Journal of Memory and Language xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

# No delays in application of perceptual learning in speech recognition: Evidence from eye tracking

Holger Mitterer[a,*], Eva Reinisch[b]

[a] Max Planck Institute for Psycholinguistics, The Netherlands
[b] Institute of Phonetics and Speech Processing, University of Munich, Germany

ARTICLE INFO

*Article history:*
Received 4 October 2012
revision received 2 July 2013
Available online xxxx

*Keywords:*
Speech perception
Perceptual learning
Eye tracking

ABSTRACT

Three eye-tracking experiments tested at what processing stage lexically-guided retuning of a fricative contrast affects perception. One group of participants heard an ambiguous fricative between /s/ and /f/ replace /s/ in s-final words, the other group heard the same ambiguous fricative replacing /f/ in f-final words. In a test phase, both groups of participants heard a range of ambiguous fricatives at the end of Dutch minimal pairs (e.g., *roos-roof*, 'rose'-'robbery'). Participants who heard the ambiguous fricative replacing /f/ during exposure chose at test the f-final words more often than the other participants. During this test-phase, eye-tracking data showed that the effect of exposure exerted itself as soon as it could possibly have occurred, 200 ms after the onset of the fricative. This was at the same time as the onset of the effect of the fricative itself, showing that the perception of the fricative is changed by perceptual learning at an early level. Results converged in a time-window analysis and a Jackknife procedure testing the time at which effects reached a given proportion of their maxima. This indicates that perceptual learning affects early stages of speech processing, and supports the conclusion that perceptual learning is indeed perceptual rather than post-perceptual.

## Introduction

Even though listeners become attuned to the typical pronunciations of the sounds of their native language during the first year of life (Werker & Tees, 1984), recent evidence shows that these established phonetic categories remain surprisingly flexible (for a review, see Samuel & Kraljic, 2009). This flexibility can be experienced in everyday life when listening to speakers with different regional and foreign accents: as we become more familiar with their pronunciation peculiarities, their speech becomes easier to understand. This has been demonstrated empirically on a global level as good recognition of foreign-accented words after some exposure (Bradlow & Bent, 2008) but also on a more fine-grained phonemic level as listeners adjust to the unusual pronunciation of a particular native-language segment (starting with the seminal study by Norris, McQueen, & Cutler, 2003). Even though it seems well established (anecdotally and empirically) that adjustment to a speaker does occur, what has not been addressed so far is when during speech processing the new knowledge about the pronunciation of a segment is applied. Once we know that a certain speaker produces a phoneme in an unusual fashion, do we immediately interpret new instances of this phoneme in relation to our prior experience? Or is early phonetic processing not affected by perceptual learning, and only the final decision about the segment's identity is influenced by the newly learned knowledge? (Similar to the conceptualization of auditory and visual processing in Massaro's, 1998, FLMP model.) The present study set out to address this question by revealing the cognitive stages of speech processing at which knowledge about pronunciation variants is taken into account. Specifically we asked whether retuned

* Corresponding author. Address: University of Malta, Department of Cognitive Science, Msida, Malta.
*E-mail address:* holger.mitterer@um.edu.mt (H. Mitterer).

0749-596X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jml.2013.07.002

phonetic categories affect early perceptual stages of speech processing. In that case, acoustic cues would be interpreted in the light of the known pronunciation variants, or whether retuned categories come into play at a later post-perceptual processing stage.

The adjustment to unusual pronunciation variants of single phonemes was first demonstrated by Norris et al. (2003). They exposed Dutch listeners to a speaker who produced an ambiguous fricative between /s/ and /f/ (transcribed from here on as [$^s/_f$]). One group of participants heard this ambiguous fricative replace /s/ in s-final words, as in [mœy$^s/_f$] ("mouse"); the other group heard the same ambiguous fricative replacing /f/ in f-final words, as in [ʃirɑ$^s/_f$] ("giraffe"). Importantly, the fricative could only be interpreted as /s/ or /f/ in these stimuli since the other possible interpretation (i.e., [mœyf] and [ʃirɑs]) are nonwords in Dutch. That is, the phonetically ambiguous fricative [$^s/_f$] was presented in a lexcially unambiguous context of an existing Dutch word. Listeners could thus use lexical information to interpret the ambiguous sounds (Ganong, 1980). In a lexical decision task, which served as an exposure phase, the words with ambiguous fricatives were mostly accepted as real words. Immediately thereafter, participants had to categorize sounds along an [ɛs] to [ɛf] continuum. The results of the categorization task were influenced by the exposure condition. Participants who had heard the ambiguous fricative in f-final words gave more /f/ responses for tokens from the [ɛs] to [ɛf] continuum than participants who had heard the ambiguous fricative in s-final words. Apparently, participants had learned, guided by lexical knowledge, that the same ambiguous fricative [$^s/_f$] can be a possible implementation of either /s/ or /f/.

Further experiments on this type of perceptual learning showed that the effect is speaker specific (at least for fricatives, Eisner & McQueen, 2005; Kraljic & Samuel, 2007), but generalizes over lexical items (McQueen, Cutler, & Norris, 2006; Mitterer, Chen, & Zhou, 2011). Moreover, the effect has been shown with a variety of tasks during exposure, ranging from simply counting words (McQueen, Norris, & Cutler, 2006) to hearing a story or watching a TV show (Eisner & McQueen, 2006; Mitterer & McQueen, 2009). So it is well established that listeners flexibly retune their phoneme categories. However, there are at least two ways in which this learning might influence perception. One possibility is that the newly acquired knowledge may immediately influence the processing of incoming information in the speech stream. That is, the knowledge that the speaker produces certain sounds in an unusual fashion could be applied during the initial stage of phonetic processing, at the time when the unfolding speech signal is being processed. Alternatively, phonetic processing may not be influenced directly. Rather, the newly acquired knowledge may only be consulted after an initial, speaker-independent phonetic processing of the input. The effect of learning would then have no influence on initial phonetic processing of incoming information, but would only be integrated with the outcome of phonetic processing at a later stage.

The distinction between early versus late integration is a common one in speech perception research. Kingston and Macmillan (1995), for instance, asked whether nasalization and the first-formant frequency are perceived integrally or independently for the perception of vowel height. The research question pursued by Kingston and Macmillan was whether integration already occurs at a phonetic level or whether the dimensions are perceived independently at a phonetic level, and are integrated late at a decision level. Kingston and Macmillan used signal-detection theory to show that listeners do not distinguish between degrees of vowel nasalization and first-formant frequency but instead integrate both dimensions at a phonetic level to form one cue for vowel height. A similar question arose in the debate relating to how listeners achieve "compensation for phonological assimilation". Phonological assimilation is a production process in which a given segment is so strongly coarticulated with its context that it "loses its identity" and takes over one property of the context segment. An example is assimilation of place of articulation of word-final nasals: an underlying /n/ in *lean bacon* /lin beɪkn/ can become an [m] in the surface form [lim beɪkn]. The underlying /n/ has then been assimilated to the labial place of articulation of the following /b/. Gaskell (2003) proposed a model of compensation for phonological assimilation in perception in which the assimilated segment (e.g., the [m] in [lim beɪkn]) is first perceived as an instance of its surface form (i.e., as /m/). Only at a later processing stage is the context taken into account, such that the [m] is treated as a possible instance of an underlying /n/. This contrasts with the proposal by Mitterer, Csépe, and Blomert (2006) who argued that the context already influences the initial perceptual processing of the assimilated segment, making the [m] "sound" like an /n/ already at an auditory level, similar to auditory backward masking (Moore, 2003).

Most prominently, the distinction between early and late integration has featured in the field of audiovisual speech perception. Proponents of gestural theories of speech perception argued that the visual and auditory information streams are integrated at an early level of speech perception (Fowler, Brown, & Mann, 2000). This contrasts with the model of Massaro (1998), in which auditory and visual sensory processing proceed independently and are only integrated at a decision stage. Massaro's (1998) proposal—independent sensory processing in the auditory and visual domains followed by integration at a decision level—resonates with a proposal for a distinction between an initial, fast, first-pass processing and a later reevaluation in visual perception. Lamme and Roelfsema (2000) argued that there is an initial fast feedforward sweep of sensory processing that is relatively stable and automatic. Visual awareness, however, seems to depend on additional horizontal and recurrent processing, that is, processing within one brain area or re-entrant processes from later areas, respectively. As this shows, a frequent distinction is made between early first-pass sensory processing and later re-evaluation and decision processes. In a way, this distinction relates to the common title "Sensation and Perception" used for textbooks in introductory psychology.

In the current paper, we ask whether the results of lexically guided retuning of phonemes are brought to bear on

Please cite this article in press as: Mitterer, H., & Reinisch, E. No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language* (2013), http://dx.doi.org/10.1016/j.jml.2013.07.002

such an initial feedforward sweep of phonetic processing or only at a later integration stage. Lexically guided retuning has been linked to adaptation to accents and individual speakers (Bradlow & Bent, 2008; Mitterer & McQueen, 2009). The literature on speaker adaptation effects provides examples for both influences on early perceptual processes and late integration, so that a priori, this literature makes both options viable. Evidence for influences on early perceptual processes stems from research using the Ladefoged and Broadbent (1957) vowel-normalization paradigm. Sjerps, Mitterer, and McQueen (2011) adapted the paradigm so that concurrently, event-related potentials could be measured. They found that "speaker identity" (in this case, the speaker's average first formant frequency) influenced early perceptual processes, reflected in the N1 component, an early ERP component liked to processing in the auditory cortex (see Reinisch & Sjerps, 2013, for converging evidence using eye-tracking). Other examples of speaker normalization are likely to influence late integration. Johnson, Strand, and D'Imperio (1999) showed that audiovisual stimuli of male versus female speakers influence phonetic perception based on the expectation that female speakers have smaller vocal tracts than male speakers. Given that even visible speech gestures fail to influence early auditory processing in the N1 time window (Colin, Radeau, Demolin, Colin, & Deltenre, 2002; Stekelenburg & Vroomen, 2007), it seems unlikely that gender expectation would be able to do just that. Moreover, Hanulíková, van Alphen, van Goch, and Weber (2012) tested the effects of grammatical errors from native and foreign-accented voices and found that the late P600 reflects speaker adaptation, with a smaller P600 for the foreign-accented speaker, but no speaker-dependent modulation of an earlier component (LAN; early left anterior negativity). Speaker adaptation hence consists of both early adaptation and later integration processes. This raises the question whether lexically guided retuning, as another example of speaker adaptation, affects early perceptual processes or late integration.

