Divide and conquer: how perceptual contrast sensitivity and perceptual learning cooperate in reducing input variation in speech perception

Matthias J. Sjerps[1] & Eva Reinisch[2]

[1]Max Planck Institute for Psycholinguistics, m.j.sjerps@gmail.com

[2]Institute of Phonetics and Speech Processing, Ludwig Maximilian University Munich, evarei@phonetik.uni-muenchen.de

running head: contrast and learning in speech perception

corresponding author:

Matthias J Sjerps, PhD

Department of Linguistics,

University of California Berkeley,

Berkeley, California, USA

Phone: +1-5106465194

Email: m.j.sjerps@gmail.com

Abstract

Listeners have to overcome variability of the speech signal that can arise, for example, due to 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.

Words: 161

Introduction

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 which 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, et al., 2003); "Perceptual contrast effects" have been argued to help 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 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 where listeners take into account multiple sources of variation at a time.

Over the last decades experimental evidence has been accumulated which suggests 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, to be reviewed in the following sections, along with modelling approaches (see Appendix A), 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 where 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 (Broadbent & Ladefoged, 1960; 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., Broadbent & Ladefoged 1960; Reinisch & Sjerps, 2013; Sjerps, Mitterer & McQueen, 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 preceding sound (Watkins, 1991). That is, if an ambiguous sound between /f/ and /s/ (here [$^{s}_{f}$]; we use the notation [$^{s}_{f}$] 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 as 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 et al, 2003, and references therein).

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 non-speech 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 see Viswanathan, Fowler, & Magnuson, 2009; Viswanathan, Magnuson & Fowler, 2010, for a discussion of speech-specific contributions in the closely related domain of compensation for coarticulation; see also Holt & Lotto, 2002; Holt, et al., 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 (e.g., Sjerps, Mitterer & McQueen, 2012, and references therein) that takes place at the level of the brainstem (Cant, 1992). As for the timecourse of perceptual contrast effects, they result from the immediate acoustic context and are observable on 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 perception 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, McQueen, & Cutler, 2003; Reinisch, Weber, & Mitterer, 2013; Sjerps & McQueen, 2010; Samuel & Kraljic, 2009, for an overview). 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/ into their /f/ category (Clarke-Davidson, Luce, & Sawusch, 2008; Kraljic & Samuel, 2005; Norris, et al., 2003). Notably, the same ambiguous sound can also be learnt to be interpreted as an instance of /s/ if the ambiguous sound occurs in words where 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; see Mitterer, Scharenborg, & McQueen, 2013; Poellmann, Bosker, McQueen, & Mitterer, 2014; and Reinisch, Wozny, Mitterer & Holt, 2014, for discussions of the units that are affected by perceptual learning). 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 do not apply across the board but rather to new items from the same speaker or tokens from speakers that produce highly similar fricative tokens (Eisner & McQueen, 2005; Kraljic & Samuel, 2005; Reinisch & Holt, 2014). Moreover, on some occasions perceptual learning

appears 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 timecourse are involved in perceptual learning. First, on each encounter of an ambiguous sound [$^{s}_{f}$] in a word like gira[$^{s}_{f}$] 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 to prelexical level (as suggested in interactive models of speech perception, e.g., TRACE, McClelland & Elman, 1986) or it affects a decision stage where prelexical and lexical information is merged (as in feedforward models such as Shortlist B, Norris & McQueen, 2008, or the implementation of this process in Merge, Norris, McQueen, & Cutler, 2000). The second type of timecourse relates to the actual "long-term" retuning during perceptual learning. It has been shown that about 10 to 20 instances of the ambiguous sound in unambiguous context have to be experienced in order to influence the interpretation in lexically ambiguous contexts (Kraljic et al. 2008; Poellmann, McQueen, & Mitterer, 2011). This is what feedforward models of speech perception would call feedback for learning (Norris & McQueen, 2008). Thereby, on encountering pronunciations like "gira[$^{s}_{f}$]" the weights or expectations associating incoming ambiguous input with one or the other segmental interpretation are gradually shifted towards the lexically supported category (note that this long-term adjustment also holds for interactive models like TRACE). This timecourse, spanning multiple encounters of learning contexts at the experiment level, and its dependence on lexical activation at the trial level are a crucial difference to perceptual contrast effects (see Appendix A for details).

While 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, and references therein). This provides strong evidence that linguistic experience, or learning, can influence processes at relatively early physiological levels of processing. Since perceptual learning hence appears not restricted to one cognitive level, the present study sets out to assess the relation of lexically-guided perceptual learning to perceptual contrast effects.

