Perceptual learning in speech

Dennis Norris,[a],* James M. McQueen,[b] and Anne Cutler[b]

[a] *MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, UK*
[b] *Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

Abstract

This study demonstrates that listeners use lexical knowledge in perceptual learning of speech sounds. Dutch listeners first made lexical decisions on Dutch words and nonwords. The final fricative of 20 critical words had been replaced by an ambiguous sound, between [f] and [s]. One group of listeners heard ambiguous [f]-final words (e.g., [wɪtlo?], from *witlof*, chicory) and unambiguous [s]-final words (e.g., *naaldbos*, pine forest). Another group heard the reverse (e.g., ambiguous [na:ldbo?], unambiguous *witlof*). Listeners who had heard [?] in [f]-final words were subsequently more likely to categorize ambiguous sounds on an [f]–[s] continuum as [f] than those who heard [?] in [s]-final words. Control conditions ruled out alternative explanations based on selective adaptation and contrast. Lexical information can thus be used to train categorization of speech. This use of lexical information differs from the on-line lexical feedback embodied in interactive models of speech perception. In contrast to on-line feedback, lexical feedback for learning is of benefit to spoken word recognition (e.g., in adapting to a newly encountered dialect).
© 2003 Elsevier Science (USA). All rights reserved.

1. Introduction

Learning the phonetic categories of the native language is one of the listener's earliest achievements. At the end of the first year of life, babies show discrimination responses only for phonetic contrasts which occur in the environmental language. Contrasts which the environmental language ignores do not elicit discrimination

---

* Corresponding author. Fax: +44-1223-359062.
E-mail address: dennis.norris@mrc-cbu.cam.ac.uk (D. Norris).

responses. The learning underlying this specialization has taken place within just a few months, since at 6 months of age, babies can discriminate foreign-language contrasts which at 10 months of age they no longer respond to (Werker & Tees, 1999). Phonetic learning early in life has far-reaching consequences also for adult listening: foreign-language contrasts which do not correspond to category distinctions in the native language are notoriously hard to learn (Strange, 1995).

This does not, however, entail that native phonetic categories are immutable once learned. The precise realization of phonetic categories changes over time for the language community as a whole (Labov, 1994) and for individual speakers (Bauer, 1985; Harrington, Palethorpe, & Watson, 2000; Yaeger-Dror, 1994). Such systematic shifts in production must cause, and in turn feed on, changes in the perceptual model presented to listeners. Adult listeners are thus able to adjust the phonetic categories of their language when required.

These adjustments presumably involve reference to lexical meaning. Phonetic categories need to be learned precisely because (by definition) they distinguish minimal pairs of words; similarly, change in phonetic categories most reasonably involves knowledge of the word in which the changing phoneme is encountered. Consider an example involving exposure to a new dialect. In British English, /t/ between two vowels is likely to be released (so that both instances of /t/ in *total* would be essentially the same), while in American English an intervocalic /t/ is likely to be realized as a flap (so that the two instances of /t/ in *total* would be different). Speakers of British English have difficulty interpreting American intervocalic /t/, but British speakers who have moved to America treat it just as American listeners do (Scott & Cutler, 1984); we assume that they have learned to do this by recognizing that their American interlocutors intended to utter words such as *total*, *writer*, *cotton*, or *computing*, rather than new, unknown words containing a new, unfamiliar phoneme.

Listeners are also highly adept in adjusting category judgements as a function of the immediate speech environment. For example, the duration of the formant transitions in the syllables [ba] and [wa] is an important cue to the [b]–[w] distinction, but the boundary between [ba] and [wa] on this dimension varies as a function of speaking rate. Even a small lengthening of the post-consonantal vowel (as a cue to speaking rate) can induce this rate-dependent adjustment (Miller & Liberman, 1979; see also, e.g., Lotto, Kluender, & Green, 1996; Sawusch & Newman, 2000). Vowel height variation in contextual vowels, consistent with differences in vocal tract size, can induce listeners to alter judgements about vowel identity (Ladefoged & Broadbent, 1957). Articulation of a given sound can differ as a function of adjacent phonetic context, and compensation for this is also reflected in changes in listeners' category boundaries (Mann & Repp, 1980, 1981). None of these short-term and short-range adjustments requires learning as a response to the current input, nor do they depend on lexical knowledge. Learned modulation of phonetic category judgements by lexical knowledge alone has not been directly investigated in the speech perception literature. The present study is a first attempt to subject this issue to empirical test.

If lexical information can play a role in modifying phonetic categorization over time this would involve higher-level (lexical) knowledge feeding back to adjust

lower-level (prelexical) processing. This is a different kind of feedback from the on-line feedback which has been the subject of much recent discussion in the speech perception literature (Norris, McQueen, & Cutler, 2000). That debate concerned immediate on-line feedback; do individual stages of language processing function autonomously, or is the operation of lower levels of the system subject to constraint from higher levels? But the feedback which would modify a listener's phonetic categories cannot be of this type; it must have long-term effect, such that phonetic processing will produce the same adjusted output regardless of the current lexical analysis. The listener's phonetic processing will be modified so that all future input, including out-of-vocabulary items such as new words or unfamiliar names, will be processed in the same way. This type of feedback is essentially feedback for learning: high level information informs the perceptual system, resulting in alteration of the long-term development of perceptual representations or processes. In this way, listeners can adjust their phonetic categories to accommodate changes in their community's pronunciation, or to adapt to a newly encountered dialect.

The contrast between perceptual (on-line) feedback and feedback for learning can be illustrated by considering the operation of a simple feedforward back-propagation network (Rumelhart, Hinton, & Williams, 1986). Simple feedforward networks behave like modules. Layers within these networks compute particular functions independently of any subsequent processing further along in the system. Their output is passed in one direction only—to the next stage in the processing system. Effectively, such networks are, in Pylyshyn's (1984) terms, cognitively impenetrable. Their operation cannot be influenced by any processes or representations in later stages of processing. In this they contrast with interactive systems with feedback; in such systems, connections are bidirectional, and the top–down connections allow the operations of each layer to be continually under the influence of decisions taken at stages which are later in the system's processing.

However, there is one important qualification to the strictly unidirectional flow of information in a feedforward network. In a network being trained by back-propagation, error-correcting feedback will be passed back down through the network. Over time, this feedback will alter the weights in the network, and alter the function it computes. This feedback thus produces retuning of the network over time; however, it is still the case that feedback does not cause modulation of the way the network processes a particular input at a given time. Despite the use of feedback for learning, on-line processing remains strictly feedforward. Restricting feedback to be an error-correcting signal thus has the computational advantage of not adding to the complexity or overheads of feedforward processing. The distinction between feedback for learning and feedback that influences the processing of the current input is discussed in more detail by Norris (1990). In this paper we will refer only to the latter kind of feedback as on-line feedback, since feedback for learning exercises no immediate effect on on-line processing. However, note that the distinction between the data paths through which activation flows in a network, and the separate error-correcting feedback pathway, illustrates an important point. Even if there is indeed feedback of a training signal, this does not entail that there is on-line feedback in the system.

