Journal of Experimental Psychology:
Human Perception and Performance
2015, Vol. 41, No. 3, 710–722

0096-1523/15/$12.00 http://dx.doi.org/10.1037/a0039028

Divide and Conquer: How Perceptual Contrast Sensitivity and Perceptual Learning Cooperate in Reducing Input Variation in Speech Perception

Matthias J. Sjerps
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Eva Reinisch
Ludwig Maximilian University Munich

Listeners have to overcome variability of the speech signal that can arise, for example, because of differences in room acoustics, differences in speakers' vocal tract properties, or idiosyncrasies in pronunciation. Two mechanisms that are involved in resolving such variation are perceptually contrastive effects that arise from surrounding acoustic context and lexically guided perceptual learning. Although both processes have been studied in great detail, little attention has been paid to how they operate relative to each other in speech perception. The present study set out to address this issue. The carrier parts of exposure stimuli of a classical perceptual learning experiment were spectrally filtered such that the acoustically ambiguous final fricatives sounded relatively more like the lexically intended sound (Experiment 1) or the alternative (Experiment 2). Perceptual learning was found only in the latter case. The findings show that perceptual contrast effects precede lexically guided perceptual learning, at least in terms of temporal order, and potentially in terms of cognitive processing levels as well.

Keywords: perceptual contrast effects, perceptual learning, speech perception, speech perception models

Understanding speech involves the rapid mapping of an acoustic signal onto lexical representations. This mapping is not straightforward, as instances of the same word may be spoken very differently on different occasions. Listeners have to continuously adjust perception to overcome the influence of multiple sources of variation. One could think, for example, of someone speaking with a strong accent in a room that happens to attenuate high frequencies somewhat more than lower ones. In this situation, listeners may rely on at least two types of adaptation processes that allow them to better understand what is being said. Perceptual learning has been argued to occur, for example, as a means to adjust to the speaker's accent (Norris McQueen, & Cutler, 2003); perceptual contrast effects have been argued to help in dealing with unusual filter properties of transmission channels (Watkins, 1991), such as, in this case, that of the room. It is, however, unclear how these two processes (co)operate in everyday listening situations. Here, we address the question of to what extent these processes may differ in the temporal or cognitive locus of their most dominant influences during speech perception in a single experimental design. The goal is to assess the speech perception process in increasingly natural situations in which listeners take into account the consequences of multiple sources of variation at a time.

Over the last decades, experimental evidence has accumulated suggesting that input variation may be dealt with by a number of functionally different processes in speech perception (Holt & Lotto, 2002; Sjerps, Mitterer, & McQueen, 2011a; Watkins, 1991). Previous research, reviewed in the following sections, along with modeling approaches (see the Appendix), have led us to hypothesize that among these processes, perceptual contrast effects and lexically guided perceptual learning may at least partly apply in a certain temporal order. Specifically, contrast effects may be more dominant at cognitive levels that precede those at which lexically guided perceptual learning in speech perception takes place.

Perceptual Contrast Effects

It has been shown that preceding acoustic context can influence the perception of a following target sound. This allows listeners to perceptually resolve variation that arises as a result of, for example, room acoustics or speakers' vocal tract differences (Ladefoged & Broadbent, 1957; Watkins, 1991). A defining characteristic of these effects is that they are mostly contrastive. In the case of vowel perception, for instance, spectral properties of a preceding context have been shown to influence the location of the category boundary between phonemes such as /ɪ/ and /ɛ/ (e.g., Ladefoged & Broadbent, 1957; Reinisch & Sjerps, 2013; Sjerps et al., 2011a; Watkins, 1991). Similar effects have been observed with filtering of a preceding context. An ambiguous sound is perceived as the perceptual inverse of the filter that is used to manipulate a pre-

This article was published Online First March 23, 2015.

Matthias J. Sjerps, Psychology of Language Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; Eva Reinisch, Institute of Phonetics and Speech Processing, Ludwig Maximilian University Munich.

Matthias J. Sjerps is now at the Centre for Cognition, Donders Institute for Brain, Cognition and Behavior, Radboud University, and the Department of Linguistics, University of California, Berkeley.

We thank the research assistants of the Psychology of Language Department at the Max Planck Institute for Psycholinguistics for their help with running the experiments. Both authors contributed equally to the experiments and to writing the paper.

Correspondence concerning this article should be addressed to Matthias J. Sjerps, Department of Linguistics, University of California, Berkeley, Berkeley, CA 94720. E-mail: m.j.sjerps@gmail.com

ceding sound (Watkins, 1991). That is, if an ambiguous sound between /f/ and /s/ (here notated as [$^{s}_{f}$] to indicate ambiguity) is preceded by a sound that is filtered with an /f/-minus-/s/ filter (a filter that suppresses those frequency regions that are more dominant in /s/ than in /f/ and excites those frequency regions that are more dominant in /f/ than in /s/), listeners will interpret the ambiguous sound as more /s/-like. In analogy, an ambiguous sound will be more often interpreted as /f/ when it is preceded by a sound filtered with an /s/-minus-/f/ filter. Perceptual contrast effects (also referred to as *acoustic context effects*, *perceptual calibration*, *compensation*, and *normalization*; e.g., Stilp, Alexander, Kiefte, & Kluender, 2010; Watkins, 1991) can be considered part of a more general class of contrastive processes that are pervasive in perceptual processing and that act to increase the dynamic range of perception across all modalities (see Kluender, Coady, & Kiefte, 2003).

Regarding the cognitive locus of perceptual contrast effects, different types of evidence point to a relatively early, general auditory locus. For example, it has been shown repeatedly that a context segment or sentence spoken by one speaker can influence the perception of a target sound spoken by another speaker (e.g., Newman & Sawusch, 2009; Watkins, 1991). Moreover, on a number of occasions, qualitatively similar contrastive effects on speech sounds have been observed with nonspeech contexts (Holt, 2005, 2006; Sjerps, Mitterer, & McQueen, 2012; Watkins, 1991; Watkins & Makin, 1994), and linguistic exposure (language-specific category structure) does not have a substantial effect on the magnitude of compensation effects (Sjerps & Smiljanić, 2013). That is, to a large extent, these types of perceptual contrast effects are not language or speech specific (but, for a discussion of speech-specific contributions in the closely related domain of compensation for coarticulation, see Viswanathan, Fowler, & Magnuson, 2009; Viswanathan, Magnuson, & Fowler, 2010; see also Holt & Lotto, 2002; Holt, Lotto, & Kluender, 2000).

The available data therefore suggest that an important part of perceptual contrast effects may operate on perceptual representations that consist of frequency or feature information. However, researchers have been able to provide a lower bound on the cognitive level of implementation of at least an important portion of perceptual contrast effects. A context sound that is presented to one ear can influence the perception of a target sound that is presented to the other ear. This suggests that acoustic context is mostly taken into account at central auditory processing levels (i.e., it is not only a result of peripheral masking), occurring after the level of interaural integration (see e.g., Sjerps et al., 2012) that takes place at the level of the brainstem (Cant, 1992). As for the time course of perceptual contrast effects, they result from the immediate acoustic context and are observable in every instance of context–target pairings (they are observed when different context conditions are presented intermixed) and as early as the unfolding speech signal is being interpreted—that is, they do not merely influence participants' judgements at a postperceptual stage (Reinisch & Sjerps, 2013).

Lexically Guided Perceptual Learning

Listeners can quickly adapt speech perception to accommodate a speaker's idiosyncratic pronunciation variants—for example, by using lexical context to map ambiguous sounds to the relevant categories (McQueen, Cutler, & Norris, 2006; Norris et al., 2003; Reinisch, Weber, & Mitterer, 2013; Sjerps & McQueen, 2010; for an overview, see Samuel & Kraljic, 2009). For example, when a particular speaker consistently produces a variant of /f/ that is ambiguous between /f/ and /s/ (e.g., producing "gira[$^{s}_{f}$]" for *giraffe*), listeners shift their phonetic category boundary so as to include that variant of /f/ in their /f/ category (Clarke-Davidson, Luce, & Sawusch, 2008; Kraljic & Samuel, 2005; Norris et al., 2003). Notably, the same ambiguous sound can also be learned to be interpreted as an instance of /s/ if the ambiguous sound occurs in words in which it replaces /s/ (Norris et al., 2003). In other words, listeners use lexical information to change, or retune, the mapping from auditory signals to prelexical representations.

