control group likely reflects the slight shift of the English relative to the Dutch pretest continua. That is, the selected ambiguous sounds for the English exposure words were more f-like than the selected sounds for the Dutch exposure words. This suggests that the specific properties of the ambiguous sounds during exposure influence the precise location of the recalibrated phoneme boundaries but not the strength of the recalibration effect.

## General Discussion

We showed in three experiments that, in order to adapt to a speaker's unusual pronunciation variants, listeners not only retune phoneme boundaries in their native language but also in a foreign language. Moreover, retuned phoneme categories are even applied across languages. Specifically, we showed that German L2 listeners' categorization of Dutch ambiguous sounds along an /f/–/s/ continuum was influenced by preceding exposure to an ambiguous fricative in Dutch words (Experiment 2). The group of German L2 listeners who had learned to interpret the ambiguous sound as /f/ (i.e., in lexical contexts such as *lokomotie*[f/s]) categorized more steps along the continuum as /f/ than the group who heard the ambiguous sound as replacement for /s/ (e.g., in *geitenkaa*[f/s]). Notably, the magnitude of the shift in the category boundary between groups was similar to the Dutch L1 listeners in Experiment 1. Finally, Experiment 3 demonstrated that Dutch listeners also adjust their category boundaries for the Dutch test continua after having heard the ambiguous fricatives in English lexical context. Again, the effect was of the same size as for Dutch listeners who had heard the ambiguous fricatives in Dutch. Category retuning thus occurs across languages. Listeners who are exposed to a speaker in one language can adapt to this speaker's articulatory idiosyncrasies even when they later hear the same speaker using a different language.

These findings add to the large literature of how listeners flexibly adjust their perception to the listening environment (e.g., Samuel & Kraljic, 2009), and they are important for at least two reasons. First, listeners' retuning of phoneme categories in a second language is not self-evident given the smaller vocabulary size of L2 learners and the putative consequence of coarser phonological representations in the mental lexicon. Second, the transfer of retuned phoneme boundaries across languages would not be expected if all types of contextual differences between exposure and test phase block adaptation (e.g., Kraljic et al., 2008; Kraljic & Samuel, 2011).

In order to address the questions of category retuning in a second language and across languages, we first had to establish a method that allowed us to optimally measure the amount of adaptation in L1. To reduce the variability in our data as compared to previous studies (e.g., Eisner & McQueen, 2006; Norris et al., 2003), we improved our material in two ways. Stimuli were generated by audio-morphing of naturally produced words (or parts thereof) rather than mixing only frication noise. The morphing ensured that the stimuli were ambiguous with regard to all cues relevant for the distinction of the trained contrast. Listeners could thus not be misled or distracted by mismatching cue information such as transitions cuing the "wrong" fricative. In addition, for the test phase we made use of continua between five minimal pairs rather than using only one nonword–nonword continuum. This deterred listeners from responding strategically based on acoustic comparisons of test stimuli. The resulting stable effects were a good basis for looking at adaptation in L2 and across languages.

In Experiment 2, German learners of Dutch adapted just as well to the odd pronunciations of the fricatives as did Dutch L1 listeners in Experiment 1. This is good news for second language learners. Phoneme categories can be flexibly adjusted in L2 listening, at least when the categories resemble the ones that are also present in L1, and when the learners have a good, though not perfect, command of the second language lexicon to guide the category retuning. Having shown that category retuning is possible in a second language opens a whole new field of questions to be addressed. Numerous studies have shown, for example, that for certain phoneme contrasts, even highly proficient learners do not obtain native-like discrimination abilities in their L2 (e.g., Cutler, Weber, Smits, & Cooper, 2004) and that difficult phoneme contrasts resist mastery even after extensive training (e.g., Bradlow, Pisoni, Yamada, & Tohkura, 1997). How would listeners treat adaptation contexts for these contrasts? If they either rigidly stick to their poorly established categories or, alternatively, just accept any instance of the sounds as reasonable tokens, no explicit category shift would be expected. Such a finding, however, can only be interpreted once the possibility of category retuning in a second language has been shown. This was the purpose of the present study.