Previous studies on perceptual learning seem to favor an account of early, first-pass phonetic processing. Sjerps and McQueen (2010) tested how "complete" perceptual learning is. Their exposure phase was similar to the one in Norris et al. (2003) but the test phase made use of cross-modal identity priming instead of phonetic categorization. They tested how strongly an ambiguous auditory stimulus primed members of minimal word pairs, for example, how well [ro$^{s}$/$_{f}$] primed visual lexical decisions to *roos* (/ros/ is the Dutch word for *rose*) vs. *roof* (/rof/ *robbery*). Critically, priming effects were dependent on exposure. Participants who heard the ambiguous sound in s-final words before (e.g., [mœy$^{s}$/$_{f}$] for /mœys/, *mouse*) showed priming from[ro$^{s}$/$_{f}$] to *roos* but not *roof*. The opposite occurred for participants who heard the ambiguous sound on f-final words (e.g., [ʃirɑ$^{s}$/$_{f}$] for /ʃirɑf/, *giraffe*). For them, the ambiguous prime only facilitated the recognition of *roof*. In a comparison of the magnitude of these priming effects with effects of an experiment in which unambiguous primes were used (i.e., [ros] for *roos*) no difference could be found. That is, after lexically-biased exposure, priming by the ambiguous tokens was just as strong as the priming by unambiguous tokens. This suggests that perceptual learning is complete, and that, after exposure, the ambiguous fricative is treated as a fully acceptable token of the respective fricative category. A similar argument can be made with regard to findings that show generalization of the retuned categories, for example, across words (McQueen, Cutler, et al., 2006; Mitterer et al., 2011), across position in the word (Jesse & McQueen, 2011), and even across languages that share the same phonemes (Reinisch, Weber, & Mitterer, 2013). Once listeners had adjusted to the new pronunciation variant, they apply this knowledge "across the board" suggesting an unspecific, hence early application during processing. Importantly, however, none of these previous studies have looked at the temporal locus of the effect, compared to the processing of acoustic cues.

Another way to view the question about the locus of perceptual learning is in terms of Signal Detection Theory (Macmillan & Creelman, 1991). Signal Detection Theory distinguishes changes in sensitivity and changes in bias, which may be mapped on an early versus late locus of perceptual learning, respectively. Following this line of thought, Clarke-Davidson, Luce, and Sawusch (2008) tested perceptual learning for an /s/-/ʃ/ contrast not only with a phonetic identification task, but also with a discrimination task comparing the discrimination of steps along the /s/-/ʃ/ continuum. The peak of a discrimination function is the indicator of a perceptual boundary. The rationale was that, if perceptual learning was simply due to a decision bias and not a change in phonetic category representations, perceptual learning should not influence performance in a discrimination task (see Kingston & Macmillan, 1995, for a similar strategy). However, Clarke-Davidson et al. found that the peak of the discrimination function over the continuum had changed depending on exposure. The group with s-biased exposure was better at discriminating pairs near the /ʃ/ end of the continuum, and the group with /ʃ/-biased exposure was better at discriminating pairs near the /s/ end of the continuum. Importantly, this result was obtained even if the exposure task did not require listeners to pay attention to the ambiguous sounds (e.g., by requiring explicit lexical decisions). This suggests that between groups the category boundaries were shifted, as after exposure the s-biased group had a category boundary closer to the /ʃ/ end of the continuum.

However, none of the abovementioned studies can really show that perceptual learning influences the first-pass phonetic analysis of the auditory input, simply due to the fact that the behavioral responses in these tasks were given well after the presentation of the fricative. Norris et al. (2000), for instance, commented extensively on the problems of using signal-detection methods for inferences about the levels of processing. One way to probe early perceptual processing is by using eye-tracking measures (Allopenna, Magnuson, & Tanenhaus, 1998). Eye-tracking measures in a visual world paradigm can provide a continuous record of the lexical hypotheses of the listener. However, the distinction between early perceptual processes and later decision processes using eye tracking may not be quite as straightforward as it seems. Even "early" eye movements due to linguistic input are necessarily the output of some kind of decision process. Some kind of "decision" has to be involved in planning and

Please cite this article in press as: Mitterer, H., & Reinisch, E. No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language* (2013), http://dx.doi.org/10.1016/j.jml.2013.07.002

executing an eye movement towards a potential referent. The question then is whether these decision processes add a task-specific dynamics of their own. Tanenhaus, Magnuson, Dahan, and Chambers (2000) provide a review of relevant studies. They argue that, though some decision has to be involved before an eye movement is executed, the eye-movement record seems to be a reflection of linguistic processing with little task-specific influences. Evidence for this assumption stems, for instance, from the finding that non-displayed lexical competitors influence fixation proportions. Hence, using an eye-tracking paradigm will allow us to address the issue whether application of perceptual learning occurs during first-pass phonetic processing or at a later decision stage.

Poellmann, McQueen, and Mitterer (2011) used a "visual-world eye-tracking task" with printed words to examine lexically-guided perceptual learning. Participants saw four printed words on a screen and heard an instruction to click on one of the words. Simultaneously, their eye movements were monitored. In this paradigm, fixations on the printed words reveal the lexical hypotheses that listeners are forming as they hear the words unfold (for the use of printed words, see McQueen & Viebahn, 2007). Perceptual learning was assessed with minimal pairs. On these trials, participants heard, for example, the ambiguous stimulus [ro$^{s}/_{f}$] and saw the printed words 'roos' and 'roof' on the screen. Poellmann et al. found the expected perceptual learning effect; participants with an /s/-biased exposure looked more to 'roos' than participants with an /f/-biased exposure, but only after hearing ten exposure trials in which the ambiguous stimulus [$^{s}/_{f}$] occurred in unambiguous positions (e.g., [ʃirɑ$^{s}/_{f}$] for /ʃirɑf/, "giraffe").

An unexpected finding from this study, however, was that the effect of exposure condition emerged only a full second after target onset, which is about 700–800 ms after the fricative had been heard. This late effect was even more surprising given that eye tracking usually picks up the processing of phonetic information 200 ms after its onset (Allopenna et al., 1998).[1] These data thus suggest that the knowledge gained during exposure is not employed in the initial feedforward sweep of sensory processing. Instead, the knowledge gained in the exposure phase seems to be integrated late with the incoming speech signal. But this study was not designed to address the level at which lexically guided retuning affects processing. The main focus was on how many examples of the ambiguous stimulus [$^{s}/_{f}$] in lexically-biased contexts were necessary to support retuning. As a consequence, the amount of data on which this conclusion is based is very limited, since learning was only observed for a subset of the test trials (i.e., 10 trials per participant). Therefore, in the present study we ran perceptual learning experiments in which learning was first well established with 20 examples of the ambiguous stimulus [$^{s}/_{f}$] in lexically unambiguous contexts before testing with a larger number of trials per participant.

The central feature of the study was that we directly pitted the effect of exposure against the effect of the fricative spectrum by presenting participants with a range of more /s/-like and more /f/-like stimuli. This provides us with two tests as to whether the knowledge gained by perceptual learning is used during first pass processing or not. The first test is "study internal". We compared the time point when participants' eye movements were influenced by the acoustic properties of the fricatives with the time point at which group differences (i.e., /s/-biased vs. /f/-biased exposure) were evident in the eye-movement record. If perceptual learning is integrated early with incoming acoustic information, these time points should not differ. If perceptual learning is integrated late, the effect of the acoustic properties should be visible earlier than the effect of perceptual learning. Next to this "study internal" test, which is about the relative timing of the effects in the current study, there was also an external criterion. A large amount of literature using eye tracking to study the immediate uptake of acoustic cues (e.g., McMurray, Clayards, Tanenhaus, & Aslin, 2008) shows that acoustic cues influence eye movements within about 150–200 ms. Because this lag is mostly attributed to the planning of eye movements, eye tracking hence provides a window on early processes in speech perception. If the effect of perceptual learning is used directly during the uptake of acoustic cues, we should find a similar time course, with effects of exposure emerging around 200 ms after hearing the onset of an ambiguous fricative.

## Experiment 1

In an exposure-test paradigm, listeners first performed a lexical decision task during which they were exposed to the unusual pronunciation of a fricative. Depending on exposure group, either /f/ or /s/ was replaced by an ambiguous sound. This part was similar to many previous studies on lexically-guided category retuning ("perceptual learning") in Dutch (see, e.g., Norris et al., 2003). The innovation of the present study was the format of the test phase. Listeners saw four printed words on the screen, from two /s/-/f/ minimal pairs. They then heard a partially ambiguous stimulus with the instruction to click on the word they heard, while their eye movements were tracked. Using a visual-world paradigm with printed words, we examined how early the effects of lexically-guided learning influence speech processing. Note, however, that there is, in contrast to typical visual-world paradigm, not a clear target and a clear competitor (as *beaker* when the target is *beetle*), since all stimuli were at least to some extent ambiguous. Listeners always chose from a target pair in which both words were possible targets (e.g., 'roos' and 'roof').

### Method

#### Participants

Twenty-six members of the Max Planck Institute for Psycholinguistics' participant pool (19 female/7 male) participated in the experiment for pay. They were recruited from the student population in Nijmegen, The Netherlands

[1] There are claims that this is a conservative estimate (Altmann, 2011), an issue to which we return in the Section 'General Discussion').

and were between 18 and 27 years of age. None reported any hearing problem and all had normal or corrected-to-normal vision. Two participants did not complete the experiment due to problems with the calibration procedure of the eye tracker. Data from these participants was discarded.

*Materials*

The materials comprised 110 Dutch words and 100 nonwords that were phonologically legal in Dutch. The set of words consisted of 50 critical items and 60 filler words. Ten of the critical items were minimal pairs ending in /f/ and /s/ and were selected as test items for categorization in the eye-tracking task (*doof-doos* "deaf"-"box", *les-lef* "lesson"-"guts", *roof-roos* "robbery"-"rose", *half-hals* "half"-"neck", *kuif-kuis*, "curl of hair"-"chaste"). Of the remaining 40 critical items, half ended in /f/ (e.g., *locomotief* "locomotive") and half ended in /s/ (e.g., *geitenkaas* "goat cheese"). Importantly, these words are nonwords if the fricatives are exchanged. That is, *locomotie[s]* and *geitenkaa[f]* are nonwords in Dutch. Except for the word-final position of the critical phonemes, none of the words or nonwords contained the sounds /f/, /s/, or their voiced counterparts /v/ and /z/. This additional constraint was imposed because many Dutch speakers use only unvoiced fricatives.

All words and nonwords were recorded by a female Dutch native speaker (aged 28) in a soundproof booth. Additionally, the critical /s/- and /f/-final items for the exposure phase (such as *locomotief* and *geitenkaas*) were recorded in their original version and as a nonword with the other fricative (i.e., *locomotie*[s] and *geitenkaa*[f]). This was necessary to create the ambiguous stimuli for the exposure phase, which were based on interpolation between the /s/- and /f/-final versions, as explained below.

For each /f/-final and /s/-final recording of the critical words, including the minimal pairs, the fricatives plus one or two preceding phonemes were spliced out and morphed in an 11 step continuum (0–100% of the /f/-final recording, in steps of 10%) using the STRAIGHT algorithm (Kawahara, Masuda-Katsuse, & de Cheveigné, 1999) in Matlab (MathWorks Inc.). Splicing points were chosen such that the resulting word sounded maximally natural when the morphed part was spliced back onto the beginning of the word. Note that we morphed the fricatives and at least one preceding segment. This ensured that not only the frication noise but also the formant transitions into the frication noise were ambiguous. Phoneme boundaries were used as temporal anchors for the morphing algorithm. In this way, different types of phonemes (i.e., fricatives vs. preceding sounds) were time-aligned, and only segments of the same type were morphed (i.e., vocalic portions of the signal with other vocalic portions and frication noise with frication noise).

To generate tokens that are phonetically ambiguous and to find a range of ambiguous stimuli for the test phase, a pretest with this stimuli was performed (see Reinisch et al., 2013, for details). Using the 10% /f/ + 90% /s/ token and 90% /f/ + 10% /s/ token, we found performance near floor (1% /f/ responses) and ceiling (95% /f/ responses). Based on this pretest, ambiguous tokens for the exposure phase were generated and subsets of four consecutive steps of the original 11-step continua for all 5 minimal pairs were selected for use in this experiment. To do this, we identified where the interpolated identification function crossed 50% (which was always between two measured points) and used the two steps above and below this point.