With respect to a cognitive processing hierarchy in speech perception, it has been found that retuned phonetic categories are not specific to the words they have been heard in: Listeners generalize the perceptual remappings across words (McQueen et al., 2006) and even across positions of a word, suggesting a prelexical locus (Jesse & McQueen, 2011; for discussions of the units that are affected by perceptual learning, see Mitterer, Scharenborg, & McQueen, 2013; Poellmann, Bosker, McQueen, & Mitterer, 2014; Reinisch, Wozny, Mitterer & Holt, 2014). This prelexical nature of adjustments provides an upper bound to the implementation of perceptual learning. In addition, however, there are also empirical arguments to assume a lower bound on these processes. Adjustments in fricative mappings apply not across the board but, rather, to new items from the same speaker or tokens from speakers who produce highly similar fricative tokens (Eisner & McQueen, 2005; Kraljic & Samuel, 2005; Reinisch & Holt, 2014). Moreover, on some occasions, perceptual learning appears to be dependent on the context situation (Kraljic, Samuel, & Brennan, 2008). This speaker or context specificity suggests a relatively higher cognitive level of implementation for perceptual learning relative to perceptual contrast effects.

One important point to consider is that two types of time course are involved in perceptual learning. First, on each encounter of an ambiguous sound like [$^{s}_{f}$] in a word like *gira*[$^{s}_{ʃ}$], lexical information has to inform the listener that the intended sound was /f/. Depending on the model of speech processing, this involves online feedback from the lexical level to the prelexical level (as suggested in interactive models of speech perception; e.g., TRACE [McClelland & Elman, 1986]) or it affects a decision stage at which prelexical and lexical information is merged (as in feedforward models like Shortlist B [Norris & McQueen, 2008] or the implementation of this process in Merge, [Norris, McQueen, & Cutler, 2003]). The second type of time course relates to the actual long-term retuning during perceptual learning. It has been shown that about 10–20 instances of the ambiguous sound in an unambiguous context have to be experienced to influence the interpretation in lexically ambiguous contexts (Kraljic et al., 2008; Poellmann, McQueen, & Mitterer, 2011). This is what feed forward models of speech perception would call *feedback for learning* (Norris & McQueen, 2008). Thereby, on encountering pronunciations like *gira*[$^{s}_{ʃ}$], the weights or expectations that associate incoming ambiguous input with one or the other segmental interpretation are gradually shifted toward the lexically supported category (note that this long-term adjustment also holds for interactive models like TRACE). This time course, spanning multiple encounters of learning contexts at the experiment level, and its

dependence on lexical activation at the trial level are a crucial difference from perceptual contrast effects (see the Appendix for details).

Although most of these observations seem to be in line with an implementation of perceptual learning at a higher level than perceptual contrast effects, there is evidence for very early learning effects in closely related domains. Krishnan, Xu, Gandour, and Cariani (2005), for example, showed that Chinese listeners exhibit stronger pitch representation and smoother pitch tracking than English listeners at the level of the auditory brainstem, and a number of other studies have observed reliable effects of learning at the level of the brainstem as well (see, e.g., Skoe, Krizman, Spitzer, & Kraus, 2013). This provides strong evidence that linguistic experience, or learning, can influence processes at relatively early physiological levels of processing. Because perceptual learning, hence, appears not to be restricted to one cognitive level, the present study set out to assess the relation of lexically guided perceptual learning to perceptual contrast effects.

The Current Project

The research just reviewed suggests that perceptual contrast effects may at least partially apply before the adjustments that are made in lexically guided perceptual learning. This cognitive ordering can be conceptualized in at least two different ways. The first is that the two processes operate at successive stages[1] in a hierarchy of neuronal populations that display sensitivity to patterns of increasing complexity. If, indeed, perceptual contrast effects operate earlier and at a lower level than the locus of retuning in perceptual learning, the learning mechanisms involved in retuning could only operate on perceptual representations that had already been adjusted by perceptual contrast effects. In a situation in which variation occurs because of steady filter properties, contrast effects may then reduce the effects of those filter properties early on. This would then require only minimal changes at the level at which perceptual learning is implemented. This interpretation is, in fact, fully in line with modeling approaches that describe these processes within the framework of TRACE (e.g., McClelland & Elman, 1986; see the Appendix for a detailed description of the two processes). It has been argued that acoustic context effects (in our case, instantiated as perceptual contrast effects) are most straightforwardly modeled at the featural level (McClelland, Mirman, & Holt, 2006; for additional modeling-based evidence in favor of a low-level implementation of contrast effects, see Apfelbaum & McMurray, 2014), whereas the retuning in lexically guided perceptual learning could best be modeled at the level of connection weights mapping from feature to phoneme units (Mirman, McClelland, & Holt, 2006). With regard to contrast effects, feature nodes are interpreted relative to the features of preceding time slices and only then map "up" to the phoneme level through the connections that—via lexical feedback—are affected in perceptual learning.

A second possible implementation is to relate the two processes without the assumption of different levels of processing in speech perception. That is, perceptual contrast effects and perceptual learning could partially be implemented in parallel. Perceptual contrast effects would still have to precede the lexical level but not necessarily the locus of prelexical remappings triggered by perceptual learning. Both retuning and contrast effects could then operate on the same ambiguous signal, but contrast effects would prevent retuning of the phoneme category by preventing a lexical mismatch signal. This option implements the same functional separation as the first one, but it does so by assuming only a difference in timing. These two possible implementations are discussed further in relation to the results of our study in the General Discussion.

Regardless of which of these two options is more likely, the current study was set up to test the shared hypothesis that perceptual contrast effects precede lexically guided perceptual learning, at least in terms of time course. Although we have already presented evidence for this assumption, so far the relation between these two processes has not been tested directly. Moreover, some evidence, such as effects of learning at the level of the brainstem (e.g., Krishnan et al., 2005), makes alternative implementations plausible. Testing this assumption directly could therefore be useful for future modeling attempts.

The present study consisted of two experiments following the classical lexically guided perceptual learning paradigm using ambiguous sounds between /f/ and /s/ in Dutch (Eisner & McQueen, 2006; McQueen et al., 2006; Norris et al., 2003; Reinisch et al., 2013; Sjerps & McQueen, 2010). For both experiments, stimuli from a previously reported perceptual learning experiment (Reinisch et al., 2013) were used as the basic stimuli and are referred to as the *no-filter condition*. These stimuli were chosen because they have been shown to elicit strong learning effects. This allowed for a comparison of effect sizes between the no-filter condition and the present study. All exposure stimuli except for the critical fricatives were filtered. Filtering provided the acoustic context expected to shift the perception of the ambiguous fricatives in a spectrally contrastive manner (i.e., elicit perceptual contrast effects). In this way, lexically guided perceptual learning could be set in relation to perceptual contrast effects.

In Experiment 1, the filters were designed to make the acoustically ambiguous fricatives used in Reinisch et al. (2013) sound less ambiguous and, hence, potentially attenuate perceptual learning. The basic logic is as follows: If perceptual contrast effects indeed resolve the input variation because of filter properties before lexically guided retuning can trigger learning, this should result in a reduction of the perceptual learning effect (relative to the no-filter condition).

In Experiment 2, the opposite type of filter was applied. This served as a control to test whether any effects observed in Experiment 1 could have been attributable to the procedure of filtering itself rather than the nature of the filter. Further, applying acoustic filters that shift perception toward the other alternative would help to explore the limits of perceptual learning. The magnitude of the learning effect may increase if the critical sounds are perceived as perceptually further away from the lexically supported target category. These combined tests allowed us to determine to what extent perceptual contrast effects and lexically guided perceptual learning operate, at least partially, in a certain temporal order.

[1] The reference to stages of processing is not meant to suggest a strict division or temporal ordering of processes—indeed, there is likely to be some overlap (details on what accounts could be predicted are provided later).

Experiment 1: Filtering to Reduce Ambiguity

On the basis of the study by Reinisch et al. (2013), which contributed the no-filter condition, lexically guided perceptual learning was tested in a between-groups design in which one group of listeners heard an ambiguous sound in the final position for words that normally end with /f/ (the /f/-trained group), whereas another group heard an ambiguous sound in the final position for words that normally end with /s/ (the /s/-trained group). In Experiment 1, for the /s/-trained group, all exposure materials (except for the critical final fricatives) from the no-filter condition (materials from Reinisch et al., 2013) were filtered such that those frequencies that are dominant in /s/ were suppressed. This should have made the sound that was ambiguous in the no-filter condition less ambiguous for the following reason. Listeners experience suppressed high frequencies in their input. What remains of the high-frequency noise in the unfiltered ambiguous [$^{s}_{f}$], which usually cues /s/, should therefore be perceptually prominent, making the sound more /s/-like. Similarly, materials for the /f/-trained group were processed with a filter that suppressed frequency regions characteristic of /f/, so the ambiguous sound became perceptually more /f/-like. We predicted that if perceptual contrast effects deal with such changes in general filtering properties first, these manipulations would cause the ambiguous sounds to no longer be perceived as (fully) ambiguous. As a result, the lexically guided updating of phoneme categories would induce no (or only a small) change in phoneme category representations. In contrast, if remappings in perceptual learning operate in parallel, then a learning effect would be found, because a mapping would be made from the ambiguous (untransformed) representation of the phoneme to the lexically supported category (i.e., the untransformed, ambiguous, representation would be associated with occurrences of a particular phoneme).