The current project

The research reviewed above 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 the hierarchy of neuronal populations that display 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 where variation occurs due to 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 where perceptual learning is implemented. This interpretation is, in fact, fully in line with modelling approaches that describe these processes within the framework of TRACE (e.g., McClelland & Elman, 1986; see Appendix A 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

[1] The term "processing stage" is not meant to suggest a strict division/temporal ordering between processes - indeed there is likely to be some overlap; see below for details on what accounts could be predicted.

straightforwardly modelled at the featural level (McClelland et al., 2006; see Apfelbaum & McMurray, 2014, for additional modelling-based evidence in favor of a low-level implementation of contrast effects) while the retuning in lexically-guided perceptual learning could best be modelled at the level of connection weights mapping from feature to phoneme units (Mirman et al., 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 by 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 by assuming *only* a difference in timing. These two possible implementations will be 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 its timecourse. Although we have presented evidence for this assumption above, 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, 2005) makes alternative implementations plausible. Testing this assumption directly would therefore be useful for future modelling attempts.

The present study consists 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 will be referred to as the no-filter condition. These stimuli were chosen as they have been shown to elicit strong learning effects. This allowed for a comparison of effect size to the present study. All exposure stimuli except for the critical fricatives were filtered. Filtering provides the acoustic context expected to shift the perception of the ambiguous fricatives in a spectrally contrastive manner (i.e., eliciting 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 due to filter properties before lexically-guided retuning can trigger learning, then 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 in Experiment 1 could have been due to the procedure of filtering itself rather than the nature of the filter. Furthermore, applying acoustic filters that shift perception even more towards the other alternative will 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 allow us to determine to what extent perceptual contrast effects and lexically-guided perceptual learning are, at least partially, in a temporal order relation to resolve different parts of input variation.

Experiment 1: Filtering to reduce ambiguity

Based on the study by Reinisch et al. (2013) which contributes the "no-filter" condition, lexically-guided perceptual learning was tested in a between-group design in which one group of listeners heard an ambiguous sound in final position for words that normally end with /f/ (the /f/-trained group), while another group of participants heard an ambiguous sound in final position for words that normally end with /s/ (the /s/-trained group). In Experiment 1, for the /s/-trained listeners 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 make 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 suppresses frequency regions characteristic of /f/, so that the ambiguous sound becomes perceptually more /f/-like. We predicted that if perceptual contrast effects deal with such changes in general filtering properties first, these manipulations should cause the ambiguous sounds to no longer be perceived as (fully) ambiguous. As a result, the lexically guided updating of phoneme categories should 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 should 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 is associated with occurrences of a particular phoneme).

Methods

Participants. 30 native speakers of Dutch were recruited from the Max Planck

Institute 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 (for details see below). We briefly summarize the stimulus set and construction of ambiguous fricatives in the no-filter condition but refer the reader to the original paper for a more detailed description.

One hundred Dutch words and 100 non-words 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 are non-words if the fricatives are exchanged (*locomotie*[s] and *geitenkaa*[f] are non-words in Dutch). None of the words or non-words contained the sounds /f/, /s/, or their voiced counterparts /v/ and /z/, except for 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"; *kuif– kuis*, "tuft of hair", "chaste"). All stimuli were recorded by a female Dutch native speaker (aged 28) 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 (The MathWorks Inc.). 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, (for example frication noise with frication noise, and 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 (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 5 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 was 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 LTAS 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 minus 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 multiplied by -1).

When a speech signal is now 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 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.

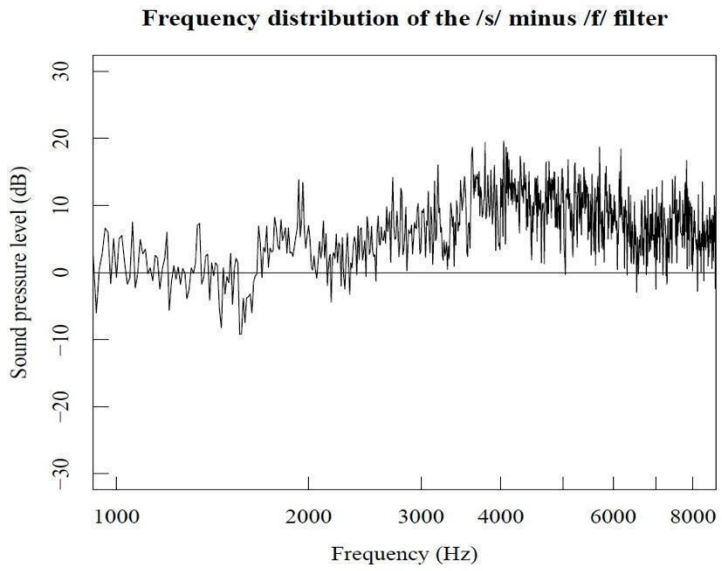

Figure 1: Filter properties of the /f/ minus /s/ filter.