*Apparatus and procedure*

During exposure, half of the participants were randomly assigned to the /f/-biased condition and the other half was assigned the /s/-biased condition. All participants heard the same 60 filler words and 100 nonwords. Participants in the /f/-biased condition were further presented with the 20 /f/-final words in which the /f/ was replaced by the ambiguous fricative, and the /s/-final words in which the /s/ was unambiguous (i.e., the /s/-endpoint of the morphed continua). Participants in the /s/-biased condition heard all /s/-final words with the ambiguous sounds, and /f/-final words with the /f/-endpoints of the continua.

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

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

Immediately following exposure, all participants completed the same visual-world eye-tracking task with the five Dutch minimal pairs. First, the eye-tracking cameras of an SR Research Eyelink 2 eye tracker were fitted. The eye tracker was calibrated with a 32.5 × 24 cm screen at a 60 cm distance. Then participants received written instructions that their task on every trial was to click with the computer mouse on the displayed word they thought they heard. After participants pressed a button to confirm that they understood the instructions, the visual-world task began. Each participant was presented with an individual random order of trials generated with the following constraints. Each of the 20 test stimuli (five word–word continua with four steps) was presented eight times. The presentation was organized in eight blocks with a random permutation of the stimuli within a block. That is, listeners were presented with all stimuli once before a repetition occurred.

On each trial, four printed words were displayed on the screen, centered in the four quadrants. Two of the words were the members of the continuum the auditory stimulus was taken from. The other two served as distractors and were taken from another randomly chosen minimal pair. That is, if an auditory stimulus from the [ros]-[rof] continuum was presented, the visual display contained the words *roof* and *roos* plus the words from another minimal pair also used in the experiment (e.g., *hals-half*, "neck"-"half"). The positions of the two words potentially matching the auditory input on the screen were counterbalanced, that is, each appeared equally often on each of the four available positions on the screen. Note that there are twelve (four times three) different combinations of positions on the screen for the two potential targets. With 160 test trials per participants, it was not possible that each combination of positions was presented to a given participant equally often (with each combination appearing at least 13 times [13 repetitions × 12 combinations = 156 trials] and four combinations appearing 14 times). Therefore, the randomization procedure additionally counterbalanced these position combinations across participants.

After every six trials, a drift correction trial was inserted to adjust for possible shifts of the cameras on the head relative to the eyes. If necessary, the calibration was adjusted on these trials. Exposure and test phase were implemented using the Experiment Builder software (SR Research). Completing the whole experiment took approximately 30 min.

### *Analyses*

After checking the acceptance of words containing ambiguous sounds in the exposure phase—that is, does a Dutch word like *giraf* give rise to a "yes" response in the lexical decision task—, three types of analyses were performed on the critical data from the test trials. In the first analysis, we tested whether there was an overall learning effect in the click responses. We used linear mixed-effects models with a logistic linking function, which takes into account the categorical nature of the dependent variable (a click on the f-final word[2] was coded as one and a click on an s-final word was coded as zero). Fixed effects were Exposure Group (/f/-biased or /s/-biased; between participant factor) and Continuum Step (within participants); both factors were coded as a numerical factor centered on zero (/s/-bias as −0.5, /f/-bias as +0.5). The random effect structure included a random intercept for participants as well as random slopes for Continuum Step over participants. This is the maximal random effect structure (Barr, Levy, Scheepers, & Tily, 2013), as a random slope of Exposure Condition over participants is not meaningful (Exposure Condition was varied between participants). This first analysis showed whether we replicated the perceptual learning effect found in many previous studies. More clicks on f-final words were expected after exposure to the ambiguous fricative in words in which it replaced /f/ (i.e., in the /f/-biased participant group) than after exposure to the ambiguous fricatives in words in which it replaced /s/ (i.e., in the /s/-biased group).

In a second analysis, we used the eye-tracking data to estimate at what point in time the learning effect influenced eye movements. In the literature on estimating the onset of an effect in timecourse data, two strategies are commonly used. The first one tests when an effect becomes significant by analyzing a number of adjacent time windows. This technique was used, for instance, by Van Turennout, Hagoort, and Brown (1998) with lateralized readiness potentials to compare the timecourse of phonological and syntactic encoding in speech production. We used this technique to compare the onset of the Continuum effect and the Exposure effect on a series of time window analyses of successive 100 ms bins, again using linear mixed effect models. Note that the successive windows are not fully independent of one another. However, Barr (2008) suggested that, for the analysis of eye-tracking data, time windows of 50 ms already are not too strongly dependent to allow for a timecourse analysis. Moreover, the point of the analysis is not to show that there is perceptual learning (this should already be evident in the analysis of the click responses), but rather from what point in time onwards this effect is robust.

The dependent variable was the logOdds transformed proportion of fixations on f-final words, normalized by the sum of the proportion of looks to s-final and f-final words. The logOdds transformation helps to prevent artifacts due to the limited range of proportions and the related co-dependence of mean and variance in raw proportions (Dixon, 2008; Jaeger, 2008). This dependent variable requires some correction when either one or both the proportion of /s/ and /f/ responses are zero. If one of the proportions is zero, the normalized proportion is zero or one, which would transform into negative and positive infinity when logOdds transformed. These values were therefore replaced by 1/48 and 47/48, based on the recommendation given in Signal Detection Theory (Macmillan & Creelman, 1991): These values are the mean between, on the one hand, the extremes of zero and one, and, on the other, the next highest or lowest observable value with 24 samples in a 100 ms time window (i.e., 1/24 and 23/24 or 2/48 and 46/48). If both proportions of looks to /s/ and /f/ are zero, a division by zero occurs in (p(f)/[p(f) + p(s)]). These cases were replaced by 0.5, indicating that there was no preference for either of the words.[3]

This dependent variable was predicted with the same fixed effects and random effect structure as in the analysis of click responses (fixed effects: Exposure Group, between participants, and Continuum Step within participants; maximally specified random effects structure over participants). We determined which time window was the first in which the signal (i.e., the factor Continuum Step; more /f/-like or /s/-like stimulus) and the Exposure Group (/f/-biased group,

[2] A note on notation: We use the notation "/s/-final word" for spoken words and "s-final" words for printed words, since /s/ necessarily refers to sound.

[3] This can be justified by considering the limit when both proportions approach zero, that is, both p(s) and p(f) are substituted by $^1/_n$ with $n$ approaching infinity: $\lim_{n\to\infty}\left(\frac{1}{n}/\frac{1}{n}+\frac{1}{n}\right)=\lim_{n\to\infty}\left(\frac{1}{n}*\frac{n}{2}\right)=\lim_{n\to\infty}\left(\frac{1}{2}\right)=\frac{1}{2}$.

/s/-biased group) significantly influenced fixations. The critical question here was whether the onset of the signal-related effect preceded the onset of the group/exposure effect.

The second strategy estimated when the effects of Continuum Step and Exposure Group reached a given proportion of their maxima following McMurray et al. (2008). We used maxima from 10% to 40% in 10% steps. Since eye-tracking data from individual participants are not reliable enough to estimate these time points, we employed a Jackknife procedure, estimating these time points by using samples of data from all participants minus one. This was done for both the signal and the exposure effect. With 24 participants, this gave us 24 estimates for both the signal and the exposure effect. The difference between these estimates was then compared with a *t*-test, for which the standard error estimate was multiplied by $n-1$ (i.e., 23) to account for the fact that each participant contributed $n-1$ times to the mean estimate.

### Results

#### Exposure

Table 1 shows that participants overwhelmingly accepted the ambiguous items as words in the lexical decision task. Previous studies (Norris et al., 2003) rejected data from participants who accepted less than 50% of the ambiguous tokens as words. All participants in the current study passed this criterion.

#### Click responses

Fig. 1 shows the proportions of clicks on f-final words in the different conditions, revealing a clear learning effect with more /f/ responses by participants in the /f/-biased group who had heard the ambiguous fricative in /f/-final words during exposure. In fact, three participants in the /f/-bias group always clicked on the f-words. The statistical analysis with Continuum Step (centered on zero) and Exposure Group (with the /s/-bias group mapped on −0.5 and the /f/-bias group on 0.5) as fixed factors and participant as random factor (including random slopes for all fixed factors varying over participants) confirmed this with a significant effect of Continuum Step ($b = 1.25$, $SE = 0.1$, $p < .001$) and a significant effect of Exposure Group ($b = 3.8$, $SE = 1.0$, $p < .001$). Note that the clicks on f-final words were coded as 1, so a positive regression weight indicates more clicks on f-final words, while a negative regression weight indicates fewer clicks on f-final words. The results hence replicate the perceptual learning effect reported in many earlier studies.

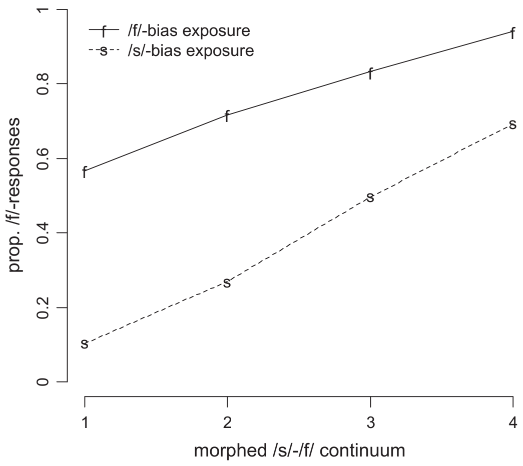

**Fig. 1.** Experiment 1: Mean proportion of /f/ responses (*y*-axis) depending on Exposure Group (solid vs. dotted line) and Continuum Step (*x*-axis; 1 = most /s/-like step, 4 = most /f/-like step).

#### Eye movements

Fig. 2 shows the eye-tracking data from the test phase. This includes data from three participants (from the /f/-bias group), who always clicked on /f/. Inspection of their eye-tracking data nevertheless showed that they were influenced by the phonetic properties of the stimuli, so that they directed their eye gaze faster towards f-final words if the stimulus was more /f/-like. Fixation proportions on the printed minimal pairs (*y*-axis) are plotted over time (*x*-axis) with zero being the onset of the frication noise. Word onset ranged from −350 ms to −272 ms before this point and the frication noises had durations of 200–250 ms. (This may seem quiet long, but one has to consider that these fricatives received to phrase-final lengthening.)