The current findings are also in line with broader learning effects demonstrated in a study on the use of subtitles for adaptation in a second language (Mitterer & McQueen, 2009). In this study, Dutch participants who had watched a Scottish or Australian English movie with English subtitles during exposure were better at repeating back new utterances from the respective main characters of the movies than participants who had seen the same shows either without subtitles or with Dutch subtitles. Mitterer and McQueen argued that this was a reflection of lexically guided learning. English subtitles provided the listeners with feedback about the intended word form. Dutch subtitles, in contrast, only provided the meaning and, indeed, those participants who watched the TV shows with subtitles in their native language Dutch showed hardly any adaptation effect. In the present study, we showed for the first time that perceptual adaptation in L2 can occur in a narrow sense. That is, listeners are able to use their second language lexicons to flexibly shift single phoneme boundaries as has been previously shown for native phoneme categories.

In this way, findings on L2 category retuning can also inform models of the architecture of the L2 speech recognition system. A possible explanation for why we found category retuning in the second language, as well as across languages, is the fact that we tested sounds that occur in the listeners' first and second languages. The fricatives /f/ and /s/ both exist in Dutch, German, and English. The L2-version of the Perceptual Assimilation Model (PAM-L2, Best & Tylor, 2007) as well as the Speech Learning Model (SLM, Flege, 1992, 1995) suggest that second language phoneme categories are perceived in relation to the native language phoneme space. L2 categories that are similar to L1 categories are thereby assimilated to native phonemes (SLM) or, in case there is no perfect match, are established as allophones of native phonemes (PAM-L2). In our case, the fricative /f/ is an excellent match across the languages tested, and hence should be assimilated easily with no need for a language-specific phonetic category. The Dutch /s/, however, would probably require a language-specific phonetic category (i.e., allophone) in terms of PAM-L2. Dutch /s/ is known to differ phonetically from German and English /s/ (see Collins, & Mees, 1996, p. 145, for a comparison to English). Thus in our case, at least for /f/ and possibly also for /s/, listeners could have shifted L1 phonemes by using L2 lexical context to guide the retuning. For phonemes that differ phonetically more strongly across languages (e.g., Dutch and English onset stops) or that even fall in different categories across languages (e.g., English /θ/ has no equivalent category in Dutch or German), however, the newly established L2 phonemes would have to be shifted. Further research will have to show whether this selective retuning is possible, that is, whether nonmatching L2 categories can be retuned.

Finally, Experiment 3 showed that second language lexical contexts can guide category retuning when the speaker has a foreign accent. Even though listeners are expected to adapt to the foreign accent and understand the speaker better after exposure (e.g., Bradlow & Bent, 2008), effects of foreign accents on native language perception have not been demonstrated yet. It is also unclear why listeners would attribute the unusual pronunciation of the fricatives to the speaker's speech characteristics in general (which are also present in native speech) rather than to the presence of the foreign accent (i.e., a transient cause). Of course, in the present study, the speaker had the same accent as the listeners (i.e., Dutch). However, if it was crucial that the listeners are able to produce the same accent as the speaker, no effect should have been found after exposure to a native speaker of the second language (Experiment 2).

The current results, therefore, show that not all differences between exposure and test conditions block category retuning. Although seeing the speaker being impeded by a pen in the mouth does block category retuning (Kraljic et al., 2008), hearing the speaker using a foreign language—a different form of impediment—does not block category retuning. It should be noted that this was found with a speaker who had a foreign accent, which suggests that the speaker is influenced by his or her L1 categories in L2 production. This may foster generalization across languages. It remains an open question whether retuning would also occur with a balanced bilingual speaker. Yet the present findings further specify the mechanisms of lexically guided perceptual learning. Our series of experiments demonstrated that not all aspects of the listening situation are stored as episodes alongside with a speaker's phoneme realizations. Language identity appears to be abstracted or disregarded. We showed that exposure and test language do not have to be the same. The fact that a possible episodic system seems very selective in what information it stores does not speak in favor of episodic accounts as being the most parsimonious.