Method

Participants. Thirty native speakers of Dutch were recruited from the Max Planck Institute for Psycholinguistics participant pool. They were between 18 and 30 years of age and were mostly sampled from the student population of Nijmegen, The Netherlands. All participants reported not having hearing or language impairments. They received a small financial reward for their participation.

Materials. The materials were the same as those used in the no-filter condition reported in Reinisch et al. (2013) except that the stimuli were filtered (details are provided later). We briefly summarize the stimulus set and construction of ambiguous fricatives in the no-filter condition but refer readers to the original article for a more detailed description.

One hundred Dutch words and 100 nonwords that were phonologically legal in Dutch were used as exposure materials for an auditory lexical-decision task. 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 were nonwords if the fricatives were exchanged (*locomotie*[s] and *geitenkaa*[f] are nonwords in Dutch). None of the words or nonwords contained the sounds /f/, /s/, or their voiced counterparts /v/ and /z/ except for in the word-final position of the critical items.

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” [in the sense of bravery]), *roof–roos* (“robbery,” “rose”), *half–hals* (“half,” “neck”), and *kuif–kuis* (“tuft of hair,” “chaste”). All stimuli were recorded by a female Dutch native speaker (age 28 years) in a soundproof booth. All critical words were recorded with the correct fricative and the respective other fricative. In this way, ambiguous stimuli could be created from natural utterances of each word.

Creating ambiguous stimuli. For each /f/-final and /s/-final recording of the critical training words, as well as the minimal pairs for testing, the fricatives plus one or two preceding phonemes (mostly corresponding to the last syllable) 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, Masuda-Katsuse, & de Cheveigné, 1999) in MATLAB (MathWorks, Natick, MA). Time anchors at phonetically salient points in the speech signal were used for the morphing procedure to morph only phonetically similar parts of the signal (e.g., frication noise with frication noise, vocalic portions of the signal with other vocalic portions). The time anchors further allowed for the interpolation of durational differences between the morphed segments. Morphing larger portions of the signal than the critical fricatives ensured that other potential cues to the fricatives, such as formant transitions, were also set to ambiguous values. The word onsets onto which the manipulated signals were spliced back were selected from the correct recordings or the recordings with the respective other fricative depending on the naturalness of the resulting tokens. All splicing was done at positive zero-crossings using Praat (Version 5.1; Boersma & Weenink, 2009).

To find the most ambiguous steps of the continua to be used in the perceptual learning experiment, all continua were subjected to a pretest (reported in Reinisch et al., 2013). For the critical items to be used during exposure, a single ambiguous token was selected. For each of the minimal pairs (the test items), four stimuli were selected from the ambiguous part of the continuum spanning the 50% /f/-response mark between the middle two steps.

Filtering the stimuli. For the present experiment, all training stimuli used in Reinisch et al. (2013) were further manipulated. Two acoustic filters were created from the most /f/-like and /s/-like fricative tokens from each of the five minimal-pair continua. First, the long-term average spectrum (LTAS) was calculated for each of the fricatives using a 10-Hz bin size as implemented in Praat. From these values, an overall average /f/ LTAS and an overall average /s/ LTAS were calculated. These LTAS values were, thus, representations of the average spectral properties of the /f/ and /s/ endpoint tokens used in the test phase. Two different LTASs were then calculated to be used as filters: an /s/-minus-/f/ LTAS and an /f/-minus-/s/ LTAS (for each frequency bin, we subtracted the number in one filter from that for the other). To increase the distinctiveness between the two filters, the value obtained for each frequency bin was multiplied by 2. The resulting frequency distribution of the /s/-minus-/f/ filter is plotted in Figure 1 (the /f/-minus-/s/ filter is its inverse; i.e., each value is multiplied by −1).

When a speech signal is passed through the filter displayed in Figure 1 (i.e., the /s/-minus-/f/ filter), the signal’s frequencies around 5000 Hz are enhanced. The peak around 5000 Hz is a result of the fact that /s/ has a higher amplitude than /f/ around that

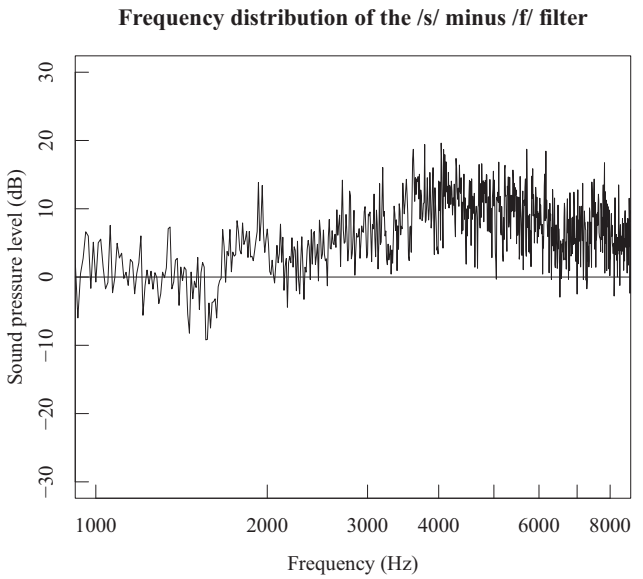

Figure 1. Filter properties of the /s/-minus-/f/ filter.

frequency. The /f/-minus-/s/ filter would be a mirror image, as each frequency bin would be multiplied by −1. That is, the /f/-minus-/s/ filter would have a trough around 5000 Hz and would attenuate the amplitude of those frequencies accordingly.

These filters were applied to all words and nonwords used for exposure. Of the critical training words, only the part up to but excluding the fricatives was filtered (the fricatives were to be interpreted relative to filtered context). All manipulated materials were filtered with both the /f/-minus-/s/ and /s/-minus-/f/ filters. The different exposure groups (described in the next section) were presented with different subsets of these items. The minimal pairs used in the test phase were left unchanged—that is, they were identical across conditions and the same as in the no-filter condition (reported in Reinisch et al., 2013). See Table 1 for an overview of the filters applied to the materials of the different experiments.

Procedure: Exposure. Participants were randomly assigned to two groups: an /f/-trained group and an /s/-trained group. Participants in the /f/-trained group were presented with the 20 critical /f/-final words with the /f/ replaced by the most ambiguous step from the morphed /f/-to-/s/ continuum. The 20 /s/-final words were presented with fricatives in their unambiguous form. Participants in the /s/-trained group were presented with all critical words ending in /s/ with ambiguous sounds and all /f/-final words with fricatives in their unambiguous form. All participants were presented with the same set of 60 filler words and 100 nonwords. Moreover, participants in the /f/-trained group were presented with all words passed through the /s/-minus-/f/ filter (i.e., all filler words, nonwords, and critical items up to the fricatives), and participants in the /s/-trained group were presented with the words passed through the /f/-minus-/s/ filter.

Participants were seated in a soundproof booth wearing Sennheiser (Wedemark, Germany) HD 280–13 headphones, over which the sounds were presented binaurally. Across the 200 trials in the training phase, the critical items, filler words, and nonwords were presented in semirandomized order. The first trials consisted of at least six filler words or nonwords before an /f/- or /s/-final word occurred. Care was taken that critical items did not directly follow one another. Overall, the stimulus lists and experimental setup were identical to those reported in Reinisch et al. (2013).

During every trial, participants were asked to indicate whether the stimulus they heard was an existing Dutch word or not by pressing one of two buttons on a button box. The response options *woord* ("word") and *geen woord* ("nonword") were displayed on the left side and right side of the screen, respectively (each corresponding to a button on the same side). Response options were displayed on the screen until the participant responded. The instructions emphasized speed as well as accuracy of listeners' responses. Nine hundred ms after a response was given, the next trial started automatically. Every 50 trials, participants were allowed to take a self-paced break.