Fixation proportions are plotted separately for the printed f-final and s-final words of the minimal pair that was used on a given trial. Fixations to the distractor pair were averaged. Note that at the onset of the fricative, participants have already processed the onset and the vowel of the word. Hence already at this point in time more fixations were directed towards the target word pair than the distractor pair. That is, having heard [ro...], participants looked more at *roof* and *roos* (solid and dashed lines) than at the distractors (e.g., *hals* and *half*, dotted lines) at frication onset. The left panel of Fig. 2 shows the effect of Exposure Group: The f-words (solid lines) received more looks from the participants in the group with /f/-biased exposure (dark lines) than the participants with /s/-biased exposure

**Table 1**
Proportion of word responses to the critical items in the exposure phases of Experiments 1 and 2. The bold numbers indicate the proportion of word responses to the critical items in their ambiguous form.

| Stimulus | Exposure group | | | |
|---|---|---|---|---|
| | /f/-biased | | /s/-biased | |
| | Experiment 1 (%) | Experiment 2 (%) | Experiment 1 (%) | Experiment 2 (%) |
| /f/-final words | **96.3** | **95.0** | 99.6 | 96.3 |
| /s/-final words | 96.3 | 97.1 | **95.4** | **95.8** |

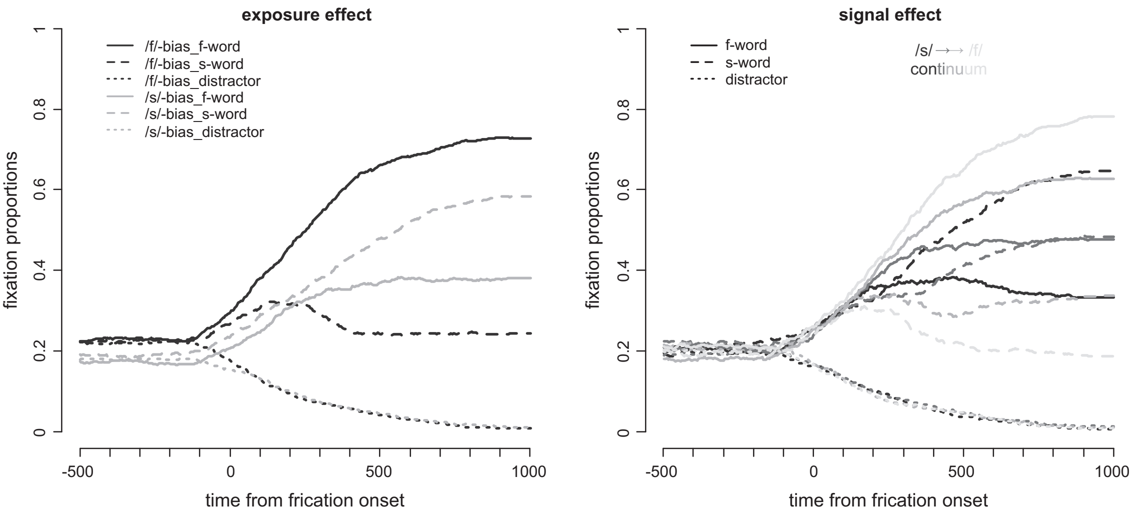

**Fig. 2.** Experiment 1: Fixation proportions for the different words. The left panel shows the effect of exposure, with more looks to f-final words by the /f/-biased than the /s/-biased group but an opposite preference for the s-final words. The right panel shows the effect of the /s/-/f/ continuum with more looks to the f-final words for more /f/-like steps of the continuum (i.e., the lighter lines).

(light lines), while the opposite was observed for the s-words (dashed lines). The right panel shows the fricative continuum with more looks to the f-final words (solid lines) for more /f/-like stimuli (dark and light gray lines).

Fig. 2 shows that the signal (i.e., Continuum Step) influenced the eye movements from about 150 ms after frication onset. Before that time, the lines for the different continuum steps are more or less on top of each other. At 150 ms, the lines for the different stimuli in the right panel start to diverge. The group effect (left panel), however, is already present at frication onset and even earlier: Participants in the /s/-bias group look more at the s-words prestimulus (the light dashed line is consistently above the light solid line), while there is no visible preference for the s-words or f-words for participants in the /f/-bias group. Note that the group effect is necessarily defined as the difference *between the groups* in the relative attractiveness of the two words and not as a preference for a word type within one group.

Additionally, the groups differed in their overall likelihood to fixate on any of the items; but note that the dependent variable (logOdd(p(f)/[p(f) + p(s)]) normalizes for this difference. The difference in overall likelihood to fixate any word is due to one participant in the /s/-bias group, who unlike all other participants, tended not to scan the array of printed words (fixation on any of the words <50% in the pre-stimulus time window, >90% for all other participants). Given that this is not a systematic difference between the groups, it is not surprising that this effect was not significant ($t(12) = 2.03$, $p = .06$). It rather reflects differences in scanning strategies in the pre-stimulus baseline between the participants.

Although listeners are likely to have some information about the upcoming fricative available due to formant transitions from the preceding vowel at the onset of the frication noise (point 0 ms), formant transitions are unlikely to be the cause for this effect. It usually takes about 150–200 ms for eye movements to reflect processing of the acoustic signal (Allopenna et al., 1998), which is much longer than the transitions in our words. Therefore, the early effect is more likely to be strategic; the /s/-biased group mostly clicked on s-final words, which led them to develop an expectancy to look at those words even before they heard the acoustic signal. That is, hearing [r…] and seeing the four words *roos*, *roos*, *hals*, and *half* on the screen, participants could rule out *hals* and *half* as potential targets on the basis of the acoustic input, and *roos* if they had developed an expectancy to click on the f-final words (due to the /f/-biased exposure condition).

As indicated above, the eye-tracking data were analyzed first with successive time-windows, evaluating the effects of Exposure Group and Continuum Step in 100 ms intervals from 300 ms before to 600 ms after frication onset. The dependent variable was the proportion of looks to the f-final words, normalized by the sum of looks to the f- and s-final word of the relevant minimal pair. Note that the effect of Exposure Group is defined as a difference in the relative attractiveness of the s- and f-words between groups, that is, it is a difference of differences measure. The difference between the looks to s-words and f-words within one group does not constitute the Exposure effect, as this difference confounds the Exposure effect with the overall stimulus quality which may be more /s/- or more /f/-like. Only the difference between the groups provides a measure of the Exposure effect that is not conflated with the stimulus properties. For the analysis, this proportion was transformed into logOdds. Fig. 3 shows the outcome of these different analyses. The effect of Exposure Group was significant even in the time window at the frication onset (0 ms) and from there onwards. Note that at time 0 ms, the information about fricative identity can in fact not yet influence the eye movements. This hence reflects an overall anticipation effect.

Note that it may seem surprising that the *t*-value for the effect of Exposure Group is relatively stable over the different time windows from frication onset, even though the

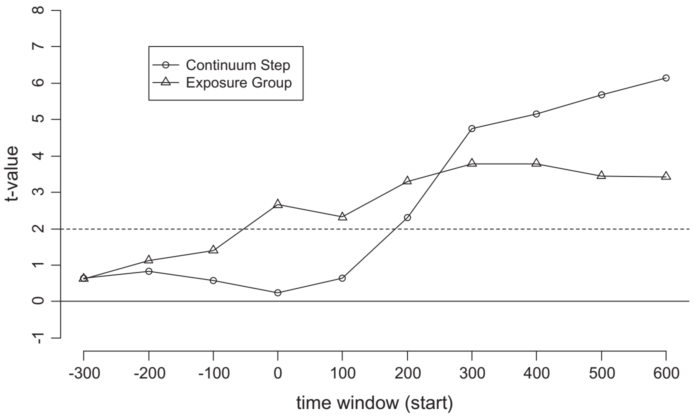

**Fig. 3.** Experiment 1: *t*-values from the mixed-effects models for the fixed factors Continuum Step and Exposure Group. The critical *t*-value of two is surpassed by the effect of Continuum Step around 200 ms after frication onset, while Exposure Group influences the fixation proportions already in earlier time windows.

group difference gets visibly larger over time (see the left panel of Fig. 2). This is reflected in the regression weight for the group difference, which rises from half a logOdds unit in the first time window to two logOdds units in the last time window. However, as the regression weight increases, so does the standard error, leading to a relatively stable *t*-statistic.

Given the anticipation effect, the time-window analysis is not able to provide an estimate of when the effect of exposure first influences the processing of the fricative. For the estimation of the onsets of the effects using a criterion of when they reach a given percentage of their maxima, however, we can at least numerically compensate for the anticipation effect. This can be achieved by normalizing the effect of Exposure Group to zero on the basis of its mean value in the time window starting at frication onset (see, Barr, 2008; Huettig & Altmann, 2007, for similar strategies). Fig. 4 shows the results of the Jackknife analysis for each subsample consisting of normalized data from all participants minus one.

Each subpanel of Fig. 4 shows data for the sample without the subject indicated in the subpanel's heading. The solid lines show the estimates of the exposure and the continuum effects over time. The continuum effect was calculated as the difference in the fixation preference for the f-final word (= proportion of looks to the f-final words minus proportion of looks to the s-final word) for the most /f/-like and the most /s/-like stimulus. The exposure effect was calculated as the difference in preference for the f-final word between the /f/-biased group and the /s/-biased group. Note that the theoretical maximum of the effects is two, with preferences ranging from −1 to 1. (If all participants in the /s/-biased group looked only at the s-final words, their preference for the f-final words would be −1, the difference in proportion of looks to f-final words [=0] and looks to s-final words [=1].) The dotted lines show 20% of the maximum effects. The average intersection of the dotted and solid lines is at 229 ms after frication onset for the exposure effect and at 278 ms for the continuum effect. The difference between these estimates was not significant ($-1 < t(20)_{\text{corrected}} < 1$) suggesting that the effects of Continuum Step and Exposure Group occur simultaneously. Also for the other criteria (10%, 30%, and 40%), there was no significant difference between the onset of continuum effect and the exposure effect ($t_{\max} = 1.43$).

### *Discussion*

The purpose of Experiment 1 was to test whether the results of lexically-guided perceptual learning influence first-pass phonetic processing or only come into play after initial first-pass phonetic processing. To this end, we first replicated the perceptual learning effect in the click-response data: The group with an /f/-biased exposure perceived the ambiguous fricatives during the test phase as /f/ more often than the group with an /s/-biased exposure.

In order to test whether the exposure influenced first-pass phonetic processing, we used the timing of the effect of Continuum Step as a benchmark. Based on earlier results with the visual-world paradigm (Allopenna et al., 1998), we assumed that the time point at which differences in the properties of the speech signal influence the eye movements (with more fixations on f-final words if the fricative spectrum is more /f/-like) reflects first-pass phonetic processing. We observed that the signal properties of the fricatives influenced eye movements around 200 ms after frication onset. This suggests that the effect occurred at the earliest possible point in time, given a 200 ms estimate to program and launch an eye movement. Moreover, the timing is in line with previous finding on the use of phonetic information in the visual-world paradigm (see Allopenna et al., 1998; and McQueen & Viebahn, 2007, for paradigms with printed words).