The kind of learning embodied in back-propagation is generally termed supervised learning. Supervised learning is under the control of an external trainer that gives the network feedback about its performance. This feedback can either be a detailed indication of how close the activation of each output unit is to the desired output, or can consist of a simple reinforcement signal, indicating whether the output is correct or not; whatever form it takes, some type of feedback is needed in supervised systems. In contrast, unsupervised systems require no feedback. Unsupervised systems are systems which organize themselves on the basis of the statistical properties of the input, irrespective of whether their outputs have the desired consequences for later stages of analysis (Rumelhart & Zipser, 1986). Note that there is often some ambiguity in the distinction between supervised and unsupervised learning. In the present context we will use the term "supervised" to refer to cases where the learning in the prelexical system is under the influence of a training signal originating from the lexicon. We will use "unsupervised" to refer to learning which takes place entirely within the prelexical system itself, without any reference to external sources of information. The viability of unsupervised systems means that learning itself is not necessarily an indication of feedback. In the early stages of speech perception learning might be entirely unsupervised, and isolated from all later processing.

There is a strong case to be made that supervised learning—that is, perceptual learning using feedback from the lexicon—could be of considerable benefit to the speech recognition system. This is in marked contrast to on-line feedback, which cannot benefit recognition. As Norris et al. (2000) point out, the best any word recognition system can do is to identify the lexical representation in long-term memory that best matches the perceptual input. The only way to improve this matching process is to improve the quality of the initial perceptual analysis. So, in order for on-line feedback to be able to improve recognition, early perceptual analysis would need to deliver a result that was, in terms of its own potential, incomplete or suboptimal. That is, early perceptual analysis would normally have to operate below its maximal level of effectiveness, and only operate at full efficiency under the instruction of feedback from the lexicon. If the perceptual analysis system normally does operate at full efficiency, there can be no possible benefit from on-line feedback. Worse, on-line feedback might actually harm recognition by creating hallucinations.

Interactive theories such as the TRACE model of spoken word recognition (McClelland & Elman, 1986), or the earlier Interactive Activation Model of visual word recognition (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), nevertheless claim that prelexical processing is strongly influenced by on-line feedback from the lexicon, and this view continues to be strongly represented in the literature (see, e.g., Samuel, 1997, 2001). But Norris et al. (2000) advanced the case that the early stages of spoken word perception involve no feedback: the perceptual mechanism responsible for prelexical processing is uninfluenced by constraints arising in the lexicon. Thus, prelexical processing is autonomous. Closely related arguments have been presented by Pylyshyn (1999) to support the case that the early stages of visual perception are also modular.

The crucial issue in the on-line feedback debate is how to explain empirical evidence that phoneme identification is influenced by lexical knowledge: first, phonemes

can be identified faster in words than in nonwords (Cutler, Mehler, Norris, & Segui, 1987; Rubin, Turvey, & Van Gelder, 1976); second, listeners are more likely to identify ambiguous phonemes such that their judgements produce words rather than nonwords (Ganong, 1980; McQueen, 1991; Pitt & Samuel, 1993); and third, noise-masking of phonemes shows lexical influence on listeners' decisions (Samuel, 1981, 1987, 1996). These lexical influences can also be observed in stimuli that do not actually form words themselves (Connine, Titone, Deelman, & Blasko, 1997; Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, 1999; Newman, Sawusch, & Luce, 1997). Connine et al., for example, found that the more similar nonwords are to real words, the more rapidly phonemes in the nonwords are detected. These studies demonstrate that even partially activated lexical representations can have an effect on phoneme identification. Interactive models account for such results in terms of on-line feedback from the lexicon to prelexical processing.

Norris et al.'s (2000) Merge model, however, explains these data without invoking on-line feedback. Merge is an elaboration of the Shortlist model (Norris, 1994), and has a completely feedforward architecture. Information from prelexical processing, whether featural, phonemic or syllabic, activates lexical representations. There is no feedback of activation from these lexical representations to the prior prelexical representations. Merge also has an additional set of decision nodes. These nodes integrate information from both prelexical and lexical levels to enable the model to make phoneme identification responses. When lexical nodes are activated, they activate the decision nodes corresponding to their constituent phonemes. That is, presentation of *meat*, for example, would activate the decision nodes corresponding to /m/, /i/, and /t/. The decision nodes corresponding to these lexically activated phonemes can therefore be activated more quickly than decision nodes receiving input from the prelexical nodes alone. The decision nodes are influenced by the lexicon, but prelexical nodes are not.

Note that Norris et al. (2000), in common with Pylyshyn (1999) and other proponents of autonomy (e.g., Forster, 1979; Massaro, 1989), thus acknowledge that the final percept, or the interpretation of a given input, is subject to influence from high-level cognitive processes. But this influence arises, according to Norris et al., when listeners integrate lexical and prelexical information at a post-lexical level of analysis. Similar assumptions are embodied in the Fuzzy Logical Model of Perception (Massaro, 1987, 1989, 1997; Oden & Massaro, 1978).

Norris et al. (2000) presented several arguments against on-line feedback. First, there was an argument from parsimony. The autonomous Merge model has no components that are not required in any other model of speech perception, and it can successfully simulate the data often cited in support of on-line feedback. If on-line feedback would thus be an additional process in a model, and is not required to explain the data, why postulate it? Second, Pitt and McQueen (1998) have shown that lexical effects in phonetic categorization can dissociate from effects on categorization which reflect the perceptual system's compensation for fricative-stop coarticulation (Elman & McClelland, 1988; Mann & Repp, 1981). This dissociation suggests that these two effects have different loci (e.g., the lexical bias at the decision stage in Merge, and the compensation mechanism at the prelexical level). If the lexical bias

were due to on-line feedback, the feedback should have caused a compensation effect. Dissociations between lexical biases and another low-level adjustment process, that of speech rate normalization, have also been observed (Miller & Dexter, 1988; Miller, Green, & Schermer, 1984).

The third and perhaps strongest argument which Norris et al. (2000) made against on-line feedback was that, as described above, feedback could not possibly help in recognizing words. Why include feedback in a model, if it cannot help the system to perform better? Although more than 30 commentaries on the Norris et al. (2000) paper were published, none offered any counter to this fundamental argument that there is simply no benefit to be derived from having on-line feedback from lexical to prelexical processes during word recognition. Perceptual learning via feedback, however, is a different matter altogether: It is not on-line, and it is an example of feedback from lexical to prelexical processing which can indeed be helpful.

Consider again the situation of encountering a talker with an unfamiliar accent. Some of the phonemes produced by the talker may not map directly onto the listener's existing phonemic categories. The example cited above concerned intervocalic /t/, which American English speakers tend to produce as the flap [D], while British English speakers do not (Scott & Cutler, 1984). Lexical information could help a British English listener adapt to an American English accent. If the listener can be sure that [toDəl] really is the word *total*, then this information could be used to direct the perceptual system to categorize [D] as an instance of /t/, and not as an instance of a new phoneme or another existing English phoneme (for example /d/, which is featurally quite similar). A training signal could feed back information from the lexicon to earlier levels of processing, to modify prelexical representations. This kind of phonetic adjustment would immediately generalize to other words. In general, whenever lexical information can tell the listener which phonemic category a particular sound maps onto, the listener can use this knowledge to retune those categories. In this way, lexical feedback could have a positive, and entirely beneficial, effect on speech recognition.

Very similar perceptual learning would also be of great benefit to infants trying to learn phonological categories of their language. Much infant learning of speech may turn out to be unsupervised learning operating entirely autonomously within the phonological system itself. In principle at least, unsupervised learning should be very effective in learning phonemic categories. For example, Kohonen (1988) showed how unsupervised learning systems could extract phonological categories from spoken input; Behnke (1998) constructed such a system to mimic the infant's acquisition of vowel categories. Indeed, there is now extensive evidence that language learning in infants is strongly influenced by the kind of statistical regularities in the input that might be expected to drive unsupervised learning (Maye, Werker, & Gerken, 2002; Saffran, Aslin, & Newport, 1996). However, one problem for unsupervised learning algorithms is determining whether a particular set of inputs should be categorized as one large category, or whether these inputs really ought to be categorized in terms of a number of smaller categories. In other words, what aspects of the variability in the input really determine how to categorize a particular stimulus, and what should be

considered simply as noise? For infants acquiring language, this corresponds to the problem of determining which aspects of the variability in the input are relevant for assigning the input to different phonemic categories. Lexical information could be of enormous benefit to the infant in this task. If the infant can determine that different utterances contain different tokens of the same word, then the infant could infer that the corresponding phonemes in the two tokens should be instances of the same phonemic categories, and that any differences between those phonemes could safely be ignored.