These filters were applied to all words and non-words used for exposure. Of the critical training words only the part up to but excluding the fricatives was filtered (the fricatives should 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 below) 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 as well as to 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.

Table 1

Overview of conditions in Experiment 1 and 2 compared to the no-filter control experiment. Note that filters were applied to all exposure materials excluding the critical fricative sounds.

Experiment	Participant group	Filter	Critical sounds	
			/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

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 non-words. Moreover participants in the /f/-trained group were presented with all words passed through the /s/ minus /f/ filter (i.e., all filler words, non-words, 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 sound-proof booth wearing Sennheiser 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 the non-words were presented in semi randomized order. The first trials consisted of at least six filler words or non-words 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* "non-word" were displayed on the left 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 instruction emphasized speed as well as accuracy of listeners' responses. 900 ms after a response was given the next trial started automatically. Every 50 trials participants were allowed to take a self-paced break.

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 4-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. 900 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. Exposure and test phase were implemented with Presentation software (Version 14.9, Neurobehavioural Systems Inc.). The whole experiment took approximately 30 minutes to complete.

Results

Exposure. As in previous studies, we set the criterion that in order to be included in the analyses participants must 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). None of the participants had to be excluded. Table 2 reports average percentage correct responses during exposure.

Table 2

Auditory lexical decision performance: mean percentage correct responses and mean reaction time in milliseconds from word offset (in brackets). Data are summarized from Experiment 1 and 2 as well as the no-filter condition reported Reinisch et al. (2013). For Experiment 2 we report the performance separately 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%"). Data are split according to participants in the /f/-trained

and /s/-trained groups.

Experiment	Group	% correct (RT in ms)			
		words with ambiguous fricatives	words with unambiguous fricatives	filler words	filler non-words
1 (reduced ambiguity)	/f/-trained	98 (960)	95 (970)	93 (934)	96 (1037)
	/s/-trained	96 (970)	95 (976)	94 (924)	95 (1048)
2 (shifted to opposite) All	/f/-trained	68 (1122)	95 (1028)	92 (989)	95 (1099)
	/s/-trained	57 (1253)	94 (1081)	94 (1043)	94 (1177)
2 (shifted to opposite)>50%	/f/-trained	85 (1088)	96 (1023)	93 (989)	94 (1116)
	/s/-trained	75 (1266)	95 (1094)	93 (1057)	94 (1204)
no-filter (control)	/f/-trained	97 (1070)	97 (1032)	95 (1001)	94 (1141)
	/s/-trained	96 (1004)	97 (990)	95 (956)	97 (1078)

Test. Figure 2 shows the results of the phonetic categorization task in Experiment 1 (top panel) compared to the no-filter condition (lower panel). 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 in the no-filter condition). This suggests that our hypothesis may be 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 ANOVAs on logit-transformed data to account for the dichotomous dependent variable (/s/ vs. /f/ response; see e.g., Jaeger, 2008, for a discussion of the need for logistic transformation of proportion data). We entered Training (/f/-trained vs. /s/-trained) as a between-participant factor and Continuum step as a within-participant factor.

For Experiment 1, a single main effect was observed for the factor Continuum ($F(3,84) = 188.16$, $p < 0.001$, $\eta_p^2 = 0.87$), reflecting the fact that stimuli were more often categorized as /f/ towards the /f/-end of the continuum. No main effect was observed for the factor Training ($F(1,28) = 0.89$, $p = 0.353$, $\eta_p^2 = 0.031$) nor was there an interaction between Continuum and Training ($F(3,84) = 0.64$, $p = 0.592$, $\eta_p^2 = 0.022$). Hence there is no evidence for perceptual learning in Experiment 1.

This is in strong contrast to the data in the no-filter condition where a significant learning effect could be 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 to the analyses. Again there was a significant main effect of Continuum ($F(3,168) = 482.4$, $p < 0.001$, $\eta_p^2 = 0.896$), indicating that participants gave more /f/ responses towards the /f/-end of the continuum. The only other significant effect was the interaction between Experiment and Training ($F(1,56) = 8.28$, $p = 0.006$, $\eta_p^2 = 0.129$), reflecting the fact that the effect of Training (/s/-trained vs. /f/-trained) differed across experiments. Non-significant results were found for the main effects of Experiment ($F(1,56) = 2.57$, $p = 0.115$, $\eta_p^2 = 0.044$) and Training ($F(1,56) = 2.17$, $p = 0.147$, $\eta_p^2 = 0.037$) and for the interactions between Experiment and Continuum ($F(3,168) = 0.18$, $p = 0.91$, $\eta_p^2 = 0.003$) Training and Continuum ($F(3,168) =$

[2] An analysis of the data reported by Reinisch et al., 2013 with the same analysis method used here resulted in the following effects: Training: $F(1,28) = 10.6$, $p = 0.003$, $\eta_p^2 = 0.275$; Continuum: $F(3,84) = 335.58$, $p = 0.00$, $\eta_p^2 = 0.923$; Training by Continuum: $F(3,84) = 1.03$, $p = 0.385$, $\eta_p^2 = 0.035$.