Estimating the onset of the exposure effect, however, was complicated by an anticipation effect. As Fig. 1 shows, there was quite a large difference in the overall proportion of /f/ and /s/ responses between the groups. Even though this is the effect of interest, that is, the consequence of exposure to an ambiguous fricative in /s/- or /f/-final words, the effect is so large that it led to an anticipation effect. The /f/-bias exposure group anticipated clicking on the /f/-final words more often, and already looked

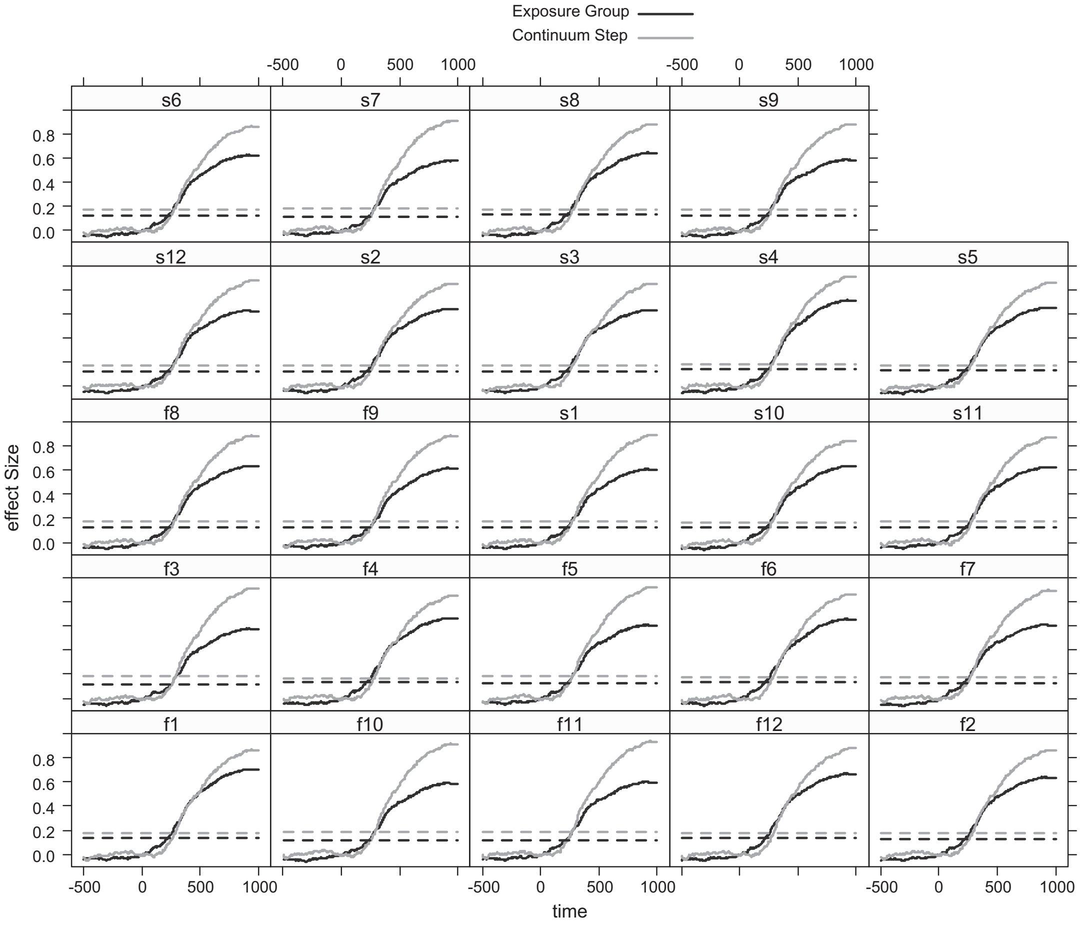

**Fig. 4.** The results of the Jackknife procedure for each subsample excluding one participant in Experiment 1. Overall, exposure effect (dark lines) and the signal effect (light lines) reach 20% of their maxima (the dotted lines) at roughly the same point in time.

significantly more often at the /f/-final words even before the critical fricative was heard. The time course analysis using time windows was hence not able to answer our main question.

In the second analysis of the time course, we tested when the effects reached a given criterion of their maximum. Here it was possible to correct for the anticipation by "baselining" the effects at the onset of the frication (see Barr, 2008; Huettig & Altmann, 2007, for similar approaches in eye-tracking). With this strategy, it seemed that the effect of exposure comes into play at the same time as the fricative is processed, that is, as participants are processing an ambiguous sound during first-pass phonetic processing. However, this conclusion would be stronger if we could show that it holds even when participants do not anticipate the most likely response. Therefore, in a second experiment we tried counteract the anticipation bias.

## Experiment 2

The purpose of this experiment was to test the effect of perceptual learning in a visual-world eye-tracking task while avoiding an anticipation bias. Experiment 1 has shown that participants in the group with an /f/-biased exposure tend to anticipate their clicks on the f-final words, and as a result, they looked towards this word even before the fricative was heard. To prevent this, we now presented the two exposure groups with only partially overlapping /s/-/f/ continua, as shown in Table 2. The /s/-biased group heard one more /f/-like stimulus, and the /f/-biased group heard one more /s/-like stimulus. In this way, the overall proportion of clicks on f- and s-final printed words should be more similar in both groups: The /f/-biased exposure should lead participants to click

**Table 2**
Design of the test phase in Experiment 2. In order to roughly equate the number of /s/ and /f/ responses in both exposure groups, the /f/-biased group was presented with one more /s/-like stimulus on the continuum than the /s/-biased group.

| Exposure | Continuum step (/s/-to-/f/) | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| /s/-bias | – | X | X | X | X |
| /f/-bias | X | X | X | X | – |

more on the f-final words, but this should be offset by hearing more /s/-like tokens, which in turn should evoke more clicks on s-final words. This should hence prevent the anticipation observed in Experiment 1.

### Method

#### Participants

Twenty-four (19 female/5 male) participants from the same population as in Experiment 1 participated in the experiment for pay. They were between 18 and 29 years old. None reported any hearing problem and all had normal or corrected-to-normal vision.

#### Materials and procedure

The materials and procedure were the same as in Experiment 1 with the exception that the test continua were shifted for the two exposure groups (see Table 2). For the /f/-biased group the test continua were moved two steps down along the overall 11 step continua. For the /s/-biased group the test continua were shifted one step in the same direction. Note that there was an overall /f/-bias in Experiment 1 (see Fig. 1). This is why we selected overall more /s/-like tokens for Experiment 2. By using non-overlapping continua, we also tried to equate the number of clicks on s- and f-final words across exposure groups.

#### Analyses

The same analyses were carried out as in Experiment 1. Due to the shift in test continua, all analyses were restricted to the three steps of the /s/-/f/ continua that were presented to both groups. First, an overall analysis of the click responses tested for the presence of the perceptual learning effect. Second, a time window analysis tested for the onset of the exposure and the continuum effect. Finally, a Jackknife method was employed to provide another estimate whether the two effects influenced the eye-movements at similar or different points in time (i.e., as measured by the point at which the effects reached 10%, 20%, 30%, or 40% of their maxima).

### Results

Table 1 shows that during exposure participants overwhelmingly accepted the ambiguous items as words, with a slight tendency for more word responses to unambiguous items, similar to Experiment 1. All participants accepted more than 50% of the ambiguous tokens as words.

Fig. 5 shows the proportions of clicks on f-final words, showing a clear learning effect with more /f/ responses by the /f/-biased group than the /s/-biased group for the overlapping part of the continua presented to both groups. The statistical analysis with Continuum Step (centered on zero) and Exposure Group (with the /s/-bias group mapped on −0.5, the /f/-bias group on 0.5) as fixed factors and participant as random factor (including random slopes for all fixed factors varying over participants) confirmed a significant effect of Continuum Step ($b = 1.24$, $SE = 0.1$, $p < .001$) and a significant effect of Exposure Group ($b = 2.1$, $SE = 0.5$, $p < .001$). Thus, again, we found an overall learning effect.

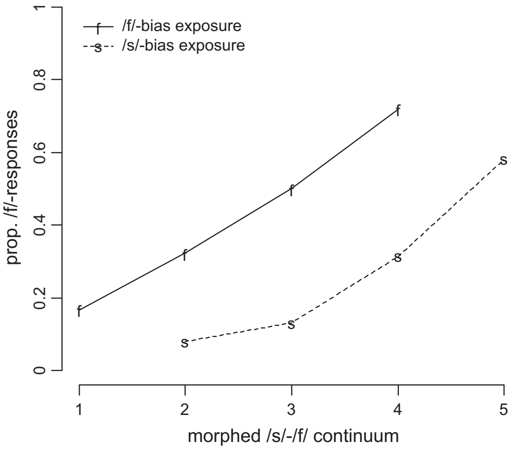

**Fig. 5.** Experiment 2: Mean proportion of /f/ responses depending on Exposure Group and Continuum Step (*x*-axis; 1 = most /s/-like step, 5 = most /f/-like step).

Fig. 6 shows the eye-tracking data from the test phase, again time-aligned to the onset of the frication noise. The left panel shows the exposure effect. The critical comparisons are the following: The f-final words (solid lines) were looked at more by the /f/-bias group than by the /s/-bias group (dark solid line above the light solid line), while the s-final words (dashed lines) were looked at more by the /s/-bias group than the /f/-bias group (dark dashed line below light dashed line). Note that the fact that the two dark lines for the /f/-bias group are close together only shows that the f-bias group did not show an overall preference for one or the other words. That is, collapsed across all steps of the fricative continuum, listeners in the /f/-bias group looked as much at the f-final words as at the s-final words. However, this does not mean that the effect of exposure is absent in half of the data as the critical comparison is *between groups* for a given type of word (i.e., for the f-final words, the dark vs. the light solid line and, for the s-final words, the dark vs. the light dashed line) rather than *between words* for a given group. The right panel shows the continuum effect with more looks to the f-final words (solid lines) for the more /f/-like stimuli (going from light to dark). This panel shows that the signal (i.e., continuum) influenced the eye-movements from about 150 ms after frication onset. In contrast to Experiment 1, the groups did not differ at frication onset. Both groups had a preference for the target pair over the distractor pair, but no preference for either the f-final or s-final word of the target pair. This allows for a more straightforward interpretation of the results of the time-course analyses than Experiment 1.

The results of the analysis of successive 100 ms time windows are shown in Fig. 7. The same linear mixed-effects models were used as described in Experiment 1. As in Experiment 1, the effect of Continuum Step reached significance in the time window starting at 200 ms after frication onset. The effect of Exposure Group reached significance in the same time window, suggesting no difference in the timing of these two effects.

The results of the time-window analysis were borne out by the results of the Jackknife procedure testing the time at

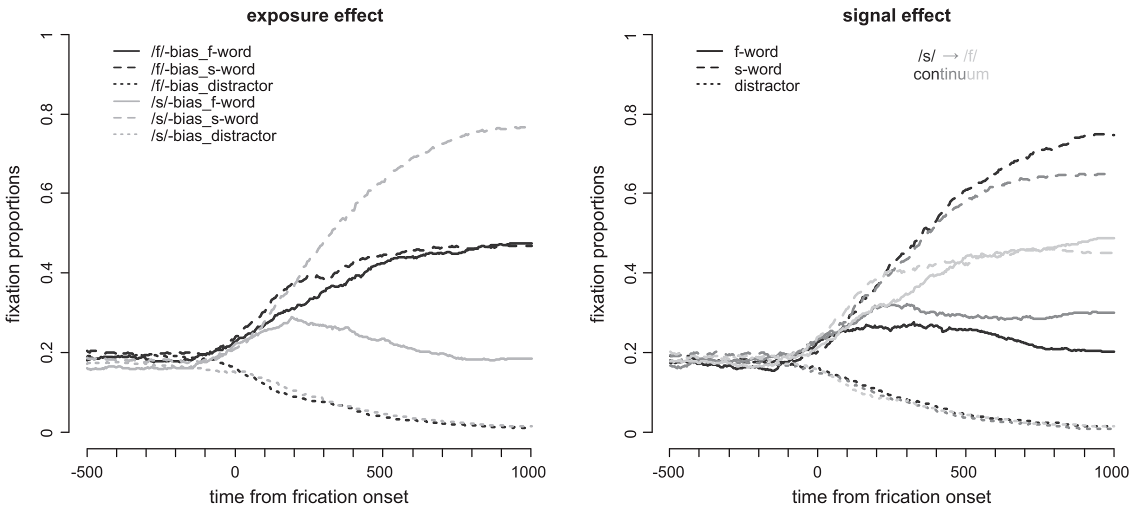

**Fig. 6.** Experiment 2: Fixation proportions for the different words. The left panel shows the effect of exposure, with more looks to f-final words by the /f/-biased group compared to the /s/-biased group and more looks to s-final words by the /s/-biased group compared to the /f/-biased group. The right panel shows the effect of the /s/-/f/ continuum with more looks to the f-final words (solid lines) the more /f/-like the stimuli (from dark to light).