Procedure: Test. The test phase immediately followed the exposure phase. The test phase involved a phonetic-categorization task in which all participants were presented with the same (unfiltered) stimuli. These stimuli consisted of selected four-step continua from the five minimal pairs ending in /f/ or /s/. A trial started with the presentation of the two written words of a minimal pair on the screen. The word ending in /f/ was always displayed on the right. After 500 ms, the audio signal was played. Participants were instructed to indicate which of the two words they heard. Nine hundred ms after their response, the next trial started. The four selected steps of each of the five continua were presented eight times in random order, resulting in a total of 160 trials per participant. Participants were allowed a self-paced break after every 40 trials. The exposure and test phases were implemented with Presentation software (Version 14.9; Neurobehavioral Sys-

Table 1
Overview of Conditions in Experiments 1 and 2 Compared With the No-Filter Control Experiment

Experiment and condition	Participant group	Filter	Critical sound	
			/f/	/s/
1: Reduced ambiguity	/s/-trained	/f/-minus-/s/	Unambiguous	Ambiguous
	/f/-trained	/s/-minus-/f/	Ambiguous	Unambiguous
2: Shifted to opposite	/s/-trained	/s/-minus-/f/	Unambiguous	Ambiguous
	/f/-trained	/f/-minus-/s/	Ambiguous	Unambiguous
No filter: Control	/s/-trained	None	Unambiguous	Ambiguous
	/f/-trained	None	Ambiguous	Unambiguous

Note. Filters were applied to all exposure materials, excluding the critical fricative sounds. No filter = no-filter condition of Reinisch et al. (2013).

Table 2
Auditory Lexical-Decision Performance: Mean Percentages of Correct Responses and Mean Reaction Times From Word Onset

Experiment and condition	Participant group	Words with ambiguous fricatives		Words with unambiguous fricatives		Filler words		Filler nonwords	
		Correct (%)	RT (ms)	Correct (%)	RT (ms)	Correct (%)	RT (ms)	Correct (%)	RT (ms)
1: Reduced ambiguity	/f/-trained	98	960	95	970	93	934	96	1,034
	/s/-trained	96	970	95	976	94	924	95	1,048
2: Shifted to opposite (all)	/f/-trained	68	1,122	95	1,028	92	989	95	1,099
	/s/-trained	57	1,253	94	1,081	94	1,043	94	1,177
2: Shifted to opposite (>50%)	/f/-trained	85	1,088	96	1,023	93	989	94	1,116
	/s/-trained	75	1,266	95	1,094	93	1,057	94	1,204
No filter: Control	/f/-trained	97	1,070	97	1,032	95	1,001	94	1,141
	/s/-trained	96	1,004	97	990	95	956	97	1,078

Note. For Experiment 2, performance is reported for the full set of participants (all) and for the set of participants who accepted more than 50% of the words with ambiguous fricatives as real words (>50%). The no-filter data are from Reinisch et al. (2013). RT = reaction time.

tems, Inc., Berkeley, CA). The whole experiment took approximately 30 min to complete.

Results

Exposure. As in previous studies, we set a criterion that to be included in the analyses, participants had to have accepted at least half of the critical exposure items with an ambiguous sound as words (following, e.g., Norris et al., 2003; Reinisch et al., 2013; Sjerps & McQueen, 2010). No participants had to be excluded. Average percentages of correct responses during exposure are reported in Table 2.

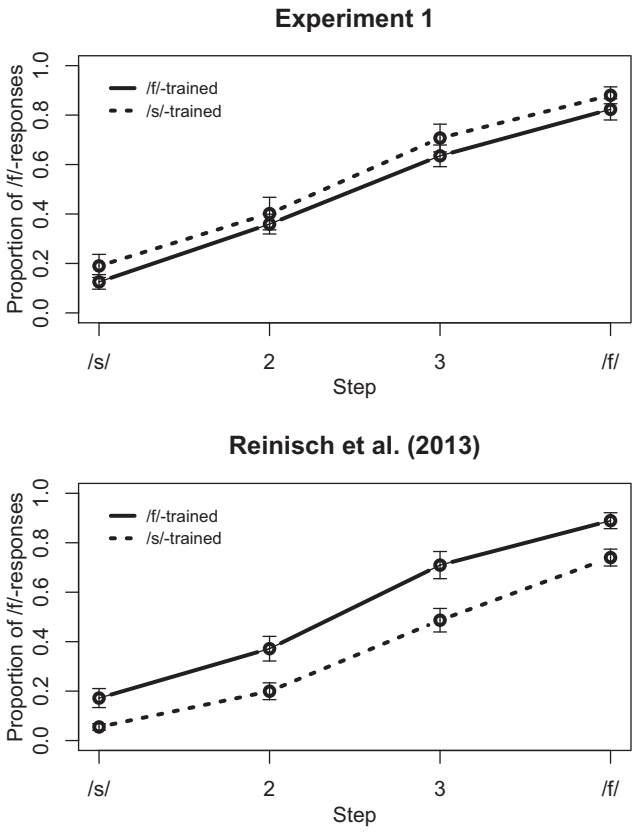

Figure 2. Categorization data for the test continua in Experiment 1 and the no-filter condition in Reinisch et al. (2013). Test stimuli were identical across the experiments. Error bars represent standard errors of the mean.

Test. The results of the phonetic-categorization task in Experiment 1 compared with the no-filter condition are shown in Figure 2. Unlike the no-filter condition, in which the categorization functions for the /s/-trained and /f/-trained groups are clearly different, the functions for the participant groups in Experiment 1 almost overlap (with a numerical trend in the opposite direction than that in the no-filter condition). This suggests that our hypothesis may have been confirmed: Perceptual learning was much reduced when the exposure stimuli were passed through filters that—through perceptual contrast effects—reduced the perceptual ambiguity of the critical fricatives. Statistical analyses confirmed this observation. Analyses were carried out using analyses of variance (ANOVAs) on logit-transformed data to account for the dichotomous dependent variable (/s/ vs. /f/ response; for a discussion of the need for logistic transformation of proportion data, see, e.g., Jaeger, 2008). We entered training (/f/-trained vs. /s/-trained participants) as a between-participants factor and continuum step as a within-participant factor.

For Experiment 1, a single main effect was observed for Continuum, $F(3, 84) = 188.16$, $p < .001$, $\eta_p^2 = .87$, reflecting the fact that stimuli were more often categorized as /f/ toward the /f/ end of the continuum. No main effect was observed for Training, $F(1, 28) = 0.89$, $p = .353$, $\eta_p^2 = .031$, nor was there a Continuum × Training interaction, $F(3, 84) = 0.64$, $p = .592$, $\eta_p^2 = .022$. Hence, there was no evidence for perceptual learning in Experiment 1.

This is in strong contrast with the data in the no-filter condition, in which a significant learning effect was found (as reported by Reinisch et al., 2013).[2] To test whether the training effects in the two experiments (no-filter condition vs. Experiment 1) were statistically different, we ran additional analyses with the factor experiment added. Again, there was a significant main effect of Continuum, $F(3, 168) = 482.40$, $p < .001$, $\eta_p^2 = .896$, indicating

[2] An analysis of the data reported by Reinisch et al. (2013) with the same analytic method used here resulted in the following effects: Training, $F(1, 28) = 10.60$, $p = .003$, $\eta_p^2 = .275$; Continuum, $F(3, 84) = 335.58$, $p < 0.001$, $\eta_p^2 = .923$; and Training × Continuum interaction, $F(3, 84) = 1.03$, $p = .385$, $\eta_p^2 = .035$.

that participants gave more /f/ responses toward the /f/ end of the continuum. The only other significant effect was an Experiment × Training interaction, $F(1, 56) = 8.28$, $p = .006$, $\eta_p^2 = .129$, reflecting the fact that the effect of training (/s/-trained vs. /f/-trained participants) differed across experiments. Nonsignificant results were found for the main effects of Experiment, $F(1, 56) = 2.57$, $p = .115$, $\eta_p^2 = .044$; Training, $F(1, 56) = 2.17$, $p = .147$, $\eta_p^2 = .037$; the Experiment × Continuum interaction, $F(3, 168) = 0.18$, $p = .91$, $\eta_p^2 = .003$; the Training × Continuum interaction, $F(3, 168) = 0.27$, $p = .85$, $\eta_p^2 = .005$; and the Training × Experiment × Continuum interaction, $F(3, 168) = 1.29$, $p = .279$, $\eta_p^2 = .023$.

As can be seen from the separate analyses just discussed, the effect of training was present only in the no-filter control condition. The interaction between Experiment and Training confirms that there was a statistically significant difference between the results of the experiments and, hence, a significant effect of the acoustic filters applied during exposure in Experiment 1.