Thus, in contrast to on-line feedback from lexical to prelexical levels during word recognition, feedback for the learning of mappings from speech input to phonemic categories could be helpful. In the experiments reported below, we investigate whether such feedback actually occurs. The experiments use a laboratory analog of the task of adapting to an unusual speaker, and the critical independent variable is the availability or otherwise of lexical information; this enables us to examine whether listeners can indeed use such feedback to modify their phonemic category decisions.

2. Experiment 1

We presented listeners with the speech from a female native speaker of Dutch. The speech was manipulated so that she seemed to produce exemplars of a phonetic category in an idiosyncratic way. The idiosyncratic sound was an ambiguous fricative [?], midway between [f] and [s], which we created by editing the talker's normal productions of [f] and [s]. The sound was presented in a list of words and nonwords. After exposure to the ambiguous fricative in one of three training conditions, listeners were asked to categorize a range of ambiguous fricatives on a five-step [ɛf]–[ɛs] continuum which included, as midpoint, the same frication noise as was used in the training phase. The question was whether categorization of these sounds would shift as a function of training condition.

In the first training condition, listeners heard twenty ambiguous [f]-final Dutch words like [wɪtlo?] (based on *witlof*, chicory; critically, [wɪtlos] is not a Dutch word) and twenty unambiguous [s]-final words (e.g., *naaldbos*, pine forest; [na:ldbof] is not a Dutch word). This was their only exposure to [f]- or [s]-like sounds (there were no occurrences of [f], [s], [v], or [z] in nonfinal position in these words, nor anywhere in the other items they heard). For these listeners, therefore, it was as if the talker's boundary between [f] and [s] had been shifted towards [s], such that [f]-final words ended with a fricative that was more [s]-like than the normal Dutch [f]. The second group of listeners heard exactly the same items, except that the final sound of the [s]-final words was ambiguous (e.g., [na:ldbo?]) and the [f]-final words were unambiguous (e.g., *witlof*). For these listeners, it was as if the boundary between [f] and [s] had been shifted towards [f], such that [s]-final words ended with a sound that was more [f]-like than usual. The third group of listeners provided a control condition: they heard the ambiguous fricative at the end of strings which would be nonwords whether they ended with [f] or [s]. We predicted that if listeners used lexical

knowledge to adjust their interpretation of the ambiguous fricative, the listeners who had heard [?] in [f]-final words would tend to categorize sounds on the continuum as [f] more often than the listeners who had heard [?] in [s]-final words. The control listeners were predicted to show a pattern of categorization intermediate between the two experimental groups. That is, their training should not produce a bias in favor of either [f] or [s].

2.1. Method

Participants. Fifty-eight members of the MPI for Psycholinguistics subject panel took part. Nine participated in a pretest; the other 49 took part in the main experiment (17 in the group who heard ambiguous [s]-final words and 16 in each of the other two groups). They were paid for their participation. They were all native speakers of Dutch, with no known hearing disorders.

Pretest. The ambiguous fricative and the test continuum for the main experiment were selected using a small phonetic categorization pretest. The syllables [ɛf] and [ɛs] were recorded by a female native speaker of Dutch in a sound-treated booth onto Digital Audio Tape, sampling at 48 kHz. These syllables are the names for the letters F and S. This recording was then redigitized (at 16 kHz) and examined with the X waves speech editor. The frication noises for an [f] and an [s] were excised from the recording, cutting in each case at a zero-crossing at the onset of frication energy and at a zero-crossing close to the end of frication, such that both were 272 ms in duration. A 41-step continuum was then made by adding the amplitudes of the two waveforms sample by sample in different proportions (McQueen, 1991; Repp, 1981; Stevenson, 1979). The proportions were equally spaced in 41 steps from 0 to 1.0 and were added pairwise so as to sum to 1.0. The amplitude of each of these steps was then scaled down to 35% of each original step, to increase the overall ambiguity of the continuum. Each step was then spliced onto an [ɛ] context (96 ms); this was the vocalic portion of the utterance from which the [f] endpoint had been taken, ending at the same splice point. Fourteen steps from this [ɛf]–[ɛs] continuum were then selected for use in the pretest. The steps were the two endpoints plus 12 steps which came from the part of the continuum which appeared to be the most ambiguous: steps 1 ([f]), 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, and 41 ([s]).

Ten blocks of stimuli were created, each with a different random ordering of the stimuli. The final stimulus list was constructed by concatenating these ten blocks. Listeners were tested in sound-damped booths. They were presented with a stimulus once every 2.6 s at a comfortable listening level over headphones. They indicated whether the final sound of each token was [f] or [s] by pressing one of two labeled buttons, "F" or "S." Half the listeners made [f] responses with their dominant hand; the other half made [s] responses with their dominant hand. They were asked to respond on every trial, as fast and as accurately as possible. The main sequence of 140 trials was preceded by a short practice block containing one token of each of the 14 stimuli. Half way through the main sequence there was a short pause.

The total proportions of [f] responses to each member of the 14-step continuum are plotted in Fig. 1. As can be seen, listeners found stimuli in the range from about
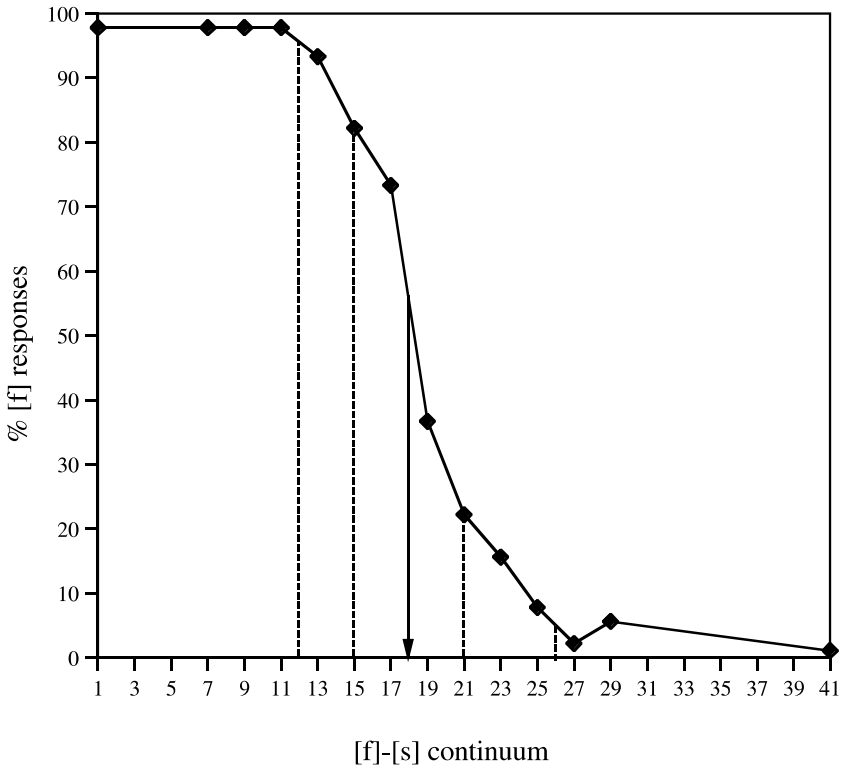
Fig. 1. Experiment 1, pretest: total proportion of [f] responses. The point on the continuum marked by a solid arrow was selected as the ambiguous sound for use in the lexical decision phase. This sound and the four indicated by dashed lines were used in the categorization phase. See text for details.

step 13 to about step 25 to be ambiguous. The continuum was most ambiguous between steps 17 and 19. As shown in Fig. 1, we used these results to select an ambiguous fricative sound for use in the lexical decision phase of the main experiment, as well as five sounds for use in the categorization phase.