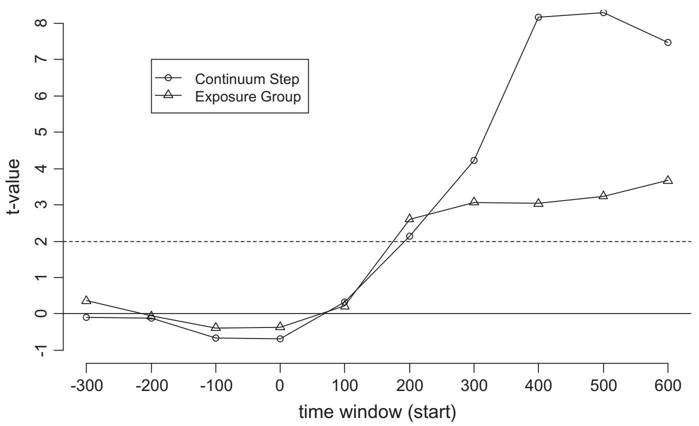

**Fig. 7.** Experiment 2: $t$-values from the mixed-effects models for the fixed factors Continuum Step and Exposure Group. The critical $t$-value of 2 is surpassed by the effects of Continuum Step and Exposure Group in the same time window, the one starting 200 ms after frication onset.

which a given percentage of the effects' maxima was reached. As thresholds, we again used 10%, 20%, 30%, and 40% of the maxima. Fig. 8 shows the results of the Jackknife analysis for each subsample of participants in Experiment 2, showing the 20%-of-maximum criterion. Given the absence of an anticipation effect, there was no need to correct the baseline for the effect of Exposure Group. Again, the solid lines show the estimates of the exposure and the continuum effect over time and the dotted lines show 20% of the maximum effects. Note that only the light dotted line is clearly visible in most of the panels. This is due to the nearly identical maxima of effects of Exposure Group and Continuum step, leading to overlapping lines for the 20% points of the maxima. The average intersect of the solid and dotted lines are at 265 ms after frication onset for the exposure effect and at 286 ms for the continuum effect ($-1 < t(23)_{\text{corrected}} < 1$). This was also the case for the other criteria (10%, 30%, or 40%, $t_{\max} = 0.35$).

*Discussion*

The purpose of the Experiment 2 was to replicate the time-course data observed in Experiment 1, but without the complication of the anticipation effect that was observed there. To achieve this, participants from the two exposure groups heard different steps of the same continua. This had the desired effect: Even though the /f/-biased group still clicked more often on the f-final words than the /s/-biased group on the overlapping part of the continuum, these responses were not anticipated. While Fig. 3 (with the data from Experiment 1) showed a difference between the groups already before frication onset, Fig. 6 (with the data from Experiment 2) shows no such difference between the groups until 200 ms after frication onset – the earliest point in time at which eye movements are expected to be driven by the speech signal. That is, in Experiment 2 there was no effect of anticipation for either

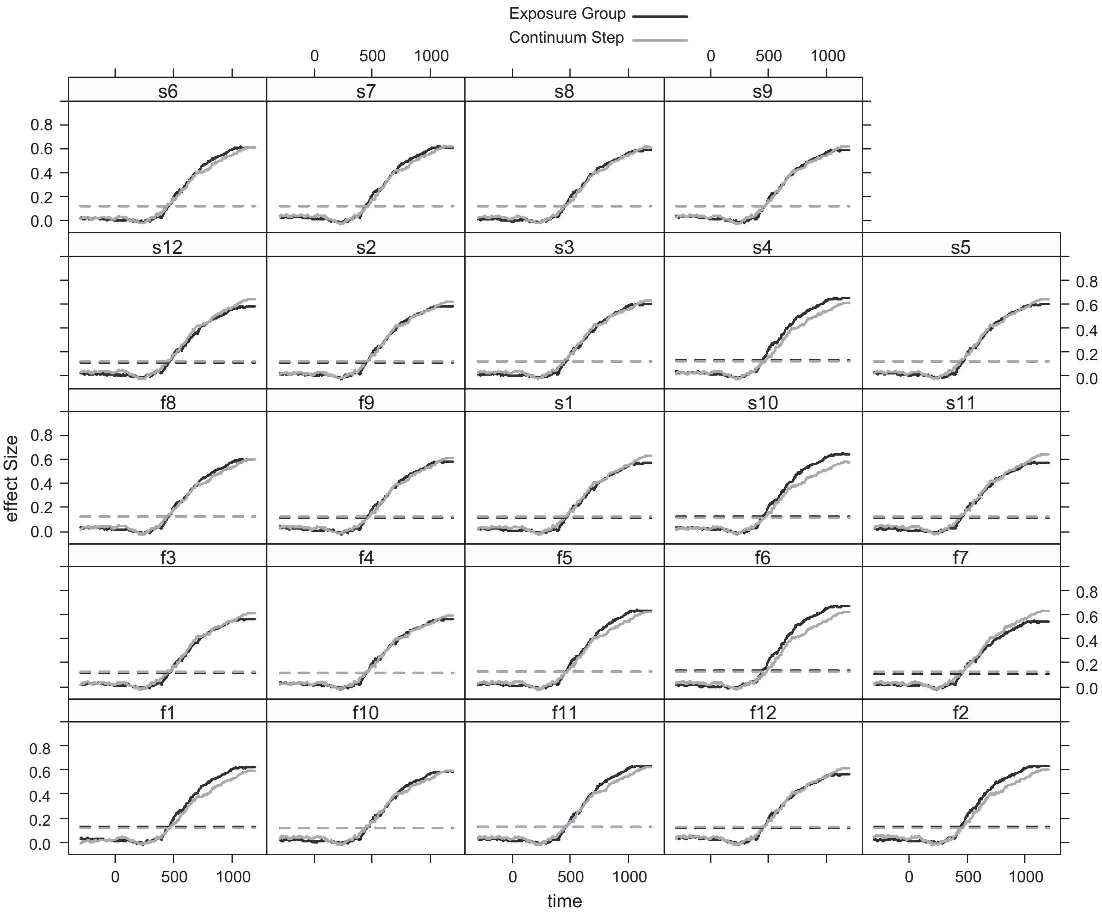

**Fig. 8.** The results of the Jackknife procedure for each subsample excluding one participant in Experiment 2. Overall, 20% of the maxima (dotted lines) were reached at roughly the same time for the exposure effect (dark lines) and the continuum effect (light lines).

of the exposure groups. In the absence of such an anticipation effect, both methods to estimate the timecourse of the perceptual learning effect lead to the same conclusion. The effect of perceptual learning exerts itself at the same time as the effect of the fricative signal.

There is, however, one potential problem with the presentation of different ranges of stimuli from an /s/-/f/ continuum to different groups of participants. This procedure might generate so-called range effects (Repp & Liberman, 1987). That is, participants tend to "home in" on the presented continuum, such that adding more prototypical examples of one phonetic category makes participants more likely to perceive ambiguous stimuli as contrasting with this prototype. These range effects, however, have mostly been reported in studies using a single nonword–nonword continuum (e.g., "ba"-"da"). The current experiment, in contrast, made use of five different word–word continua. Since range effects are often attributed to local contrast effects in which stimuli are compared from one trial to the next (Diehl & Kluender, 1987), the presentation of different word pairs should have reduced the likelihood of finding range effects between our groups. This is because it is easier to compare [ba] with [da] than it is to compare [ro:s] with [hɑlf]. Nevertheless, the presence of such contrast effects is straightforward to test: it can be done by replicating Experiment 2 without any exposure. If the differences between the groups are driven by the differences in the ranges of the continua, we should replicate the results obtained in Experiment 2 even without exposure.

## Experiment 3

### Method

#### Participants

Twenty-five (18 female/7 male) participants from the same population as in Experiment 1 participated in the experiment for pay. None of them had participated in the previous experiments. They were between 18 and 29 years old. None reported any hearing problem and all had normal or corrected-to-normal vision.

#### Materials and procedure

The materials and procedure matched the test phase of Experiment 2. One group of listeners was presented with one more s-like step of the continuum (i.e., matching the /f/-bias group in Experiment 2) whereas the other group was presented with one more /f/-like step (i.e., matching the /s/-bias group in Experiment 2). No exposure phase was given. However, for better comparison of the results with Experiment 2 we will refer to the two participant groups in terms of the group they matched in Experiment 2.

### *Results*

Fig. 9 shows the proportions of clicks on f-final words, with slightly more /s/ responses by the group that was exposed to the same range as the /f/-biased group in Experiment 2. There are overall fewer clicks on the f-words, reflecting the use of slightly more /s/-like tokens in comparison with Experiment 1. With this range, now all participants click on both f- and s-words. The statistical analysis over the shared continuum range with Continuum Step (centered on zero) and Exposure Group (with the /s/-bias group mapped on −0.5, the /f/-bias group on 0.5) as fixed factors and participant as random factor (including random slopes for continuum step which was varied over participants) confirmed a significant effect of Continuum Step ($b = 1.31$, $SE = 0.11$, $p < .001$) but no significant effect of Group ($b = -0.76$, $SE = 0.84$, $p > .1$). Note that the numerical difference between the groups is in the opposite direction to the group differences in the previous experiment.

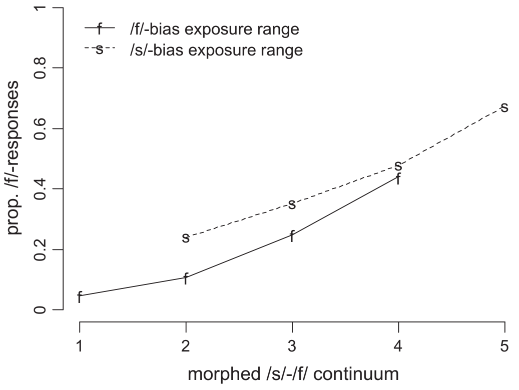

**Fig. 9.** Experiment 3: Mean proportion of /f/ responses depending on Continuum range and Continuum Step (*x*-axis; 1 = most /s/-like step, 5 = most /f/-like step).

Fig. 10 shows the eye-tracking data, again, time-aligned to the onset of the frication. The left panel shows the Group effect (i.e., the effect of continuum range). Overall, there are more looks to the s-final words (dashed lines) than to the f-final words (solid lines). This is in line with the click-response data, in which /s/ responses were overall more likely (see Fig. 9). The f-final words received more looks from the /s/-bias group (solid light line) than the /f/-bias group (solid dark line). The opposite pattern was observed for the looks to s-final words (dashed lines: dark > light). This numerical difference is in line with the behavioral data, which also showed a numerical preference for the s-final words by participants in the /f/-biased group.