Discussion

Experiment 1 showed that perceptual learning effects are reduced or absent when listeners are presented with a filtered speech signal that causes acoustically ambiguous fricatives to be perceived as unambiguous—that is, matching the intended lexical option. This is in contrast with the no-filter condition (Reinisch et al., 2013), which used the same fricatives used here (during exposure and test). In the no-filter condition, listeners did shift their category boundaries to accommodate the ambiguous sound in the intended category (see Figure 2). These findings suggest that in the filtered-context condition, the occurrence of perceptual contrast effects prevented lexically guided perceptual learning from occurring. Experiment 1 therefore provides a first insight into how these two processes apply relative to each other. Perceptual contrast effects exert their influence somewhat earlier than perceptual learning.

However, there is at least one alternative explanation for the fact that the perceptual learning effects differed between Experiment 1 and the no-filter condition: the presence of the filter itself (regardless of its nature or perceptual consequences). It has been shown that in cases in which participants can attribute an unnatural pronunciation to an incidental property of the speaker (such as holding a pen in his or her mouth), perceptual learning is blocked (Kraljic et al., 2008). It may be that in the present experiment, all learning was blocked because participants attributed any unnaturalness of the fricatives to the unusual filter properties of the materials.

Therefore, in Experiment 2, we again used filtered materials during exposure, but this time the filters were applied such that the perception of the ambiguous fricatives was pushed in the other direction. That is, the fricatives should be perceived as sounding more like the other category. If the lack of perceptual learning in Experiment 1 was mostly attributable to the filters having made the ambiguous sounds unambiguous, then we would expect to find a learning effect in Experiment 2. This was predicted because the ambiguous sounds would no longer become unambiguous as a result of the filtered precursors. In fact, we might expect to observe an even larger learning effect than in the no-filter condition, because to interpret the words correctly, listeners would have to extend their existing categories relatively far to include the ambiguous sounds in the target categories (because their representations would have become even more unlike the intended sounds). Hence, we could even test the bounds of perceptual learning—that is, whether effects got larger as the critical sounds were perceived to be more like the other alternative. If, however, our filtering manipulation in Experiment 1 blocked learning because listeners attributed any ambiguity to unusual sound properties related to the experimental setting, we would not expect to find an effect in Experiment 2 either.

Experiment 2: Filtering to Shift Sounds Away From the Target Category

Experiment 2 was similar in setup to Experiment 1 and the no-filter condition in Reinisch et al. (2013), with the exception that now the /f/-trained group heard all words passed through the /f/-minus-/s/ filter, which reduced amplitude of the spectral regions that are characteristic of /s/ in the context. As a result, an ambiguous sound in an /f/-biasing lexical context should have sounded more /s/-like and, thus, less like the lexically supported fricative. To accommodate a sound that was rather far from the ideal category in the perceptual space, a large shift in the boundary would be necessary. Hence, if acoustic context information already has a significant influence on representations before lexically guided perceptual learning, we would expect to find a learning effect. One alternative that has to be kept in mind (and which is discussed in more detail later) is the option that, in some cases, the perceptual shift of the fricatives in the opposite-to-intended direction may have been "too far." In such cases, participants may reject the critical exposure items as nonwords, and for those cases, no learning effect should be observed. Overall, however, any learning effect would discard the option that the lack of learning in Experiment 1 was attributable to the filter itself rather than the nature of the filter.

Method

Participants. Thirty native speakers of Dutch were recruited from the same population and according to the same criteria as in the previous experiment. None had participated in Experiment 1. They received a small financial reward for their participation.

Materials and procedure. The materials were again the same as in the no-filter condition, and, hence, the same word set as in Experiment 1 was used. However, now the stimuli used during exposure were filtered with the filters opposite from those in Experiment 1 (see Table 1). That is, the stimuli used for the /f/-trained listener group were now passed through the /f/-minus-/s/ filter, and the stimuli for the /s/-trained group were passed through the /s/-minus-/f/ filter. The /f/-minus-/s/ filter (/f/-trained group) attenuated the frequencies that are characteristic for /s/ in the fillers and the initial parts of the critical words. Therefore, as a result of contrastive context effects, the ambiguous fricatives replacing /f/ should have sounded more like /s/—that is, more ambiguous or even more similar to the wrong category (i.e., the category not supported by the lexical information). The opposite should have held for the /s/-trained listener group, whose stimuli were passed through the /s/-minus-/f/ filter. The minimal-pair continua used for the test phase were the same as in Experiment 1 and remained unfiltered. The experimental procedure was the same as for Experiment 1.

Results

Exposure. The same criterion as in Experiment 1 was used for participants to be included in the analyses (at least 50% of the words with an ambiguous fricative needed to be accepted as real words). In contrast to Experiment 1, nine out of the 30 participants failed to meet this criterion (four in the /f/-ambiguous group). Mean overall percentages correct and reaction times for the full sample of participants and the sample with the nine participants excluded are reported in Table 2. Implications of this finding are discussed later.

Test. Participants who failed the 50% acceptance criterion were excluded from all analyses. Categorization performance of the remaining participants is displayed in Figure 3. It can be observed that, in contrast with Experiment 1, the categorization functions of the /f/-trained and /s/-trained groups are clearly different. Statistical analyses were again performed on logit-transformed data using ANOVAs with Training, Continuum, and their interaction as factors. This analysis resulted in a main effect of Continuum, $F(3, 57) = 81.36$, $p < .001$, $\eta_p^2 = .811$, reflecting reliable use of the acoustic properties of the stimuli along the continuum, and a main effect of Training, $F(1, 19) = 8.24$, $p = .01$, $\eta_p^2 = .302$. As can be seen in Figure 3, listeners in the /f/-trained group gave more /f/ responses than listeners in the /s/-trained group; hence, perceptual learning took place. The Continuum × Training interaction was not significant, $F(3, 57) = 1.02$, $p = .389$, $\eta_p^2 = .051$.

Because an effect of training was found in Experiment 2 but not in Experiment 1, we ran an additional analysis to test whether the effect of training statistically differed between experiments. Therefore, we included the factor experiment (Experiment 1 vs. Experiment 2) in our analysis. This analysis showed a main effect of Continuum, $F(3, 141) = 262.30$, $p < .001$, $\eta_p^2 = .848$ (again reflecting reliable categorization performance), and, critically, an Experiment × Training interaction, $F(1, 47) = 9.61$, $p = .003$, $\eta_p^2 = .17$. The inverted-filtering manipulation in Experiment 2 resulted in a significant increase in the learning effect compared with Experiment 1. Nonsignificant results were found for the main effects of Experiment, $F(1, 47) = 0.06$, $p = .808$, $\eta_p^2 = .001$; Training, $F(1, 47) = 2.32$, $p = .134$, $\eta_p^2 = .047$; the Experiment × Continuum interaction, $F(3, 141) = 1.40$, $p = .245$, $\eta_p^2 = .029$; the Training × Continuum interaction, $F(3, 141) = 1.52$, $p = .213$, $\eta_p^2 = .031$; and the Experiment × Training × Continuum interaction, $F(3, 141) = 0.22$, $p = .884$, $\eta_p^2 = .005$.

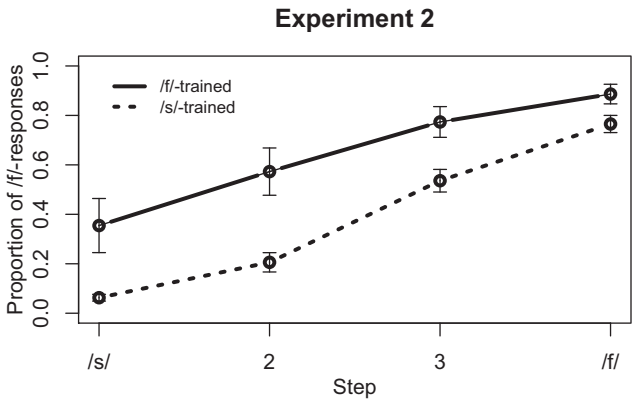

Figure 3. Categorization data for the test continuum in Experiment 2. Error bars represent standard errors of the mean.

Comparing the categorization data for the no-filter condition in Reinisch et al. (2013; see the bottom panel of Figure 2) with those of Experiment 2 (see Figure 3), it can be observed that Experiment 2 led to a numerically larger learning effect. Analyses were performed to test this pattern. These revealed main effects for the factors Continuum, $F(3, 141) = 344.89$, $p < .001$, $\eta_p^2 = .88$, and Training, $F(1, 47) = 18.36$, $p < .001$, $\eta_p^2 = .281$. No main effect was observed for Experiment, $F(1, 47) = 2.09$, $p = .155$, $\eta_p^2 = .043$. Critically, no Experiment × Training interaction was observed, $F(1, 47) = 0.99$, $p = .324$, $\eta_p^2 = .021$, indicating that, although the training effect was numerically larger in Experiment 2 than in the no-filter condition, the increase was only numerical. In addition, no Experiment × Continuum interaction, $F(3, 141) = 2.41$, $p = .069$, $\eta_p^2 = .049$; Training × Continuum interaction, $F(3, 141) = 0.26$, $p = .857$, $\eta_p^2 = .005$; or Experiment × Training × Continuum interaction, $F(3, 141) = 1.95$, $p = .125$, $\eta_p^2 = .04$, was observed.