0.27, p = 0.85, η_p^2 = 0.005) and Training by Experiment by Continuum (F (3,168) = 1.29, p = 0.279, η_p^2 = 0.023).

As can be seen from the separate analyses discussed above, the effect of Training was present only in the no-filter control condition. The interaction between Experiment and Training confirms that there is 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.

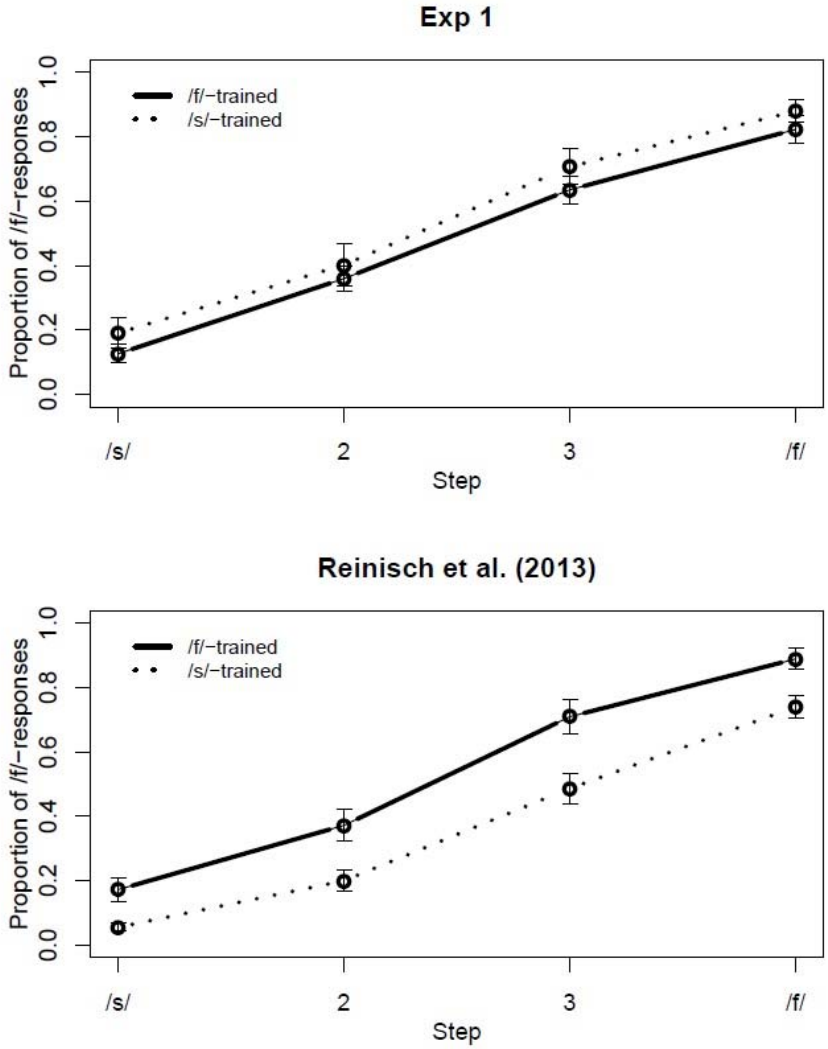

Figure 2: Categorization data for the test continua. The two panels display the data obtained in Experiment 1 (top panel) and in Reinisch et al (2013) for the no-filter condition. Each

panel displays data for the /f/-trained participants (solid lines) and the /s/-trained participants (dotted lines). The test stimuli were identical across the experiments. Error bars indicate the size of the standard error of the mean.