Materials and stimulus construction. Step 18 on the continuum was selected as the most ambiguous fricative, and was used as the ambiguous sound [?] in the lexical decision phase. It was also used in the categorization phase, along with steps 12, 15, 21, and 26. These five steps were chosen to span the range from a sound which was likely to be labeled almost always as [f] to one which was likely to be labeled mostly as [s]. They are shown in Fig. 1.

Forty Dutch words were selected. Twenty contained no [s], [v], or [z] sounds, and no [f] sounds except that they ended in [f]. There were five with one syllable (e.g., *druif*, grape), five with two syllables (e.g., *witlof*, chicory), five with three syllables (e.g., *biograaf*, biographer), and five with four syllables (e.g., *locomotief*, locomotive). The other twenty contained no [f], [v], or [z] sounds and no [s] sounds except that they ended in [s]. There were again five of each length. Word frequency was matched between these two sets of words (means: [f]-words, 13 per million; [s]-words, 14 per

million). A further twenty pairs of nonwords were then constructed (five of each of the four lengths). These pairs differed only in whether they ended with [f] or [s]. They contained no [v] or [z] sounds and no [f] or [s] sounds except their final sounds. The [f]-final words, the [s]-final words and the nonwords formed triplets which were matched on their final vowels. Stress patterns were also matched across the sets (primary stress fell on the final syllable of 11/20 items, including the monosyllables, in each set). These items are listed in Appendix A. One hundred filler words were selected and 100 filler nonwords were constructed (in each case, 25 of each length). The sounds [f], [s], [v], and [z] did not occur in any of these items. The nonwords tended to become nonwords (i.e., were no longer consistent with any real Dutch words) before their final phonemes.

The items were recorded by the same talker who recorded the [ɛf]–[ɛs] continuum (during the same recording session). The [f]-final words and the fillers were recorded as such, the [s]-final words were recorded with a final [f] as well as with a final [s] (e.g., *naaldbof* and *naaldbos*) and the experimental nonwords were recorded with a final [f]. The filler items were redigitized at 16 kHz and made into individual speech files. The experimental items were also redigitized at 16 kHz, and then examined with X waves. Ambiguous ([?]-final) versions of the [f]-final words were made by replacing the [f] with [?], which was spliced onto the final vowel of each word (at a zero crossing at the onset of frication). [?]-Final versions of the [s]-final words were made by replacing the [f] in the [f]-final versions of these words with [?]. [?]-Final nonwords were made by replacing the [f] in the original versions with [?]. The ambiguous fricative in all [?]-final items was therefore spliced onto vowels taken from natural [f]-final tokens (as indeed were the fricatives in the [ɛf]–[ɛs] continuum). Any cues to place of articulation in the vowels in the [?]-final items thus consistently cued labiodental place. Individual speech files for the experimental items were then made: natural and [?]-final versions of each of the 40 words, and [?]-final versions of the 20 nonwords.

Procedure. Three lists were constructed, each with 100 words and 100 nonwords. One list contained the [?]-final versions of the 20 [f]-final words, the natural versions of the 20 [s]-final words, 60 filler words (15 of each length), and all the filler nonwords. The second list was identical, except that it contained the [?]-final versions of the [s]-final words (instead of the natural versions) and the natural versions of the [f]-final words (instead of the ambiguous versions). The third list contained the 20 [?]-final experimental nonwords, 80 filler nonwords (20 of each length), and all the filler words.

A pseudo-random running order of the two experimental-word lists was prepared, such that the order of presentation of all items was identical in the two lists. The only difference between the lists was whether, on a given experimental-word trial, the natural or ambiguous version of that word was presented. The natural and [?]-final experimental words were spread evenly through the running order, except that these items did not occur in the first 12 trials. There were never more than four words or four nonwords in a row. A second running order was then made for each these two lists: the same first 12 items in the same order, followed by the remaining items in an order which was exactly the reverse of that in the first running order. Two

versions of the experimental-nonword list were then made, based on the two running orders for the two experimental-word lists. The fillers which were the same as in these other two lists appeared in the same sequential positions. The 20 [?]-final nonwords were divided across the positions taken by the 40 experimental words in the other lists, such that each nonword appeared in one of the two positions where its matched words appeared, and such that they were spread evenly through the running order. Filler words appeared in the remaining twenty "experimental" positions, and filler words appeared in twenty positions which contained filler nonwords in the other two lists, thus maintaining the balance of 100 words and 100 nonwords.

Listeners were tested individually in a sound-damped booth. Written instructions informed them that they would hear a list of words and nonwords over headphones, and that they were to indicate whether each item was a real Dutch word or a nonsense word by pressing one of two labeled buttons, "JA" (yes) or "NEE" (no). They were asked to respond as fast and as accurately as possible. They were not informed about the presence of ambiguous sounds. They were told that there would be a second part to the experiment, but not told what that would be. Each listener then heard the items from one running order of one of the three lists, presented at a comfortable listening level with an inter-onset interval of 2.6 s. Reaction times (RTs) were measured from item onset and adjusted by subtraction of item durations prior to analysis so as to measure from item offset. After this lexical decision phase listeners were given written instructions for phonetic categorization (as in the pretest), while the experimenter placed "F" and "S" labels on the response buttons. Half the listeners made "S" (and "JA") responses with their dominant hand and "F" (and "NEE") responses with their nondominant hand; the other half did the reverse. The five sounds on the [ɛf]–[ɛs] continuum were then each presented six times, in pseudo-random order in the same way as in the pretest, except that there was no practice block, and no pause during the test. Note that, by the normal standards of phonetic categorization tasks, we used a relatively small number of test trials. We avoided a long test session because we were concerned that lengthy exposure to a range of fricative sounds in an unbiased context could lead to unlearning (i.e., an undoing of the learning we predicted would take place during the lexical decision phase in the experimental groups).

2.2. Results

Lexical decision. We first examined performance during the training phase. We excluded from all further analyses all participants in the experimental groups who identified more than 50% of the [?]-final items as nonwords. Three of the 17 listeners who heard ambiguous [s]-final and natural [f]-final words (e.g., [na:ldbo?] and *witlof*) failed to pass this criterion (since they were unwilling to treat the majority of the [?]-final items as tokens of [s]-final words, it is unclear how to interpret their categorization data). All 16 listeners who heard ambiguous [f]-final and natural [s]-final words (e.g., [wɪtlo?] and *naaldbos*) passed the criterion. The [?]-final items in the control condition were labeled by the 16 listeners in this group as nonwords on 97% of trials. One reason why only listeners who heard ambiguous [s]-final words identified

more than 50% of those items as nonwords might be that the ambiguous [s]-final words were all derived from tokens originally recorded with a final [f]. Coarticulatory information in the vowel may therefore have biased listeners to treat the ambiguous fricative in [s]-final words as [f].