Discussion

Experiment 2 tested perceptual learning in a condition in which the acoustic context surrounding the ambiguous fricatives should have caused the fricatives to be perceived as more ambiguous, or even closer to the other endpoint on the /f/-to-/s/ continuum, than the lexically supported category. In contrast with Experiment 1, here we did find a learning effect, and this effect was statistically different from the finding of Experiment 1. This suggests that the lack of learning in Experiment 1 cannot be explained by the filtering per se but, rather, must have been a result of the nature of the filter. In Experiment 2, we also expected to find an increased learning effect relative to the no-filter condition. We reasoned that an ambiguous sound that was far away from the lexically supported category would lead to a stronger shift in the category boundary. However, the training effect in Experiment 2 relative to the no-filter condition was only numerically larger, not statistically so. This shows that there is an upper limit to the magnitude of the learning effect.

Such a limit seems reasonable given the fact that in extreme cases, acoustic context could have shifted the perception of the ambiguous sound across the natural category boundary toward the wrong interpretation. This would lead to perception of the critical training words as nonwords. Given that nonwords do not provide lexical information about the interpretation of the critical sound (Eisner & McQueen, 2005; Norris et al., 2003; unless there are other sources of information such as phonotactics [see Cutler, McQueen, Butterfield, & Norris, 2008]), learning may not occur. The rather large number of participants failing our 50% inclusion criterion supports this interpretation.

We carried out additional analyses to test whether, indeed, there may be a relation between the acceptance of critical words during exposure and the location of the category boundary at test. If our just-posited interpretation is correct, we would predict that the more words a participant accepted in the training phase, the larger the shift in category boundary in the test phase. The relation between the proportion of critical words that were accepted in the training phase and the proportion of /f/ responses in the test phase is shown in Figure 4. Consider, first, the participants in the /f/-trained group. It can be observed that those participants who accepted many critical items during training also gave many /f/ responses at test. That is, these participants indeed expanded their /f/ category. However, those

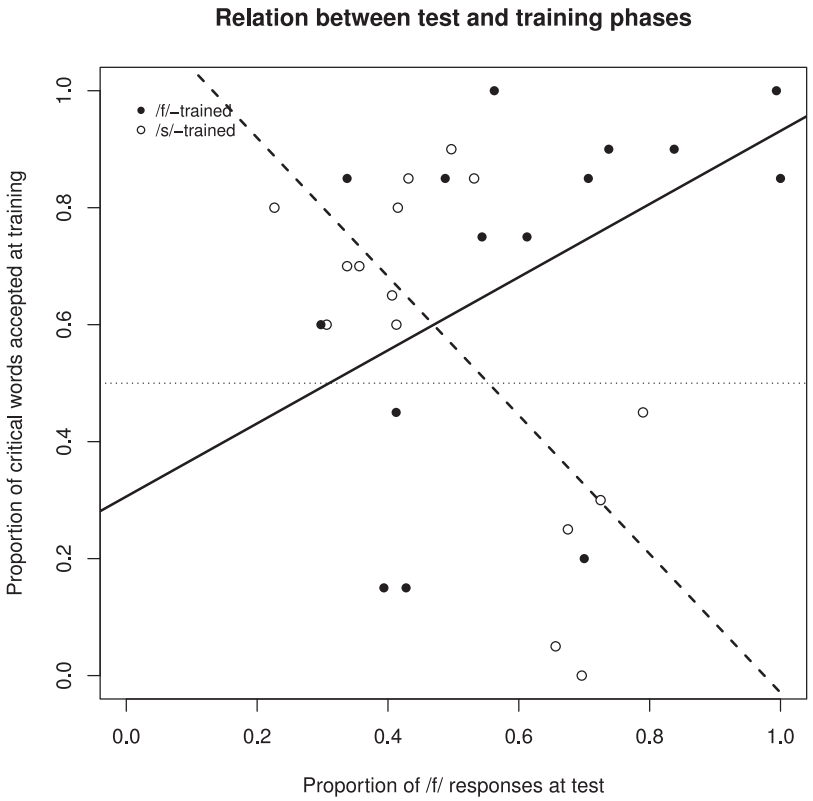

Figure 4. Dot plot displaying the relation between the acceptance rates for the critical training items in the training phase of Experiment 2 and the proportion of /f/ responses that an individual gave in the test phase. Points represent an individual participant's combined scores. Lines are fitted to these data (solid line = /f/-trained; dashed line = /s/-trained). The dashed horizontal line at 0.5 indicates the criterion value used to exclude participants from analysis.

participants who did not accept most critical items at training (see the data points below the 50% criterion indicated by the dotted horizontal line) tended to give fewer /f/ responses at test. The regression line reflects this pattern for the /f/-trained group as it has a positive slope. In contrast, for participants in the /s/-trained group, the pattern was reversed. As expected, those participants who accepted the majority of the critical items during training gave few /f/ responses at test, indicating that these participants learned to expand their /s/ category through exposure. Those participants who rejected the majority of the critical training items gave relatively more /f/ responses at test than did those participants who accepted most critical training items. These two patterns show that the size of the training effect was dependent on the proportion of critical items accepted during the training phase.

A linear regression analysis confirmed these patterns. The dependent variable was the per-participant proportion of *yes* responses to critical items in the training phase. The independent variables were training (/f/-trained vs. /s/-trained participants) and boundary. *Boundary* was defined as the per-participant proportion of /f/ responses across the continua in the test phase.

For the regression analysis, the /f/-trained group was assigned the reference level for training. Therefore, main effects for the other factors reflected patterns for the /f/-trained group only. Interactions between an effect with the factor training indicated how the /f/- and /s/-trained groups differed for that factor. The regression analysis ($R^2 = 0.37$, $F(3, 26) = 5.15$, $p = 0.006$) revealed a small main effect for category boundary—($b_{\text{boundary}} = 0.63$), $t = 2.09$, $p = .046$; ($b_{\text{intercept}} = 0.31$), $t = 1.60$, $p = .12$—indicating a positive relation between the proportion of /f/ responses in the test phase and the proportion of critical items accepted during the training phase for the /f/-trained group. Critically, there was a Boundary × Training interaction ($b_{\text{Boundary}\times\text{Training}} = -1.81$), $t = -3.71$, $p = .001$, indicating that the relation between the number of /f/ responses in the test phase and acceptance in the training phase differed between the two training groups. That is, although increased acceptance of critical items led to more /f/ responses at test for the /f/-trained group, it led to fewer /f/ responses for the /s/-trained group. A main effect was observed for Training ($b_{\text{training}} = 0.85$), $t = 3.05$, $p = .005$, indicating that the intercept for the /s/-trained group was positioned higher.

These additional analyses for the data of Experiment 2 show that there is indeed an upper limit to the magnitude of the learning effect. With an increased distance between the target category and the ambiguous signal, there is a point at which listeners fail to accept a token as a word, and hence the lexicon cannot guide perceptual learning.

General Discussion

In two experiments, we investigated the combined operation of two processes that are known to be used in dealing with variation in speech perception: perceptual contrast effects and lexically guided perceptual learning. We found evidence that, in line with predictions from previous literature, at least a portion of perceptual contrast effects apply before lexically guided perceptual learning. This study thus starts to expand our understanding of how listeners deal with multiple sources of information during speech processing.

The application of both processes was tested within a single experimental setup. Conditions for perceptual contrast effects in the form of acoustic context manipulations were added to a perceptual learning experiment in which lexical information was expected to guide phonetic category retuning. Critically, in Experiment 1, acoustic and lexical context were expected to shift the perception of ambiguous fricatives in the same direction. The logic was that if the perceptual contrast effects (here achieved through filtering of the context words) precede the application of perceptual learning, this should result in no or only minimal remapping of the phonetic categories. In line with this prediction, the data of Experiment 1 showed no effect of perceptual learning but, in fact, a numeric difference in the opposite direction.[3]

The purpose of Experiment 2 was twofold. The first purpose was to provide a proof-of-principle replication by changing the effect of perceptual contrast in the other direction. That is, during training, perceptual contrast effects were predicted to induce a perceptual shift of the target fricatives away from the lexically supported target category. Significant learning effects were observed at test, and these were significantly different from those of Experiment 1. This suggests that in Experiment 2, because of perceptual contrast effects, participants had to remap their /f/ and /s/ categories for perceptual learning to a greater extent than in Experiment 1. Thus, the direction of the filters did indeed matter.