Discussion

Experiment 1 showed that perceptual learning effects are reduced or absent when listeners are presented with a filtered speech signal that causes the acoustically ambiguous fricatives to be perceived as unambiguous, that is, matching the intended lexical option. This is in contrast to the no-filter condition (Reinisch et al. 2013), which had used the same fricatives as were 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 where participants can attribute an unnatural pronunciation to an incidental property of the speaker (such as holding a pen in 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 because the filters had made the ambiguous sounds unambiguous, then we expect to find a learning effect in Experiment 2. This is predicted because the ambiguous sounds will no longer "become unambiguous" as a result of the filtered precursors. In fact, we might expect an even larger learning effect than in the no-filter condition, because in order 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 representation has become even more unlike the intended sounds). Hence we can even test the bounds of perceptual learning, that is, whether effects get larger as the critical sounds are perceived to be more like the other alternative. If, however, our filtering manipulation in Experiment 1 had blocked learning because listeners attributed any ambiguity to unusual sound properties related to the experimental setting, we should not 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 sound more /s/-like, and thus less like the lexically-supported fricative. To accommodate a sound that is 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 should find a learning effect. One alternative that has to be kept in mind (and which will be

further discussed below) 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 non-words, 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 due to the filter itself rather than the nature of the filter.

Methods

Participants 30 native speakers of Dutch were recruited from the same population and according to the same criteria as before. 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) attenuates 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 sound more like /s/, that is, more ambiguous or even more similar to the "wrong" category (i.e., the category that is not supported by the lexical information). The opposite should hold for the /s/-trained listener group whose stimuli were passed through the /s/ minus /f/ filter. The minimal pair continua used for test were the same as in Experiment 1 and remained unfiltered. The experimental procedure was the same as for Experiment 1.

Results

Exposure. The same criteria as in Experiment 1 were 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, here 9 out of the 30 participants failed to meet this criterion (4 in the /f/-ambiguous group). Table 1 reports overall percentage correct and mean reaction times for the full sample of participants and with the 9 participants excluded. Implications of this finding will be discussed below.

Test. Participants who failed the 50% acceptance criterion were excluded from all analyses. Figure 3 displays categorization performance of the remaining participants. It can be observed that, in contrast to 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 group, Continuum and their interaction as factors. This analysis resulted in a main effect of Continuum ($F (3,57) = 81.36$, $p < 0.001$, $\eta_p^2 = 0.811$), reflecting reliable use of the acoustic properties of the stimuli along the continuum, and a main effect of Training ($F (1,19) = 8.24$, $p = 0.01$, $\eta_p^2 = 0.302$). 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 interaction between Continuum and Training was not significant ($F (3,57) = 1.02$, $p = 0.389$, $\eta_p^2 = 0.051$).

Since 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. 2) into our analysis. This analysis showed a main effect of Continuum ($F (3,141) = 262.3$, $p < 0.001$, $\eta_p^2 = 0.848$), again reflecting reliable categorization performance, and, critically, an interaction between Experiment and Training ($F (1,47) = 9.61$, $p = 0.003$, $\eta_p^2 = 0.17$). The inverted filtering manipulation in Experiment 2 resulted in a significant increase in the learning effect compared to Experiment 1. No other main effects or interactions were significant: Experiment ($F (1,47) = 0.06$, $p = 0.808$, $\eta_p^2 = 0.001$), Training ($F (1,47) = 2.32$, $p = 0.134$, $\eta_p^2 = 0.047$),

the two-way interactions between Experiment and Continuum ($F(3,141) = 1.4$, $p = 0.245$, $\eta_p^2 = 0.029$), and Training and Continuum ($F(3,141) = 1.52$, $p = 0.213$, $\eta_p^2 = 0.031$) and the three-way interaction Experiment by Training by Continuum ($F(3,141) = 0.22$, $p = 0.884$, $\eta_p^2 = 0.005$).

Comparing the categorization data for the no-filter condition in Reinisch et al. (2013; bottom panel of Figure 2) to those of Experiment 2 (Figure 3) it can be observed that Experiment 2 led to a numerically larger learning effect. Analyses were performed to test this pattern. The analysis revealed main effects for the factors Continuum ($F(3,141) = 344.89$, $p < 0.001$, $\eta_p^2 = 0.88$) and Training ($F(1,47) = 18.36$, $p < 0.001$, $\eta_p^2 = 0.281$). No main effect was observed for Experiment ($F(1,47) = 2.09$, $p = 0.155$, $\eta_p^2 = 0.043$). Critically, no interaction was observed between Experiment and Training ($F(1,47) = 0.99$, $p = 0.324$, $\eta_p^2 = 0.021$), 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 interactions were observed between Experiment and Continuum ($F(3,141) = 2.41$, $p = 0.069$, $\eta_p^2 = 0.049$), Training and Continuum ($F(3,141) = 0.26$, $p = 0.857$, $\eta_p^2 = 0.005$) or for the three-way interaction between Experiment, Training and Continuum ($F(3,141) = 1.95$, $p = 0.125$, $\eta_p^2 = 0.04$).