Lexical decision performance in the two word training conditions is summarized in Table 1. Analyses of Variance (ANOVAs) on the RT data were carried out with either participants ($F1$) or items ($F2$) as the repeated measure. Means were taken over the available data in each cell of each design. In each case the ANOVAs had two factors: training condition (between participants but within items) and final fricative (whether the original words ended with [f] or [s], which was within participants but between items). Because of the between-participants design, the ambiguity effect took the form of an interaction between training condition and final fricative: $F1(1, 28) = 484.52$, $p < .001$, $MSe = 1218$; $F2(1, 38) = 266.31$, $p < .001$, $MSe = 2795$. On average, listeners identified the natural fricative-final items as words 199 ms faster than the ambiguous items. Neither of the main effects was significant. Note also that the listeners in the control condition were as slow to decide that the [?]-final items were nonwords (266 ms, on average) as the listeners in the word condition were to decide that the [?]-final items were words (262 ms, on average).

Although listeners were thus slower to make decisions about the [?]-final items, those in the word training conditions nevertheless usually identified them as tokens of the intended words. The overall proportion of "no" responses to [?]-final words was only 10%. Most listeners therefore accepted most of the [?]-final versions of words as tokens of those words. The proportion of "no" responses to the original words was 5%. This ambiguity effect was significant: $F1(1, 28) = 10.96$, $p < .005$, $MSe = 0.0048$; $F2(1, 38) = 6.48$, $p < .05$, $MSe = 0.0110$. While there was again no main effect of training condition, there was a main effect of type of fricative: there were more "no" responses to the [s]-final words than to the [f]-final words ($F1(1, 28) = 27.26$, $p < .001$, $MSe = 0.0048$; $F2(1, 38) = 7.40$, $p < .01$, $MSe = 0.0239$). This appears to have been due to two factors: first, three of the natural [s]-final words had mean error rates of 31%; and second, four of the ambiguous [s]-final words (all of them monosyllabic) were judged to be nonwords more than 35% of the time.

We also examined whether performance on the [?]-final items in the lexical decision task changed as a function of extent of exposure to the fricatives. We correlated the rank order of presentation with mean RT for each of these items, in four different

Table 1
Experiment 1: Lexical decision performance (mean reaction times, RTs, for "yes" responses, in ms measured from word offset, and mean percentage "no" responses) in the two word training conditions, for the natural and ambiguous versions of the [f]- and [s]-final words

	Natural fricatives		Ambiguous fricatives	
	[f]-final words	[s]-final words	[f]-final words	[s]-final words
Mean RT "yes" responses	62	63	254	269
Mean % "no" responses	1	8	4	16

subsets of data (two running orders by two sets of words, [f]- and [s]-final). None of these correlations was significant. We also carried out a more sensitive test, where rank order of presentation of each word was correlated with the mean difference in RT between the [?]-final and unambiguous versions of that word (thus removing the variability in RT across items due to their length, frequency etc.; note, however, that for each item the ambiguous-natural RT difference was computed between subjects). There was only very weak evidence that listeners' responses to the ambiguous items became faster over the course of the exposure phase. There were negative correlations between the mean ambiguous-natural RT difference and rank order in all four subsets of the data, but only one of these correlations was significant (for the [f]-final words, in one of the running orders: $r(19) = -0.45$, $p < .05$). These results suggest that there were no strong tendencies to accept the [?]-final items more rapidly as words later in the lists (i.e., after more exposure).

The lexical decision results taken as a whole suggest that the ambiguous fricative tended to be interpreted as [f] by the listeners who heard that sound at the end of words which normally end with [f], and as [s] by the listeners who heard it at the end of words which are normally [s]-final. They also suggest that this tendency was somewhat stronger for the listeners who heard the ambiguous [f]-final words than for those who heard the ambiguous [s]-final words.

Categorization. A clear effect of training condition on phonetic categorization was observed. Fig. 2 shows that listeners who had heard ambiguous [f]-final and natural [s]-final words (e.g., [wɪtlo?] and *naaldbos*) were very strongly biased to label all the sounds on the continuum as [f], while those who had heard ambiguous [s]-final and natural [f]-final words (e.g., [na:ldbo?] and *witlof*) were the least likely to label the sounds as [f]. The performance of the listeners in the control ([?]-final nonword) group was intermediate.

In an ANOVA on the proportion of [f] responses for each of the five fricative sounds there was a significant effect of training condition ($F1(2, 43) = 6.56$, $p < .005$, $MSe = 8.955$; note that it was impossible to carry out by-item analyses of the categorization data, both here and in Experiment 2, because only a single continuum was tested in each experiment). There was also an effect of fricative sound ($F1(4, 172) = 48.06$, $p < .001$, $MSe = 1.163$) and an interaction of these two factors ($F1(8, 172) = 4.73$, $p < .001$, $MSe = 1.163$). This interaction reflects the fact that the training effect varied in magnitude across the continuum. Planned comparisons compared the three training conditions. There were reliably more [f] responses after exposure to ambiguous [f]-final and natural [s]-final words than after exposure to ambiguous [s]-final and natural [f]-final words ($F1(1, 28) = 17.40$, $p < .001$, $MSe = 6.596$). Neither word-training condition, however, was significantly different from the nonword-training condition.

2.3. Discussion

The interpretation of an ambiguous sound in a lexical decision session influenced the interpretation of ambiguous sounds in a subsequent phonetic categorization task: categorization responses shifted to reflect the lexically induced interpretation
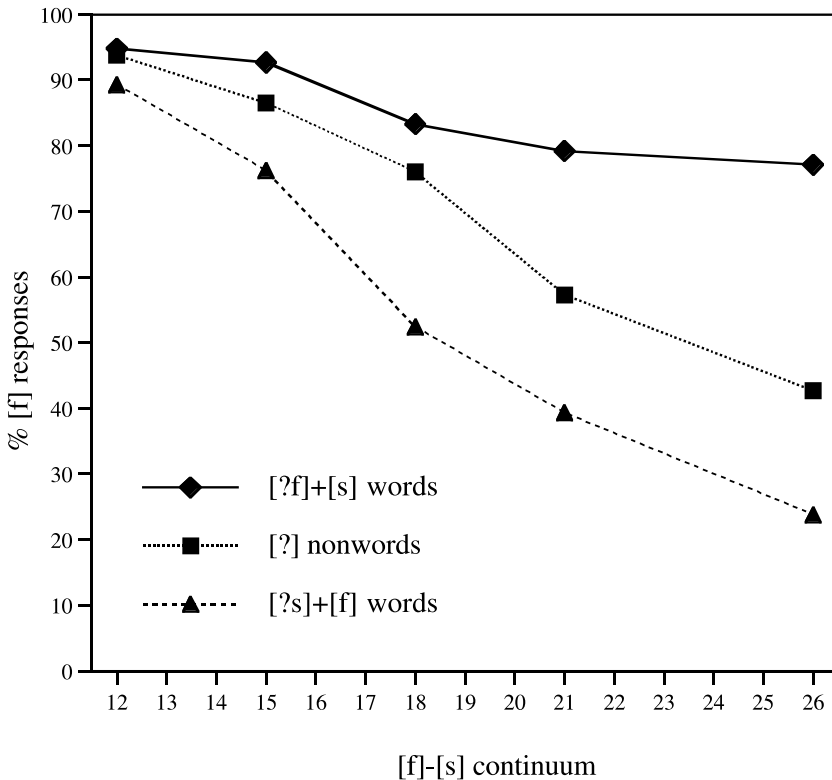
Fig. 2. Experiment 1: total proportion of [f] responses in each of the three training conditions: that with ambiguous [f]-final and natural [s]-final words ([?f] + [s] words); that with ambiguous [s]-final and natural [f]-final words ([?s] + [f] words); and that with [?]-final nonwords ([?] nonwords).

of the sound. These results suggest that listeners can rapidly use their lexical knowledge to adjust the way they evaluate the acoustic-phonetic information in the speech of a particular talker. The control condition shows that mere exposure to the ambiguous fricative (in contexts which did not have any lexical biases) does not cause the shifts in fricative identification seen in the experimental conditions. This suggests that the lexicon was involved in the training effect.