The right panel shows the continuum effect with more looks to the f-final words (solid lines) and fewer looks to the s-final words (dashed lines) as the stimuli become more /f/-like (=lighter lines, the asymptotes for the solid lines, the f-final words, increase, while the asymptotes for thee dashed lines, the s-final words, decrease). The Figure shows that the signal (i.e., continuum) influenced the eye movements from about 200 ms after frication onset, because at that point in time, the lines for the different stimuli start to diverge.

The results of the analysis with linear mixed-effects models of successive 100 ms time windows are shown in Fig. 11. The effect of Continuum Step reached significance in the time window starting at 300 ms after frication onset and stayed significant for the rest of the time period. The

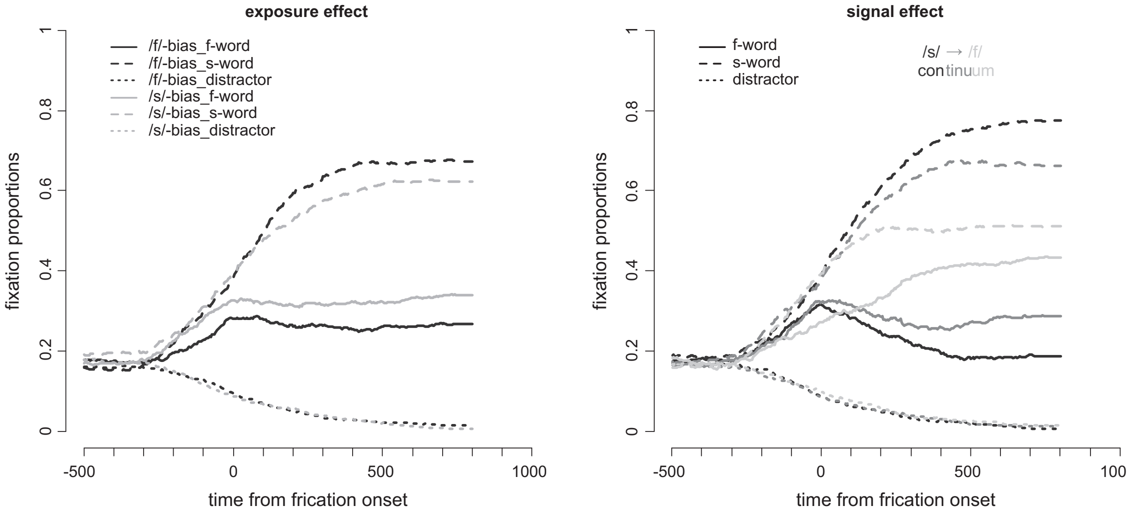

**Fig. 10.** Experiment 3: Fixation proportions for the different words. The left panel shows the effect of Group (i.e., continuum range), with more looks to f-final words for the group that matched the /s/-biased group in Experiment 2 and more looks to the s-final words for the group matching the /f/-biased group. The right panel shows the effect of the /s/-/f/ continuum with more looks to the f-final words (solid lines) and less looks to the s-final words (dotted lines) the more /f/-like the stimuli (dark-to-light).

Please cite this article in press as: Mitterer, H., & Reinisch, E. No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language* (2013), http://dx.doi.org/10.1016/j.jml.2013.07.002

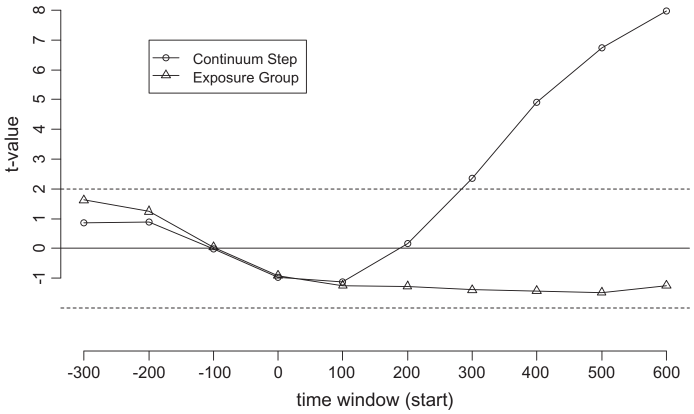

**Fig. 11.** *t*-Values from the mixed-effects models for the fixed factors Continuum Step and Exposure Group in Experiment 3.

effect of Group (continuum range) was not significant in any of the time windows.

*Discussion*

The purpose of Experiment 3 was to test whether differences between the groups in Experiment 2 could be explained by the presentation of different ranges of the fricative continua rather than by differences in exposure. In Experiment 2, the groups differed both in the exposure condition (/s/-bias vs. /f/-bias) and the ranges of the continua heard during test. Experiment 3 did not make use of the exposure condition so that the only remaining difference was the different range of the continua. Under these conditions, no significant group effect emerged in the click responses. If anything, there was a numerical difference in the opposite direction to that observed in Experiment 2.

The main purpose of Experiment 3 was to test whether the group differences in the test phase of Experiment 2 were driven by differences in the exposure phase or differences in the continua heard during test. The answer to this question is clear: The difference in the continua heard during test does not trigger a group difference as observed in Experiment 2; that is, we can be confident that the differences in Experiment 2 are due to the perceptual recalibration of the /s/-/f/ contrast during exposure.

Then why did we not find a range effect? Range effects tend to be described as quite robust in the literature (Harnad, 1987). A possibility already highlighted in the discussion of Experiment 2 is that range effects may be obliterated by the use of multiple continua. Diehl and Kluender (1987) argued that range effects may largely be attributed to local contrast effects, in which in a series of stimuli, stimulus $n$ is compared to stimulus $n - 1$. This leads to obvious contrasts when only CV syllables are involved (e.g., [ba] vs. [da]). In our case, participants had to decide between, for instance, [ro:s] and [ro:f] on one trial and [hɑls] and [hɑlf] on the next, so that a direct comparison was discouraged by different vowel contexts and more intervening speech material. It remains to be seen, whether such stimulus variability is always sufficient to prevent range effects from arising.

## General discussion

In three experiments, we used eye-tracking measures to establish the processing stage at which lexically-guided retuning of a fricative contrast affects perception. Two possibilities were evaluated: The knowledge gained during exposure might directly affect first-pass phonetic processing or it might be integrated with the results of this first-pass phonetic processing at a later decision stage. To compare these two hypotheses, we first presented participants in an exposure phase with ambiguous fricatives in lexically unambiguous contexts, such as the ambiguous fricative [$^{s}/_{f}$] in [ʃɪrɑ$^{s}/_{f}$] (/ʃɪrɑf/, "giraffe"), where, in Dutch, it can only be interpreted as /f/, since /ʃɪrɑs/ is not a word in Dutch. In the immediately following test phase, participants heard a range of ambiguous fricatives in *lexically-ambiguous* contexts, that is, in stimuli based on minimal pairs (e.g., [ro$^{s}/_{f}$], from *roos-roof*, "rose"-"robbery"). Participants saw both words on a computer screen with the instruction to click on the word they heard while their eye movements were tracked. Responses in the test phase revealed a learning effect: participants who had previously heard the ambiguous fricative replacing /f/ clicked on f-final words more often than participants who had heard the ambiguous fricative replacing /s/.

To address the main question of the early versus late integration of the retuned categories with first-pass phonetic processing, listeners' eye movements were analyzed. Two benchmarks were used to evaluate the time course of the perceptual learning effect. The first benchmark was "study-internal". We presented listeners with a range of stimuli from more /s/-like to more /f/-like stimuli to compare the onset of the learning effect with the onset of a signal-driven effect. Two different analysis methods were employed to evaluate the relative timing of the continuum and group effects: a successive time-window analysis, testing when these effects reached significance, and a Jackknife procedure, testing the time points at which the effects reached a given percentage (10–40%) of their respective maxima.

A second benchmark was provided by the absolute rather than relative timing of the perceptual-learning effect. The previous eye-tracking literature (Allopenna et al., 1998; McMurray et al., 2008) showed that the speech signal influences eye movements around 200 ms after being presented. Given that this lag is mostly due to the planning of the eye movement itself (Matin, Shao, & Boff, 1993), an onset of the perceptual learning effect at this point in time would reflect an influence on first-pass phonetic processing.

Perceptual learning effects were found in the eye-movement record throughout the study. In Experiment 1, the perceptual learning effect was quite strong. For instance, three participants in the /f/-bias group heard all fricatives as /f/. Due to this strong learning effect, the group difference triggered an additional strategic anticipation effect. Participants with /s/-biased exposure learned that they mostly clicked on the s-final words and therefore looked at these words more than the /f/-biased group, even before the fricative was heard. This led to a significant anticipation effect in the timecourse analysis. Therefore, in the analysis testing when the effects reach a certain percentage of their respective maxima, we normalized for this anticipation effect. We found that, with this correction, the effect of Exposure Group emerged at the same point in time as the effect of Continuum Step.

To further explore the timing of the exposure effect, Experiment 2 introduced a procedural change that managed to control this anticipation strategy: Participants from the /f/-biased exposure group heard more /s/-like tokens over the course of the experiment than the /s/-biased group. This had the desired effect and virtually eliminated the anticipation effect, as it diminished the difference in the overall number of /s/ and /f/ responses between the groups. Nevertheless, an effect of exposure similar to that observed in Experiment 1 was still observed. For the stimuli that allow a group comparison (i.e., those presented to both groups) the group with /f/-bias exposure gave more /f/-responses than the /s/-bias exposure group. This allowed us to interpret the results of both the time-window analysis and the analysis of when the effects reached a given percentage of their maximum. In both cases, the effect of Exposure occurred at the same point in time as the effect of Continuum Step. This is in line with the assumption that the knowledge gained from lexically-guided phonetic retuning during exposure is integrated with incoming phonetic information at an early processing level, that is, during first-pass phonetic processing. The absolute criterion set above leads to a similar conclusion. The effect of exposure arose in the time window 200–300 ms after frication onset, that is, at a point in time during which the eye-tracking record provides a window on early phonetic processing.

A third experiment tested an alternative interpretation of the results in Experiment 2. In Experiment 2, the groups not only differed in the exposure condition but also in the range of continua heard during the test phase. Experiment 3 therefore tested whether a difference in the range of the continua presented during phonetic categorization (i.e., the test phase in Experiment 2) was sufficient to trigger the same group differences as observed in Experiment 2. The results showed that it clearly was not. There was no significant difference between the groups in Experiment 3 in terms of the number of /f/-responses on the overlapping part of the continuum. A slight numerical difference between groups was in the opposite direction compared to the group difference obtained in Experiment 2. The absence of a stable effect of group in Experiment 3 rules out the potential confound of continuum range. The results of Experiment 2 thus indeed reflect lexically-guided perceptual learning and show that the effects of perceptual learning are integrated early with the incoming speech signal.