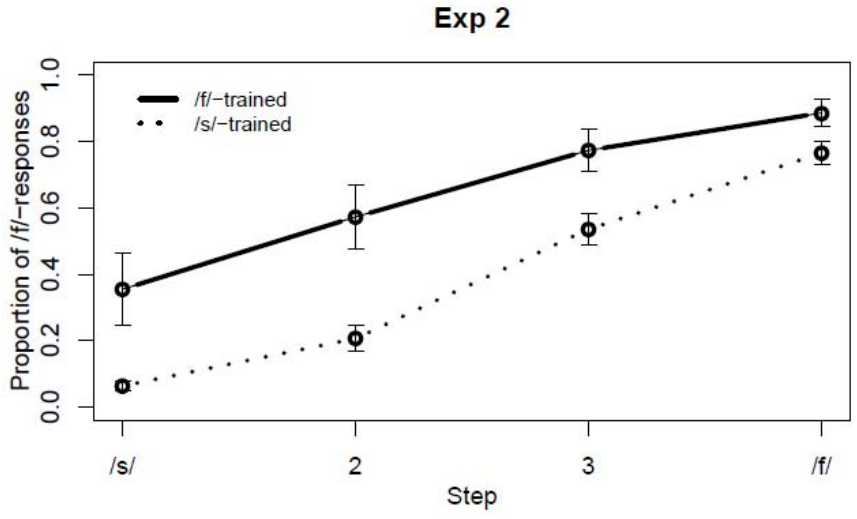

Figure 3: categorization data for the test continuum. The panel displays data for the /f/-trained participants (solid lines) and the /s/-trained participants (dotted lines). Error bars indicate the size of the standard error of the mean.

Discussion

Experiment 2 tested perceptual learning in a condition in which the acoustic context surrounding the ambiguous fricatives should cause the fricatives to be perceived as more ambiguous, or even closer to the other endpoint on the /f/-/s/ continuum, than the lexically-supported category. In contrast to Experiment 1, here we did find a learning effect and this effect was statistically different from 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 but 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 towards the wrong interpretation. This would lead to the perception of the critical training words as non-words. Given that non-words 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 interpretation above 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. Figure 4 shows 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. Each dot represents the combined scores per participant. Open circles represent participants from the /s/-trained group, and closed circles represent participants from the /f/-trained group. Regression lines are fitted to the data of the two separate groups to visualize the differences in this relation between the two groups. 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 participants that 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 is 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 learnt 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 those participants who accepted most critical training items. These two patterns show that the size of the training effect is dependent on the proportion of critical items that are 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.

Independent variables were Training (/f/-trained vs. /s/-trained) and Boundary. Boundary was defined as the per-participant proportion of /f/ responses across the continua in the test phase.

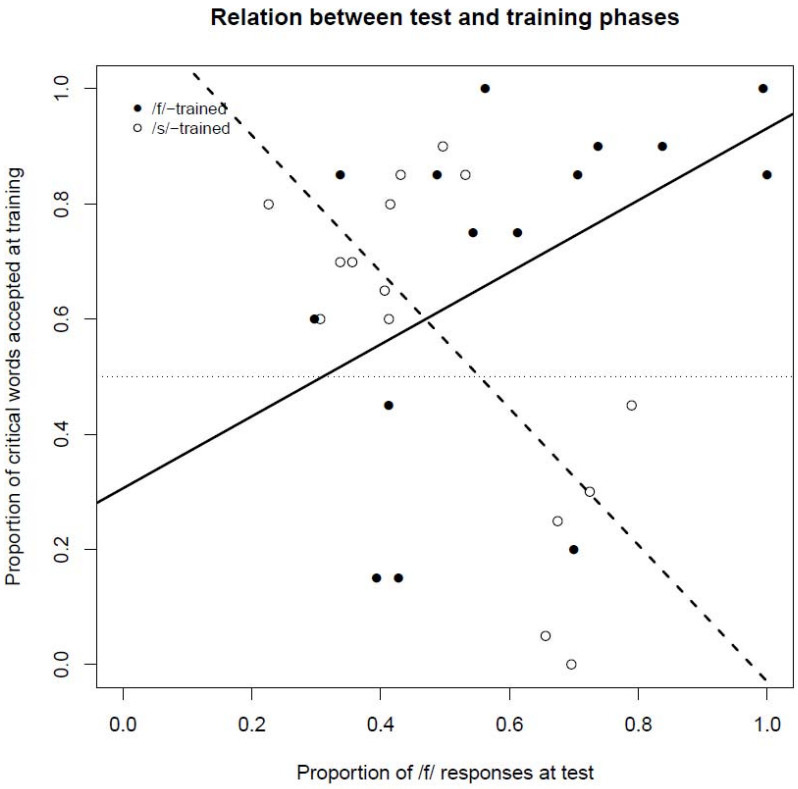

Figure 4: Dot-plot displaying the relation between the acceptance rates for the critical training items in the training phase of Experiment 2 (y-axis), and the proportion of /f/ responses that an individual gave in the test-phase (x-axis). Points represent individual participant scores, both for the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