But alternative accounts of these data are possible. The training effect could be a nonlexical selective adaptation effect (Eimas & Corbit, 1973; Samuel, 1986). In selective adaptation, repeated exposure to one endpoint of a phonetic continuum results in an identification bias in favor of the other endpoint of the continuum. For example, because listeners who heard ambiguous versions of [f]-final words also heard natural [s]-final words during the lexical decision phase, this exposure to unambiguous [s] could cause adaptation of [s]. Listeners would then become more likely to interpret ambiguous fricatives as [f]. Likewise, exposure to [f]-final words could cause an increase in [s] responses. On this view, the [?]-final words would play no role in the training effect observed in Experiment 1.

Another possibility is that the training effect could simply be due to the fact that listeners were exposed to both unambiguous and ambiguous fricatives during training. In the experimental conditions in Experiment 1, listeners heard tokens of one unambiguous fricative and tokens of an ambiguous sound. Listeners might initially have been unsure whether to interpret the ambiguous sound as [f] or [s]. Those who also heard an unambiguous [s], however, could infer that the ambiguous sound must therefore have been [f], while those who heard an unambiguous [f] could infer that it should have been [s]. That is, the effect might be due to the contrast between the ambiguous phoneme and the unambiguous endpoint, and not due to the use of lexical information.

In order to exclude explanations in terms of selective adaptation or contrast, Experiment 2 included two new control conditions. The selective adaptation explanation was tested by including two control conditions where the fricative exposure of one group of control listeners was limited to the 20 natural [s]-final words, and in another group was limited to the 20 natural [f]-final words. If selective adaptation to the unambiguous endpoints were the cause of the training effect observed in Experiment 1, a similar effect ought to be found in a comparison between these two groups of listeners.

The contrast explanation was tested by including one control group who heard the natural [s]-final words plus the 20 [?]-final nonwords from Experiment 1, and another group who heard the natural [f]-final words and the same 20 [?]-final nonwords. If contrast in the fricative exposure set, rather than lexical knowledge, were the source of the training effect, these two groups of listeners should show the same training effect as listeners in the two original experimental training groups. Note that although the adaptation controls provide the simplest and most straightforward control for adaptation, the contrast control groups also incorporate a control for adaptation. In these groups, listeners hear exactly the same critical phonemes (both unambiguous and ambiguous) as the experimental groups; any adaptation effects should therefore be equivalent in experimental and contrast-control groups.

In Experiment 2, therefore, there were six training groups: two pairs of control groups, testing the selective adaptation and contrast explanations, respectively, and two experimental groups equivalent to those tested in Experiment 1. One further change was also made. There was an [f]-bias in the main part of Experiment 1, relative to the pretest (compare Figs. 1 and 2). Listeners as a whole were more likely to label the ambiguous fricatives as [f] after lexical decision than in the categorization-only pretest. The same bias was also seen in the lexical decision data, where three listeners were excluded from the [na:ldbo?] plus *witlof* condition because they were unwilling to label the majority of the [?]-final items as words, and where the remaining listeners in this group were unwilling to accept [?]-final tokens as instances of monosyllabic words. (Note that the fricative information carried, proportionally, the most weight in the monosyllabic words, and became available sooner than in the longer words, which, in contrast, could often be identified before offset.)

One possible explanation for this bias is that it is the result of a change in the weight listeners gave to vocalic cues to the place of articulation of the fricatives in the pretest and in the main experiment. Listeners in the pretest heard a range of sounds including unambiguous natural tokens of both [f] and [s] noises, in a vowel

context that always signaled [f] (i.e., the vowel came from a token of [ɛf]). They may thus have learned to ignore the coarticulatory information in the vowel (since it was not informative, and was combined with even the most [s]-like sound). The listeners in the main experiment, however, heard at best only one endpoint during lexical decision (listeners in the control group heard neither endpoint), and, during categorization, only fricatives from the more ambiguous region of the continuum. Furthermore, during the lexical decision phase, they heard the items with ambiguous sounds (which always had vocalic information favoring [f]) embedded among many natural utterances. They may therefore have assigned more weight to this coarticulatory vocalic information than the pretest listeners, both during the lexical decision phase (as we suggested earlier, this may be why three listeners tended not to accept the ambiguous versions of [s]-final words as tokens of those words) and during the test phase. This would result in a shift in the [f]–[s] boundary closer to the [s] endpoint than in the pretest.

We therefore ran a new pretest in which only more ambiguous fricatives were presented. Since there were now no trials in which a vowel that was predictive of an [f] was combined with a fricative noise that was unambiguously an [s], we predicted listeners would be less likely to ignore the information in the vowel. This pretest did indeed result in selection of a new [?] sound which was more [s]-like. Furthermore, the pretest was now equivalent to the categorization phase in the main experiment, thus providing a baseline (no training) condition which could be compared with the six training conditions in the main experiment.

3. Experiment 2

3.1. Method

Participants. One hundred and twelve members of the MPI for Psycholinguistics subject panel who had not participated in Experiment 1 took part. Sixteen participated in the pretest; the other 96 took part in the main experiment, 16 in each of six groups. They were all native speakers of Dutch, with no known hearing disorders. They were paid for their participation.

Pretest. Five equally spaced steps were selected from the 41-step [f]–[s] continuum which was made in Experiment 1: steps 12, 18, 24, 30, and 36 (see Fig. 1). Step 18 was the ambiguous fricative used in the lexical decision phase in Experiment 1. We chose one sound that was more [f]-like than the old ambiguous sound, and three more [s]-like sounds. The results of Experiment 1 suggest that none of these sounds was completely unambiguous. Each step was then spliced onto the same [ɛ] context as was used in Experiment 1. Five randomizations of these 5 syllables were made and concatenated into a single presentation order. Listeners were asked to categorize each fricative sound as either [f] or [s]. The procedure was identical to the categorization phase in the main part of Experiment 1. The total proportions of [f] responses to each of the five steps on the continuum are plotted in Fig. 3. Listeners identified the sounds in a systematic fashion. Even the endpoints were not judged to be
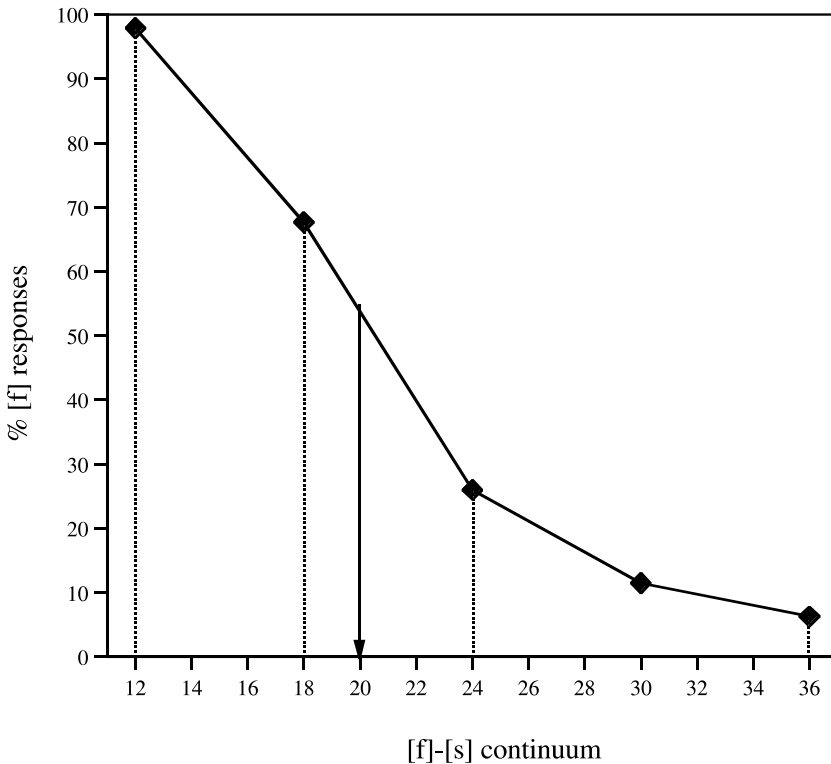
Fig. 3. Experiment 2, pretest: total proportion of [f] responses. The point on the continuum marked by a solid arrow was selected as the ambiguous sound for use in the lexical decision phase. This sound and the four indicated by dashed lines were used in the categorization phase. See text for details.