## References


Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science, 14,* 592–597. doi:10.1046/j.0956-7976.2003.psci_1470.x

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam, The Netherlands: John Benjamins.

Boersma, P., & Weenink, D. (2009). *PRAAT: Doing phonetics by computer* (version 5.1) [software application]. Retrieved from http://www.praat.org

Bohn, O.-S., & Munro, M. J. (Eds.). (2007). *Language experience in second language speech learning: In honor of James Emil Flege.* Amsterdam, The Netherlands: John Benjamins.

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

Bradlow, A. R., Pisoni, D., Yamada, R. A., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101,* 2299–2310. doi:10.1121/1.418276

Bryant, S. A. (1997). Ability to encode phonological representations: An underlying difficulty for poor readers. In B. A. Blachman (Ed.), *Foundations of reading acquisition and dyslexia* (pp. 21–46). Mahwah, NJ: Earlbaum.

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition, 33,* 433–461. doi:10.1017/S0272263111000040

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

Collins, B., & Mees, I. M. (1996). *The phonetics of English and Dutch.* Leiden, The Netherlands: Brill.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America, 116,* 3668–3678. doi:10.1121/1.1810292

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67,* 224–238. doi:10.3758/BF03206487

Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *Journal of the Acoustical Society of America, 119,* 1950–1953. doi:10.1121/1.2178721

Flege, J. E. (1992). Speech learning in a second language. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development. Models, research, implications* (pp. 565–604). Timonium, MD: York Press.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 110–125. doi:10.1037/0096-1523.6.1.110

Hanulikova, A., & Weber, A. (2010). Production of English interdental fricatives by Dutch, German, and English speakers. In K. Dziubalska-Kołaczyk, M. Wrembel, & M. Kul (Eds.), *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010* (pp. 173–178). Poznan, Poland: Adam Mickiewicz University.

Jesse, A., & McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic Bulletin & Review, 18,* 943–950. doi:10.3758/s13423-011-0129-2

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. doi:10.1016/S0167-6393(98)00085-5

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. doi:10.1111/j.1467-9280.2008.02090.x

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

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language, 56,* 1–15. doi:10.1016/j.jml.2006.07.010

Kraljic, T., & Samuel, A. G. (2011). Perceptual learning evidence for contextually-specific representations. *Cognition, 121,* 459–465. doi:10.1016/j.cognition.2011.08.015

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology, 55,* 306–353. doi:10.1016/j.cogpsych.2007.01.001

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

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

McQueen, J. M., Norris, D., & Cutler, A. (2006). The dynamic nature of speech perception. *Language and Speech, 49,* 101–112. doi:10.1177/00238309060490010601

Metsala, J. L., & Walley, A. C. (1997). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Mahwah, NJ: Erlbaum.

Mitterer, H., Chen, Y., & Zhou, X. (2011). Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm. *Cognitive Science, 35,* 184–197. doi:10.1111/j.1551-6709.2010.01140.x

Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS One, 4,* e7785. doi:10.1371/journal.pone.0007785

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

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, China: Department of Chinese, Translation and Linguistics, City University of Hong Kong.

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics, 71,* 1207–1218. doi:10.3758/APP.71.6.1207

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. doi:10.1037/a0016803

Wester, F., Gilbers, D., & Lowie, W. (2007). Substitution of dental fricatives in English by Dutch L2 speakers. *Language Sciences, 29,* 477–491. doi:10.1016/j.langsci.2006.12.029

Received November 21, 2011
Revision received February 14, 2012
Accepted February 17, 2012 ■