For the regression analysis the /f/-trained group was assigned the reference level for Training. Therefore, main effects for the other factors reflect patterns for the /f/-trained group only. Interactions between an effect with the factor Training indicate how the /f/ and /s/-

trained groups differ for that factor. A small main effect was observed for category Boundary ($b_{Boundary}$ = 0.63, t = 2.09, p = 0.046; $b_{Intercept}$ = 0.31, t = 1.60, p = 0.12), 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 an interaction between Boundary and Training ($b_{Boundary*Training}$ = -1.81, t = -3.71, p = 0.001). This interaction indicates 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, while 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 ($b_{Training}$ = 0.85, t = 3.05, p = 0.005), 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 part 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, due to perceptual contrast effects in Experiment 2, participants had to re-map 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 due to the fact that participants associated 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 the critical sounds can be attributed to external circumstances such as the speaker putting a pen in her mouth while articulating the

[3] The possibility of an opposite learning effect in fact also follows our manipulation. Even though the focus is 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 makes the ambiguous fricative sound perceptually more /s/-like, reducing the size of the boundary shift for /s/. However, the filter also makes the /f/ sound perceptually more like /s/, moving the sound perceptually into an ambiguous region. This could then induce an extension of a participant's /f/ category.

critical words, perceptual learning does not occur. Here the filters could have served as the external circumstance (e.g., the speaker is located in a room with unusual room acoustics). These explanations were disproved because a learning effect was found in Experiment 2 where the same filters were applied to the training materials as in Experiment 1, with the only difference that the acoustic and lexical context now supported the opposite sound category. A way to reconcile the present data with findings such as Kraljic et al. (2008) is to look at the issue of external evidence from a slightly different angle. If the lack of learning in Kraljic et al. (2008) is explained such that the ambiguity has been "taken care of" through the attribution to the pen - hence 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 is "taken care of" by the acoustic context, which shifts the fricatives perceptually towards the intended unambiguous category. In this case the acoustic context in the present study would not be circumstantial evidence but just another factor that takes care of the critical sounds' ambiguity, reducing the amount of lexically-guided perceptual learning[4].

Cognitive implementation

In Experiment 1 of the present study 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

[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.

unambiguous input at the level where retuning is implemented. This signal would then be mapped onto the lexically supported phoneme category, and any resulting lexical feedback would lead 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 while perceptual learning is implemented in the connections between features and phoneme representations (McClelland et al., 2006; see Appendix A). Shortlist B (Norris & McQueen, 2008), despite its lack of explicit description 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, in Experiment 1, on any single trial during exposure (i.e., the lexical decision task), contrast effects would shift the perceptual representation towards the lexically supported alternative. This would have prevented a lexical mismatch, which would in turn prevent an error signal to be sent from the lexicon to the phoneme representations. Then, although, in principle, the prelexical processing had access to the perceptually ambiguous fricative (i.e., an "untransformed" representation), no error signal is sent since it had already been blocked by the contrast effects. Therefore, perceptual learning could not associate the ambiguous sound with the lexcially-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 has specifically been 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 of the expected cue distribution for this 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 "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 only 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 are investigated here are only part of the total set of contrastive and compensatory 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 & Lotto, 2002; Holt, 2005; 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 precede lexically-guided perceptual learning.

An interesting final aspect about 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, is 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 were able to map out how lexically-guided perceptual learning and perceptual contrast effects due to acoustic context operate relative to each other. By explicitly testing this cognitive ordering for the first time it was shown that perceptual contrast effects have to at least partially precede lexically-guided

perceptual learning.

ACKNOWLEDGMENTS

We would like to thank the research assistants of the Psychology of Language Group at the MPI for Psycholinguistics for help with running the experiments. The corresponding author is now employed by the Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Radboud University, Nijmegen, The Netherlands, and is seconded at the Department of Linguistics, University of California Berkeley, Berkeley, California, USA.

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*, 1-28.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1).

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.

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*(4), 604-618.

Cutler, A., McQueen, J. M., Butterfield, S., & Norris, D. (2008). *Prelexically-driven perceptual retuning of phoneme boundaries.* Paper presented at the INTERSPEECH.

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

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

Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science, 16*(4), 305-312.

Holt, L. L. (2006). Speech categorization in context: Joint effects of nonspeech and speech precursors. *Journal of the Acoustical Society of America, 119*(6), 4016-4026.

Holt, L. L., & Lotto, A. J. (2002). Behavioral examinations of the level of auditory processing of speech context effects. *Hearing Research, 167*(1-2), 156-169.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences

vowel identification. *Journal of the Acoustical Society of America, 108*(2), 710-722.