completely unambiguous. As shown in the figure, we used these results to select a new ambiguous fricative sound for use in lexical decision, one which was more [s]-like than the sound used in Experiment 1.

Materials and stimulus construction. New versions of the test continuum and of each of the 60 [?]-final items were constructed. Otherwise, the materials were identical to Experiment 1. As shown in Fig. 3, linear interpolation suggested that step 20 on the continuum was the closest to the 50% crossover point on the fricative identification function. This sound was used as the new ambiguous sound, and was spliced onto the final vowel of each of the 60 fricative-final items in the same way as in Experiment 1. The test continuum consisted of fricative steps 12, 18, 20, 24, and 36, each spliced onto the [ɛ] sound that was used in the pretest and in Experiment 1. Four of the stimuli in this continuum had been used in the pretest (see Fig. 3); the fifth contained the ambiguous sound which was used in the lexical decision phase.

Procedure. Six lists were constructed. The first two of these lists were identical to the two experimental-word lists used in Experiment 1 (including the same two running orders), except that the versions of the [?]-final words with the new ambiguous fricative were used. There was thus one list with [?]-final versions of the 20 [f]-final

words, natural versions of the 20 [s]-final words, and 160 fillers, and one list with [?]-final versions of the [s]-final words, natural versions of the [f]-final words, and the same fillers. The other four lists consisted of two pairs. The first pair was made by substituting, in the running orders of each of the two experimental lists, each occurrence of a [?]-final item with a filler word. Thus, one list was made by substituting the [?]-final versions of [f]-final words with filler words (making a list in which the only critical fricatives were the natural [s]-final words), while the other was made by substituting the [?]-final versions of [s]-final words with the same filler words (so that the only critical fricatives were the natural [f]-final words). The other pair was made in the same way, except that [?]-final nonwords were used in the substitution instead of filler words. One list therefore contained [?]-final nonwords and natural [s]-final words, the other [?]-final nonwords and natural [f]-final words.

The instructions and the testing procedure were the same as in Experiment 1. The only other change from Experiment 1 was that the new test continuum was used in the categorization phase.

3.2. Results

Lexical decision. We again removed from all further analyses all participants in the experimental groups who identified more than 50% of the [?]-final items as nonwords. Only one listener failed to pass this criterion (in the group which heard ambiguous [s]-final and natural [f]-final words; note that three listeners were excluded from this group in Experiment 1). Lexical decision performance in the two word training conditions is summarized in Table 2. Analyses were performed in the same way as in Experiment 1.

The RT results were very similar to those in the preceding experiment. There was again a significant interaction between training condition and final fricative: $F1(1, 29) = 290.76$, $p < .001$, $MSe = 2145$; $F2(1, 38) = 555.76$, $p < .001$, $MSe = 1397$. Listeners labeled the natural fricative-final items as words faster than the ambiguous items (by 200 ms on average). Neither of the main effects was significant by both subjects and items.

Most of the ambiguous items were however again identified as words. This is true regardless of whether or not we include the data from the one excluded subject. The overall proportion of "no" responses to the original words was 6%, while that for the [?]-final versions was 12%. This difference was again significant: $F1(1, 29) = 14.01$,

Table 2
Experiment 2: Lexical decision performance (mean reaction times, RTs, for "yes" responses, in ms measured from word offset, and mean percentage "no" responses) in the two word training conditions, for the natural and ambiguous versions of the [f]- and [s]-final words

	Natural fricatives		Ambiguous fricatives	
	[f]-final words	[s]-final words	[f]-final words	[s]-final words
Mean RT "yes" responses	96	69	274	292
Mean % "no" responses	4	7	5	18

$p < .001$, $MSe = 0.0043$; $F2(1, 38) = 7.01$, $p < .05$, $MSe = 0.0111$. As in Experiment 1, there was a main effect of type of fricative: there were more "no" responses to the [s]-final words than to the [f]-final words ($F1(1, 29) = 23.72$, $p < .001$, $MSe = 0.0043$; $F2(1, 38) = 4.89$, $p < .05$, $MSe = 0.0269$). In contrast to Experiment 1, however, there was also a main effect of training condition: the listeners who heard natural [f]-final words and ambiguous [s]-final words made more "no" responses than those who heard natural [s]-final words and ambiguous [f]-final words ($F1(1, 29) = 9.91$, $p < .005$, $MSe = 0.0042$; $F2(1, 38) = 4.90$, $p < .05$, $MSe = 0.0111$). These effects appear to be because one of the natural [s]-final words had a mean error rate of 56% (the monosyllable *roes*, intoxication, which also was one of the three natural [s]-final words with high error rates in Experiment 1), and because four of the ambiguous [s]-final words were judged to be nonwords on 40% or more trials (the same four monosyllables as in Experiment 1, i.e., all of them except *muis*, mouse). Listeners again appeared to be unwilling to treat [?]-final versions of [s]-final monosyllabic words as tokens of those words. Nevertheless, the longer [?]-final versions of the [s]-final words and all the [?]-final versions of the [f]-final words were labeled as words by the majority of listeners.

We again correlated the rank order of presentation with mean RT (for each [?]-final item in each set of words in each of the two running orders). As in Experiment 1, none of these correlations was significant. We also again correlated rank order with the mean difference in RT between the [?]-final and unambiguous versions of each word. None of these correlations was significant either. These results thus again suggest that greater exposure to the [?]-final items did not encourage listeners to be faster to accept these items as tokens of real words.

The performance of the listeners in the control groups was then compared with that of the listeners in the experimental groups. Mean RTs and error rates on the natural [f]- and [s]-final words (i.e., the experimental items that were heard in each case by three of the six groups of listeners) are given in Table 3. By-participant ANOVAs were carried out comparing latencies and error rates for the [f]-final words (i.e.,

Table 3
Experiment 2: Lexical decision performance (mean reaction times, RTs, for "yes" responses, in ms measured from word offset, and mean percentage "no" responses) in the six training conditions, for the natural versions of the [f]- and [s]-final words

	Group:	[f]-final words			[s]-final words		
		[?s] + [f] words	[f] words only	[f] words & [?] nonwords	[?f] + [s] words	[s] words only	[s] words & [?] nonwords
Mean RT "yes" responses		96	69	68	69	73	67
Mean % "no" responses		4	3	4	7	6	8

Note. [?s] + [f] Words, ambiguous [s]-final and natural [f]-final words; [f] words only, natural [f]-final words only; [f] words & [?] nonwords, natural [f]-final words and ambiguous nonwords; [?f] + [s] words, ambiguous [f]-final and natural [s]-final words; [s] words only, natural [s]-final words only; and [s] words & [?] nonwords, natural [s]-final words and ambiguous nonwords.

the data from the listeners in the experimental group who heard natural [f]-final words and ambiguous [s]-final words, the listeners in the control group who heard only natural [f]-final words, and the listeners in the control group who heard natural [f]-final words and [?]-final nonwords) and, separately, for the [s]-final words (i.e., the data from the other three groups of listeners). There were no significant differences among the groups (all four $F1$'s < 1).