Our conclusions partly rest on the assumption that eye movements starting at 200 ms after hearing an acoustic cue reflect first-pass phonetic processing of this cue. A recent paper (Altmann, 2011) argues that eye movements may already be influenced by the speech signal after only 100 ms, which would mean that processes 200 ms after the presentation of a cue may in fact not provide a measure of early first-pass phonetic processing. However, the estimate of 100 ms is based on a task in which participants listened to sentences and saw pictures with possible agents (e.g., “the girl”, “the women”, “the man”, etc.). Participants apparently looked towards these agents already 100 ms after the onset of the agents’ name. However, this onset may be difficult to determine, since the agent is preceded by the determiner “the”. The schwa at the end of the determiner is a segment that is strongly colored by the surrounding segments and hence carries information about the upcoming segment much more than other segments. This coarticulatory information that is available before the point where a phonetician would set the boundary between the schwa and the following consonant may have influenced the eye movements. Altmann acknowledged this problem but argued that similar coarticulatory anticipation should then be observed in other experiments. However, this argument is based on the oversimplifying assumption that all coarticulation effects are equal. The agent *man* used in the study by Altmann affords a lowering of the velum (to produce the nasal /m/) while the alternative agent *the girl* does not. Anticipatory velum lowering can arise quite early, leading to a pronunciation of the determiner *the* with nasalization throughout the word, which in turn can be used by listeners (Ohala & Ohala, 1995). This leads to much stronger coarticulatory differences between the stimuli than in other studies in which targets differed in phonetic features with much weaker coarticulatory cues. Further research with phonetically well-controlled materials will be necessary to claim that language can mediate eye movements at such short latencies. In a similar vein, others (Salverda & Tanenhaus, 2012) have also argued that such early effects may be due to coarticulation.

Based on these considerations we can confidently assume that eye movements 200 ms after the onset of a stimulus (note that onsets are notoriously difficult to define with speech stimuli) are likely a reflection of early perceptual processing. In this timeframe, eye movements were influenced by the phonetic properties of the stimuli and

Please cite this article in press as: Mitterer, H., & Reinisch, E. No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language* (2013), http://dx.doi.org/10.1016/j.jml.2013.07.002

perceptual learning at about the same point in time. Our data hence support the conclusion that perceptual learning affects early phonetic processing. However, this conclusion is based on finding no difference in the time course between the effect of phonetic properties and exposure bias. Traditionally, psychology has had the tendency to disregard findings of no difference as uninformative null results, a tendency which caused others to view psychology as a second-rate science (Fanelli, 2010). While the evaluation of null-results may be changing, the problem remains that current statistical techniques make it difficult to assess the validity of the null-hypothesis.

At this juncture, it is fortunate that findings from different experimental paradigms also suggest an early locus of perceptual learning in speech. Clarke-Davidson et al. (2008) used signal-detection techniques, and came to the same conclusion as we did, varying both the task at test (identification and discrimination) as well as the task during exposure. Individually, the efforts of Clarke-Davidson et al. and ours are subject to possible alternative interpretations. As laid out in the introduction, there has been a lively debate as to what extent signal-detection measures can differentiate perceptual effects from decision effects. Our conclusion is based on a no-difference finding. However, in combination, both results are well explained by the assumption that perceptual learning affects early phonetic processing. The alternative hypothesis now requires quite a few "auxiliary assumptions" to be upheld, so that it becomes, in the combination of the results, close to untenable.

One serendipitous finding of the current study is that participants may anticipate referents in an eye-tracking paradigm based on the task setting (Barr, 2008). In Experiment 1, we observed differences between the groups in time windows before the fricative could influence the eye movements. This was most likely due to anticipation; the groups already looked more at the member of the minimal pair they were more likely to click on over the course of the experiment. In Experiment 1, the /s/-bias group looked more at the s-final words, because they were more likely to click on these words due to the perceptual-learning effect. As this shows, it seems crucial in visual-world eye-tracking experiments to eliminate all *incidental* cues about the identity of the eventual target.

How do our findings relate to early versus late integration of other information sources, such as phonological and speaker context? The fact that phonological context influences the interpretation of a given speech signal has long been known as "compensation for coarticulation" (Mann, 1980), which has more recently been extended to "compensation for assimilation" (Gaskell & Marslen-Wilson, 1996; Mitterer & Blomert, 2003). The timecourse of these effects has been studied with eye tracking and electrophysiological measures. Gow and McMurray (2007) used eye tracking and found that the context effect arises rather late, occurring more than 500 ms after onset of the context. This suggested a late effect of context in compensation for assimilation. However, this conclusion contrasted with three independent data sets using electrophysiological measures that demonstrated early context effects in compensation for assimilation, that is, in a time window 100–300 ms after the context (Gow & Segawa, 2009; Mitterer & Blomert, 2003; Mitterer, Csépe, Honbolygo, & Blomert, 2006). These three studies indicate that phonological context is integrated early with the incoming speech signal. One possible reason why the eye-tracking study did not reveal early effects may be due to the nature of the paradigm. If participants look at a particular object, they may be more likely to perceive the incoming speech in line with this object, just as looking at the written syllable /ba/ makes it slightly more likely to hear an ambiguous syllable as /ba/ (Massaro, Cohen, & Thompson, 1988).

Our data shows that lexically-guided perceptual learning of speaker idiosyncrasies also influences early stages of processing. It should be stressed, though, that our finding of "no delay" in the use of the newly acquired knowledge is contingent on the presence of another delay, the delay between exposure and test. We do not argue that listeners immediately change their representations based on lexical feedback when they first hear an unusual pronunciation. Instead, noticing a mismatch between the lexically prescribed sound and the actually perceived sound is crucial for learning to occur. The data by Poellmann et al. (2011) show that listeners indeed need multiple presentations of an ambiguous fricative for perceptual learning to occur. Once this learning is completed, however, acoustic cues are already interpreted in the light of the known pronunciation variants. Then, there is no delay in the application of perceptual learning in speech perception.

## Acknowledgments

We would like to thank Wendy van Ginkel for help with running the experiments and Dan Hufnagle, James McQueen, Ian Thornton and three reviewers for helpful comments on earlier versions of this manuscript.

## References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*, 419–439. http://dx.doi.org/10.1006/jmla.1997.2558.

Altmann, G. T. M. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica, 137*, 190–200. http://dx.doi.org/10.1016/j.actpsy.2010.09.009.

Barr, D. J. (2008). Analyzing "visual world" eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*, 457–474. http://dx.doi.org/10.1016/j.jml.2007.09.002.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001.

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

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.

Colin, C., Radeau, M. A. S., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk–MacDonald effect: A phonetic representation within short-term memory. *Clinical Neurophysiology, 113*, 495–506.

Diehl, R. L., & Kluender, K. R. (1987). One the categorization of speech sounds. In S. Harnard (Ed.), *Categorical perception: The groundwork of cognition* (pp. 226–253). Cambridge, Mass: Cambridge University Press.

Dixon, P. (2008). Models of accuracy in repeated-measures design. *Journal of Memory and Language, 59*, 447–456.

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.

Fanelli, D. (2010). "Positive" results increase down the Hierarchy of the Sciences. *PLoS One, 5*, e10068. http://dx.doi.org/10.1371/journal.pone.0010068.

Fowler, C. A., Brown, J. M., & Mann, V. A. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 877–888. http://dx.doi.org/10.1037//0096-1523.26.3.877.

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

Gaskell, M. G. (2003). Modeling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics, 31*, 447–463. http://dx.doi.org/10.1016/S0095-4470(03)00012-3.

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 144–158. http://dx.doi.org/10.1037//0096-1523.22.1.144.

Gow, D. W., & McMurray, B. (2007). Word recognition and phonology: The case of English coronal place assimilation. In J. Cole & J. Hualde (Eds.). *Laboratory phonology* (vol. 9, pp. 173–200). New York: Mouton de Gruyter.

Gow, D. W., & Segawa, J. A. (2009). Articulatory mediation of speech perception: A causal analysis of multi-modal imaging data. *Cognition, 110*, 222–236. http://dx.doi.org/10.1016/j.cognition.2008.11.011.

Hanulíková, A., van Alphen, P. M., van Goch, M. M., & Weber, A. (2012). When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience, 24*, 878–887. http://dx.doi.org/10.1162/jocn_a_00103.

Harnad, S. (1987). *Categorical perception: The groundwork of cognition*. Cambridge.: Cambridge University Press.

Huettig, F., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition, 15*, 985–1018. http://dx.doi.org/10.1080/13506280601130875.

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.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics, 27*, 359–384.

Kawahara, H., Masuda-Katsuse, I., & de Cheveigné, A. (1999). Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency based F0 extraction. *Speech Communication, 27*, 187–207. http://dx.doi.org/10.1016/S0167-6393(98)00085-5.

Kingston, J., & Macmillan, N. A. (1995). Integrality of nasalization and f1 in vowels in isolation and before oral and nasal consonants – A detection-theoretic application of the Garner paradigm. *Journal of the Acoustical Society of America, 97*, 1261–1285.

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

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 27*, 98–104.

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences, 23*, 571–579. http://dx.doi.org/10.1016/s0166-2236(00)01657-x.

Macmillan, N. A., & Creelman, D. (1991). *Detection theory: A user's guide*. Oxford: Blackwell Publishers.

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics, 28*, 407–412.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

Massaro, D. W., Cohen, M. M., & Thompson, L. A. (1988). Visible language in speech perception: Lipreading and reading. *Visible Language, 22*, 8–31.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics, 53*, 372–380. http://dx.doi.org/10.3758/BF03206780.

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review, 15*, 1064–1071. http://dx.doi.org/10.3758/pbr.15.6.1064.

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.

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

McQueen, J. M., & Viebahn, M. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology, 60*, 661–671. http://dx.doi.org/10.1121/1.419865.

Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics, 65*, 956–969. http://dx.doi.org/10.3758/BF03194826.

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

Mitterer, H., Csépe, V., & Blomert, L. (2006). The role of perceptual integration in the recognition of assimilated word forms. *Quarterly Journal of Experimental Psychology, 59*, 1395–1424. http://dx.doi.org/10.1080/17470210500198726.

Mitterer, H., Csépe, V., Honbolygo, F., & Blomert, L. (2006). The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science, 30*, 451–479. http://dx.doi.org/10.1207/s15516709cog0000_57.

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

Moore, B. C. J. (2003). *An introduction to the psychology of hearing*. Acad. Press.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*, 299-324. http://dx.doi.org/ 10.1017/S0140525X00003241.

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.

Ohala, J. J., & Ohala, M. (1995). Speech perception and lexical representation: The role of vowel nasalization in Hindi and English. In B. Cornell & A. Arvanti (Eds.), Phonology and phonetic evidence. Papers in laboratory phonology IV (pp. 41–60). Cambridge, UK: Cambridge University Press.

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

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.

Repp, B. H., & Liberman, A. M. (1987). Phonetic categories are flexible. In Stevan Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 89–112). Cambridge, Mass: Cambridge University Press.

Salverda, A. P., & Tanenhaus, M. K. (2012). Very fast effects of language on eye-movement control are due to anticipatory coarticulatory information. Presented at the AMLaP 2012, Riva Del Garda.

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. (2011). Listening to different speakers: On the time-course of perceptual compensation for vocal-tract

Please cite this article in press as: Mitterer, H., & Reinisch, E. No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language* (2013), http://dx.doi.org/10.1016/j.jml.2013.07.002

characteristics. *Neuropsychologia, 49*, 3831–3846. http://dx.doi.org/10.1016/j.neuropsychologia.2011.09.044.

Stekelenburg, J. J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience, 19*, 1964–1973.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29*, 557–580.

Van Turennout, M., Hagoort, P., & Brown, C. M. (1998). Brain activity during speaking: From syntax to phonology in 40 milliseconds. *Science, 280*, 572–574.

Werker, J. F., & Tees, R. C. (1984). Cross-Language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49–63. http://dx.doi.org/10.1016/S0163-6383(84)80022-3.

Please cite this article in press as: Mitterer, H., & Reinisch, E. No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language* (2013), http://dx.doi.org/10.1016/j.jml.2013.07.002