The second motivation for Experiment 2 was to control for potential alternative explanations for the pattern observed in Experiment 1. First, the lack of learning in Experiment 1 could have been a result of the fact that participants regarded the ambiguity of the final fricatives merely as a circumstantial aspect of the situation, here due to the filtering. Kraljic et al. (2008) have shown that if the ambiguity of critical sounds can be attributed to external circumstances, such as the speaker putting a pen in his or her mouth while articulating critical words, perceptual learning does not occur. Here, the filters could have served as the external circumstance (e.g., the speaker was located in a room with unusual room acoustics). These explanations were disproved, because a learning effect was found in Experiment 2, in which the same filters were applied to the training materials as were applied in Experiment 1, the only difference being that the acoustic and lexical contexts now supported the opposite sound category. A way to reconcile the present data with findings such as Kraljic et al.'s is to look at the issue of external evidence from a slightly different angle. If the lack of learning in Kraljic et al. is explained such that the ambiguity has been taken care of through the attribution to the pen—thus making category remapping superfluous, because it is not a property of the speaker—then one could say that in Experiment 1, the ambiguity of the fricatives was taken care of by the acoustic context, which shifted the fricatives perceptually toward the intended unambiguous category. In this case, the acoustic context in the present study would not be circumstantial evidence but just another factor that took care of the critical sounds' ambiguity, reducing the amount of lexically guided perceptual learning.[4]

Cognitive Implementation

In Experiment 1, perceptual learning only occurred if perceptual contrast effects did not already take care of the critical sounds' ambiguity. Therefore, our results support the suggestion that perceptual contrast effects at least partially preceded perceptual learning. As argued in the introduction, however, this order could be implemented in at least two different ways. First, perceptual contrast effects could precede both the lexical level and the prelexical remappings (i.e., the locus of retuning) in perceptual learning. On a particular training trial, the input signal would then first be transformed through contrast effects before information could reach the levels of representation involved in lexically guided perceptual learning. For Experiment 1, this transformation would have led to an unambiguous input at the level at which retuning is implemented. This signal would then have been mapped onto the lexically supported phoneme category, and any resulting lexical feedback would have led to only minimal changes to the input distribution associated with that phoneme. As discussed in the introduction, this cognitive ordering aligns with previous attempts to model these effects in the framework of the interactive activation model TRACE. In that model, perceptual contrast effects affect a feature level, whereas perceptual learning is implemented in the connections between features and phoneme representations (McClelland et al., 2006; see the Appendix). Shortlist B (Norris & McQueen, 2008), despite its lack of explicit description of how to deal with perceptual contrast effects, could implement the present findings in a similar way (then using the long-term feedback for learning rather than online lexical feedback).

The second account put forward in the introduction assumed that perceptual contrast effects and retuning in lexically guided perceptual learning operate at the same level of processing. Perceptual contrast effects would then precede the lexical level but not necessarily the locus of prelexical remappings triggered by perceptual learning. To exemplify this, in Experiment 1, on any single trial during exposure (i.e., the lexical-decision task), contrast effects would have shifted the perceptual representation toward the lexically supported alternative. This would have prevented a lexical mismatch, in turn preventing an error signal from being sent

[3] The possibility of an opposite learning effect in fact also follows from our manipulation. Although the focus was on the ambiguous target fricatives in this design, the context effects also operated on the unambiguous fricatives. Consider the manipulation in the /f/-trained condition of Experiment 1. The /f/-minus-/s/ filter made the ambiguous fricative sound perceptually more /s/-like, reducing the size of the boundary shift for /s/. However, the filter also made the /f/ sound perceptually more like /s/, moving the sound perceptually into an ambiguous region. This could, then, have induced an extension of a participant's /f/ category.

[4] This is not to say that contrast effects and the effect reported by Kraljic et al. (2008) are implemented at the same level of processing.

from the lexicon to the phoneme representations. Then, although, in principle, the prelexical processing would have had access to the perceptually ambiguous fricative (i.e., an "untransformed" representation), no error signal would have been sent because it would have already been blocked by the contrast effects. Therefore, perceptual learning could not have associated the ambiguous sound with the lexically supported category. In this way, contrast effects and retuning in perceptual learning could operate at the same cognitive level, but contrast effects would, in terms of their temporal relation, have to apply (or end) their effects slightly earlier.

Although the current project cannot ultimately distinguish between these potential implementations (locus and timing vs. timing only), Bayesian models such as the belief updating model (Kleinschmidt & Jaeger, 2012) provide an additional angle on these two hypotheses. This model was specifically designed to capture the workings of both perceptual learning and a type of contrast effect that is similar in nature to the contrast effects under investigation here—namely, selective adaptation (Samuel, 1986). According to this model, both effects occur because repeated exposure to a particular realization of a sound category results in a change in the expected cue distribution for the category and, hence, its interpretation on future encounters. This shows that phoneme distributions are continuously updated to optimally reflect input distributions (see, e.g., Kleinschmidt & Jaeger, 2012). Feedback for learning, therefore, is likely to be a continuous process that operates regardless of the size or occurrence of an outright mismatch at the lexical level. Such a continuous updating mechanism aligns more closely with the locus and timing hypothesis than with timing only, because only the latter assumes a dichotomous lexical-mismatch error signal. The Bayesian belief updating model does not aim to address the mechanistic implementation of the processes under investigation here. However, the continuous updating of category representations appears to favor a cognitive order for our two processes that is not just a difference in its temporal relation (i.e., one that is not dependent on an all-or-nothing distinction).

A further important note is that we do not expect a complete division between contrast effects and perceptual learning. It is most plausible that different processes begin as soon as they can and need not be finished before the next process begins (i.e., processing in a cascading fashion). In addition, those contrastive processes that were investigated here are only part of the total set of contrastive processes that operate throughout the processing stream in speech perception. Several researchers have argued that contrastive effects in speech perception may arise at a number of levels in the processing hierarchy. Effects such as forward masking are known to arise in the periphery of the auditory system (Summerfield, Haggard, Foster, & Gray, 1984; Wilson, 1970), and a body of research has demonstrated that there are also contextual influences that occur at later processing stages than in the auditory periphery, because they occur with longer precursor–target intervals and with contralateral presentation (Holt, 2005; Holt & Lotto, 2002; Sjerps, Mitterer, & McQueen, 2011b; Sjerps et al., 2012). In addition, there is evidence that higher level (language-specific) context effects also play an important role in speech perception (Sjerps et al., 2011a, 2012; Viswanathan et al., 2009, 2010). Therefore, the current research only describes how an important subpart of these contrast effects precedes lexically guided perceptual learning.

An interesting final aspect of the results presented here is that the difference between perceptual learning and contrast effects allows them to divide the workload in dealing with different sources of variation. Because perceptual contrast effects precede perceptual learning, they manage to take care of any signal differences that are reflected as predictable overall changes in the long-term average speech spectrum. More specific sources of variation, such as lisping, or more generally the variance that affects the production of individual sounds in a specific way, are left unchanged so that learning can apply to accommodate those sources of variation.

The current research has demonstrated how two different processes in speech perception cooperate to compensate for different types of variation. Through the exploration of different types of effects within the same paradigm, we mapped out how lexically guided perceptual learning and perceptual contrast effects from acoustic context operate relative to each other. Through explicit testing of this cognitive ordering for the first time, it was shown that perceptual contrast effects have to at least partially precede lexically guided perceptual learning.