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

Kawahara, H., Masuda-Katsuse, I., & de Cheveigné, A. (1999). Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication, 27*(3), 187-207.

Kleinschmidt, D., & Jaeger, T. F. (2011). A Bayesian belief updating model of phonetic recalibration and selective adaptation. Paper presented at the Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics.

Kleinschmidt, D., & Jaeger, T. F. (2012). A continuum of phonetic adaptation: Evaluating an incremental belief-updating model of recalibration and selective adaptation. Paper presented at the Proceedings of the 34th Annual Conference of the Cognitive Science Society.

Kluender, K. R., Coady, J. A., and Kiefte, M. (2003). "Sensitivity to change in perception of speech," Speech Commun. 41, 59–69.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive psychology, 51*(2), 141-178.

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First Impressions and Last Resorts How Listeners Adjust to Speaker Variability. *Psychological science, 19*(4), 332-338.

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*(1), 161-168.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29*(1), 98-104.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, *10*(8), 363-369.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science, 30*(6), 1113-1126.

Mirman, D., McClelland, J. L., & Holt, L. L. (2006). An interactive Hebbian account of lexically guided tuning of speech perception. *Psychonomic Bulletin & Review, 13*(6), 958-965.

Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition, 129*(2), 356-361.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America, 85*(5), 2088-2113.

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

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*(2), 204-238.

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.

Poellmann, K., McQueen, J. M., & Mitterer, H. (2011). The time course of perceptual learning. *training*, *9*, 9-5.

Reinisch, E., & Holt, L. L. (2013). Lexically Guided Phonetic Retuning of Foreign-Accented Speech and Its Generalization. *Journal of Experimental Psychology: Human Perception and Performance. online first publication*

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*(2), 101-116.

Reinisch, E., Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance, 39*(1), 75-86.

Reinisch, E., Wozny, D., Mitterer, H. & Holt, L. L. (2014). Phonetic category recalibration: What are the categories? Journal of Phonetics, 45, 91-105

Samuel, A. G. (1986). Red herring detectors and speech perception: In defense of selective adaptation. *Cognitive Psychology*, *18*(4), 452-499.

Sjerps, M. J., & McQueen, J. M. (2010). The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 36*(1), 195-211.

Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011a). Constraints on the processes responsible for the extrinsic normalization of vowels. *Attention , Perception, & Psychophysics, 73*(4), 1195-1215.

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*(14), 3831– 3846.

Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2012). Hemispheric differences in the effects of context on vowel perception. *Brain and Language, 120*(3), 401-405.

Sjerps, M. J., & Smiljanić, R. (2013). Compensation for vocal tract characteristics across native and non-native languages. *Journal of Phonetics, 41*(3), 145-155.

Skoe E, Krizman J, Spitzer E, Kraus N (2013) The auditory brainstem is a barometer of rapid auditory learning. Neuroscience 243:104–114

Stilp, C. E., Alexander, J. M., Kiefte, M. J., & Kluender, K. R. (2010). Auditory color constancy: Calibration to reliable spectral properties across nonspeech context and

targets. *Attention, Perception, & Psychophysics, 72*, 470-480.

Summerfield, Q., Haggard, M., Foster, J., & Gray, S. (1984). Perceiving vowels from uniform spectra: Phonetic exploration of an auditory aftereffect. *Perception & Psychophysics, 35*(3), 203-213.

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*(1), 74-79.

Viswanathan, Navin, James S. Magnuson, and Carol A. Fowler. (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 (4), 1005–1015.

Watkins, A. J. (1991). Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion. *Journal of the Acoustical Society of America, 90*(6), 2942-2955.

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*(3), 1263-1282.

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: Sijthoff.

APPENDIX A

Although computational modelling is not the focus of the present paper a discussion of our design in light of such a model (here TRACE, McClelland & Elman, 1986) may help to lay out our predictions and results in more detail. 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. Based on this architecture a version of TRACE has been established for modelling perceptual learning data by adding a Hebbian learning algorithm to account for long-term phoneme adjustments (Hebb-TRACE, Mirman, McClelland, & Holt, 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, Mirman & Holt, 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 argued in the current paper as a potential implementation. Consider a case where 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 is 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 come available that favours the lexically-consistent interpretation of [s]. The combination of acoustic and lexical information will point the listener towards recognizing [s] and accept 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 is already in line with those at the phoneme unit level. That is, little to no learning should occur. In contrast, in cases where information from the feature level mismatches the lexically-supported phoneme (as in Experiment 2), the feature-to-phoneme connections should gradually be tuned towards the lexically consistent alternative, resulting in perceptual learning.