References

Apfelbaum, K. S., & McMurray, B. (2014). Relative cue encoding in the context of sophisticated models of categorization: Separating information from categorization. *Psychonomic Bulletin & Review*. Advance online publication. http://dx.doi.org/10.3758/s13423-014-0783-2

Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (Version 5.1) [Computer program]. Retrieved from http://www.praat.org

Cant, N. B. (1992). The cochlear nucleus: Neuronal types and their synaptic organization. In D. B. Webster, A. N. Popper, & R. R. Fay (Eds.), *The mammalian auditory pathway: Neuroanatomy* (pp. 66–116). New York: Springer-Verlag. http://dx.doi.org/10.1007/978-1-4612-4416-5_3

Clarke-Davidson, C. M., Luce, P. A., & Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception & Psychophysics, 70,* 604–618. http://dx.doi.org/10.3758/PP.70.4.604

Cutler, A., McQueen, J. M., Butterfield, S., & Norris, D. (2008, September). *Prelexically-driven perceptual retuning of phoneme boundaries*. Paper presented at the Ninth Annual Conference of the International Speech Communication Association, Brisbane, Queensland, Australia.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67,* 224–238. http://dx.doi.org/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. http://dx.doi.org/10.1121/1.2178721

Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science, 16,* 305–312. http://dx.doi.org/10.1111/j.0956-7976.2005.01532.x

Holt, L. L. (2006). Speech categorization in context: Joint effects of nonspeech and speech precursors. *Journal of the Acoustical Society of America, 119,* 4016–4026. http://dx.doi.org/10.1121/1.2195119

Holt, L. L., & Lotto, A. J. (2002). Behavioral examinations of the level of auditory processing of speech context effects. *Hearing Research, 167,* 156–169. http://dx.doi.org/10.1016/S0378-5955(02)00383-0

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America, 108,* 710–722. http://dx.doi.org/10.1121/1.429604

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. http://dx.doi.org/10.1016/j.jml.2007.11.007

Jesse, A., & McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic Bulletin & Review, 18,* 943–950. http://dx.doi.org/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: Possible role of a repetitive structure in sounds. *Speech Communication, 27,* 187–207. http://dx.doi.org/10.1016/S0167-6393(98)00085-5

Kleinschmidt, D., & Jaeger, T. F. (2012). A continuum of phonetic adaptation: Evaluating an incremental belief-updating model of recalibration and selective adaptation. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Building bridges across cognitive sciences around the world: Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 605–610). Austin, TX: Cognitive Science Society.

Kluender, K. R., Coady, J. A., & Kiefte, M. (2003). Sensitivity to change in perception of speech. *Speech Communication, 41,* 59–69. http://dx.doi.org/10.1016/S0167-6393(02)00093-6

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

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science, 19,* 332–338. http://dx.doi.org/10.1111/j.1467-9280.2008.02090.x

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research, 25,* 161–168. http://dx.doi.org/10.1016/j.cogbrainres.2005.05.004

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29,* 98–104. http://dx.doi.org/10.1121/1.1908694

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

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences, 10,* 363–369. http://dx.doi.org/10.1016/j.tics.2006.06.007

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

Mirman, D., McClelland, J. L., & Holt, L. L. (2006). An interactive Hebbian account of lexically guided tuning of speech perception. *Psychonomic Bulletin & Review, 13,* 958–965. http://dx.doi.org/10.3758/BF03213909

Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition, 129,* 356–361. http://dx.doi.org/10.1016/j.cognition.2013.07.011

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. http://dx.doi.org/10.1016/j.wocn.2008.09.001

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115,* 357–395. http://dx.doi.org/10.1037/0033-295X.115.2.357

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

Poellmann, K., Bosker, H. R., McQueen, J. M., & Mitterer, H. (2014). Perceptual adaptation to segmental and syllabic reductions in continuous spoken Dutch. *Journal of Phonetics, 46,* 101–127. http://dx.doi.org/10.1016/j.wocn.2014.06.004

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.

Reinisch, E., & Holt, L. L. (2014). Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance, 40,* 539–555. http://dx.doi.org/10.1037/a0034409

Reinisch, E., & Sjerps, M. J. (2013). The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. *Journal of Phonetics, 41,* 101–116. http://dx.doi.org/10.1016/j.wocn.2013.01.002

Reinisch, E., Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 75–86. http://dx.doi.org/10.1037/a0027979

Reinisch, E., Wozny, D. R., Mitterer, H., & Holt, L. L. (2014). Phonetic category recalibration: What are the categories? *Journal of Phonetics, 45,* 91–105. http://dx.doi.org/10.1016/j.wocn.2014.04.002

Samuel, A. G. (1986). Red herring detectors and speech perception: In defense of selective adaptation. *Cognitive Psychology, 18,* 452–499. http://dx.doi.org/10.1016/0010-0285(86)90007-1

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

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

Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011a). Constraints on the processes responsible for the extrinsic normalization of vowels. *Attention, Perception, & Psychophysics, 73,* 1195–1215. http://dx.doi.org/10.3758/s13414-011-0096-8

Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011b). Listening to different speakers: On the time-course of perceptual compensation for vocal-tract characteristics. *Neuropsychologia, 49,* 3831–3846. http://dx.doi.org/10.1016/j.neuropsychologia.2011.09.044

Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2012). Hemispheric differences in the effects of context on vowel perception. *Brain and Language, 120,* 401–405. http://dx.doi.org/10.1016/j.bandl.2011.12.012

Sjerps, M. J., & Smiljanić, R. (2013). Compensation for vocal tract characteristics across native and non-native languages. *Journal of Phonetics, 41,* 145–155. http://dx.doi.org/10.1016/j.wocn.2013.01.005

Skoe, E., Krizman, J., Spitzer, E., & Kraus, N. (2013). The auditory brainstem is a barometer of rapid auditory learning. *Neuroscience, 243,* 104–114. http://dx.doi.org/10.1016/j.neuroscience.2013.03.009

Stilp, C. E., Alexander, J. M., Kiefte, M., & Kluender, K. R. (2010). Auditory color constancy: Calibration to reliable spectral properties across nonspeech context and targets. *Attention, Perception, & Psychophysics, 72,* 470–480. http://dx.doi.org/10.3758/APP.72.2.470

Summerfield, Q., Haggard, M., Foster, J., & Gray, S. (1984). Perceiving vowels from uniform spectra: Phonetic exploration of an auditory aftereffect. *Perception & Psychophysics, 35,* 203–213. http://dx.doi.org/10.3758/BF03205933

Viswanathan, N., Fowler, C. A., & Magnuson, J. S. (2009). A critical examination of the spectral contrast account of compensation for coarticulation. *Psychonomic Bulletin & Review, 16,* 74–79. http://dx.doi.org/10.3758/PBR.16.1.74

Viswanathan, N., Magnuson, J. S., & Fowler, C. A. (2010). Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1005–1015. http://dx.doi.org/10.1037/a0018391

Watkins, A. J. (1991). Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion. *Journal of the Acoustical Society of America, 90,* 2942–2955. http://dx.doi.org/10.1121/1.401769

Watkins, A. J., & Makin, S. J. (1994). Perceptual compensation for speaker differences and for spectral-envelope distortion. *Journal of the Acoustical Society of America, 96,* 1263–1282. http://dx.doi.org/10.1121/1.410275

Wilson, J. P. (1970). An auditory after-image. In R. Plomp & G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 303–318). Leiden, The Netherlands: Sijthoff.

Appendix

TRACE

Although computational modeling is not the focus of the present article, a discussion of our design in light of such a model (in this case, TRACE; McClelland & Elman, 1986) may help to lay out our predictions in more detail and clarify the results. TRACE is an interactive activation model that consists of three layers: an acoustic/articulatory feature layer, a phoneme layer, and a lexical layer, each connected with interactive, excitatory between-layer connections. Same-layer connections are inhibitory. The lexical layer, hence, enhances lexically consistent phoneme interpretations, which in turn decrease activation of lexically inconsistent phonemes through lateral inhibition. On the basis of this architecture, a version of TRACE has been established for modeling perceptual learning data by adding a Hebbian learning algorithm to account for long-term phoneme adjustments (Hebb-TRACE; Mirman et al., 2006). The addition of a Hebbian learning algorithm ensures that connection weights between different units are continuously updated such that an ambiguous feature input is associated with the lexically consistent phoneme. Hebb-TRACE thus assumes that the retuning in perceptual learning occurs in the connections between low-level feature representations and phonemic representations.

Importantly, perceptual contrast effects have also been discussed in TRACE (McClelland et al., 2006). They can be accounted for by allowing lateral interactions across time slices within the feature level. In this architecture, perceptual contrast effects may thus precede the locus of retuning, as indeed is argued for in the current article as a potential implementation. Consider a case in which, during exposure, listeners hear an acoustically ambiguous fricative, [$^{s}_{f}$], that replaces an intended /s/ at the end of an /f/-minus-/s/ filtered word (as in Experiment 1). The recent feature activity (earlier time slices) would suppress features that are specific to /f/, because the history would be more /f/-like than /s/-like. The acoustic context would then give /s/ an advantage over its competitor, /f/. At the same time, lexical information is likely to have become available that favors the lexically consistent interpretation of /s/. The combination of acoustic and lexical information will point the listener toward recognizing /s/ and accepting the word as a real word in the lexical-decision task. The Hebbian learning algorithm, however, will cause minimal changes to the feature-to-phoneme weights, because the level of activity of the feature units will already be in line with those at the phoneme unit level. That is, little to no learning should occur. In contrast, in cases in which information from the feature level mismatches the lexically supported phoneme (as in Experiment 2), the feature-to-phoneme connections should gradually be tuned toward the lexically consistent alternative, resulting in perceptual learning.

Received November 11, 2013
Revision received January 13, 2015
Accepted February 3, 2015 ■