Categorization. The difference between the two experimental groups observed in Experiment 1 was replicated. As shown in Fig. 4, listeners who heard ambiguous [f]-final and natural [s]-final words during training (e.g., [wɪtloʔ] and *naaldbos*) labeled most of the sounds on the continuum as [f], and those who had heard the reverse during training (i.e., ambiguous [s]-final and natural [f]-final words like [naːldboʔ] and *witlof*) labeled the fewest sounds as [f]. The listeners in the four control groups produced an intermediate number of [f] responses.
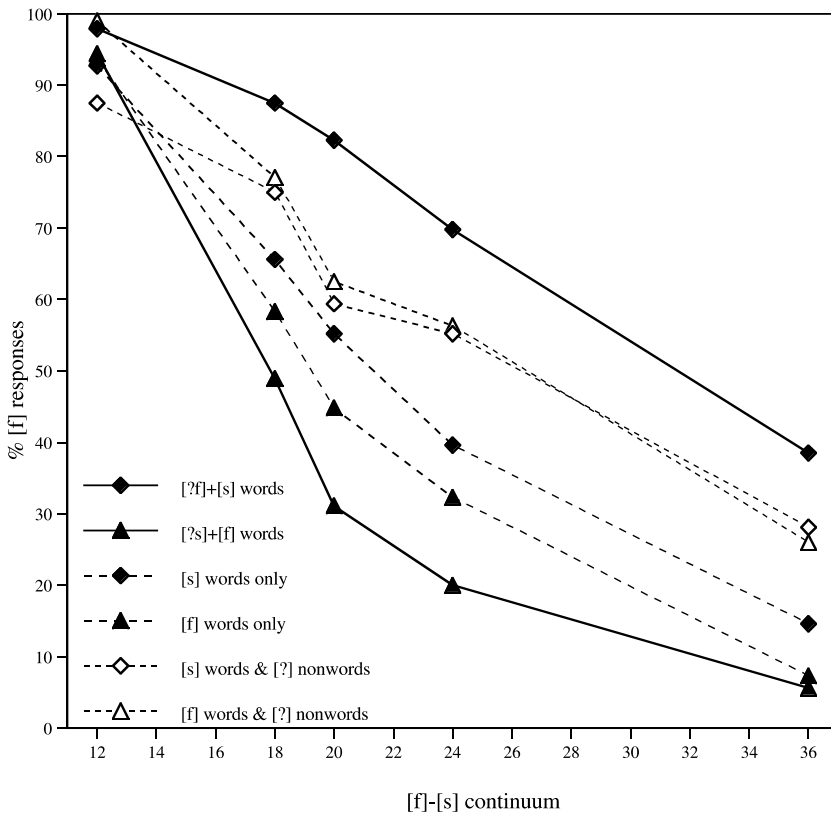

Fig. 4. Experiment 2: total proportion of [f] responses in each of the six training conditions: ambiguous [f]-final and natural [s]-final words ([?f] + [s] words); ambiguous [s]-final and natural [f]-final words ([?s] + [f] words); natural [s]-final words only ([s] words only); natural [f]-final words only ([f] words only); natural [s]-final words and ambiguous nonwords ([s] words & [?] nonwords); and natural [f]-final words and ambiguous nonwords ([f] words & [?] nonwords).

An overall ANOVA on the proportion of [f] responses for each of the five fricative sounds treated the six training groups in a 3×2 design, with one factor being whether the ambiguous fricatives were presented in words, were absent, or were presented in nonwords and the other factor being whether the listeners heard natural [f]- or [s]-final words during training. There was a significant effect of only the latter of these two factors ($F1(1, 89) = 8.39$, $p < .005$, $MSe = 8.310$): there were more [f] responses in the groups which had heard the natural [s]-final words than in the groups which had heard the natural [f]-final words. But these two factors interacted ($F1(2, 89) = 6.84$, $p < .005$, $MSe = 8.310$): fricative labeling varied according to the presence or absence of ambiguous fricatives during lexical decision and/or on the contexts in which [?] was heard. There was also an effect of fricative sound (the five-step continuum; $F1(4, 356) = 185.84$, $p < .001$, $MSe = 1.390$), reflecting the fact that fricative labeling varied systematically across the continuum. The difference between the groups which had heard either the natural [s]- or the natural [f]-final words also varied as a function of the fricative continuum ($F1(4, 356) = 5.14$, $p < .001$, $MSe = 1.390$). No other interactions were significant.

Planned pairwise comparisons were then carried out. Crucially, the listeners who were exposed to ambiguous [f]-final and natural [s]-final words made reliably more [f] responses than those who were exposed to ambiguous [s]-final and natural [f]-final words ($F1(1, 29) = 30.11$, $p < .001$, $MSe = 5.738$). But reliable differences were not found within either of the two pairs of control groups. The listeners who had heard only natural [s]-final words made somewhat more [f] responses than those who had heard only natural [f]-final words (as would be predicted by the selective adaptation account), but this difference was not significant ($F1 < 1$). Contrary to the contrast account, there was no difference in proportion of [f] responses between those who had heard natural [s]-final words and [?]-final nonwords and those who had heard natural [f]-final words and [?]-final nonwords ($F1 < 1$).

The three groups who had heard natural [s]-final words during lexical decision were also compared: those listeners who had also heard ambiguous [f]-final words made more [f] responses than those listeners who had heard only [s]-final words ($F1(1, 30) = 9.23$, $p < .005$, $MSe = 7.326$), and more [f] responses than those listeners who had heard the [s]-final words and the [?]-final nonwords, though this difference was not significant ($F1(1, 30) = 2.83$, $p = .10$, $MSe = 10.196$). Likewise, the three groups who had heard natural [f]-final words during training were compared: the listeners who had heard the [f]-final words and the ambiguous [s]-final words made fewer [f] responses than those who had heard only the [f]-final words, though this difference was not significant ($F1(1, 29) = 1.67$, $p = .21$, $MSe = 4.444$), and fewer [f] responses than those who had heard the [f]-final words and the [?]-final nonwords ($F1(1, 29) = 9.39$, $p < .005$, $MSe = 8.669$).

It was also possible to compare the performance of the two experimental groups with that of the group of listeners who took part in the pretest. Four of the five fricative sounds used in the main experiment were also used in the pretest (i.e., all sounds except the most ambiguous one; see Fig. 3). An ANOVA on proportion of [f] responses to these four stimuli in each of these three groups showed that there was a significant effect of exposure: $F1(2, 44) = 16.46$, $p < .001$, $MSe = 3.655$. There

was again an effect of fricative sound (the four-step continuum; $F1(3, 132) = 178.03$, $p < .001$, $MSe = 1.153$), and an interaction of these two factors ($F1(6, 132) = 6.56$, $p < .001$, $MSe = 1.153$), supporting the observation that the training effect varied in magnitude at different steps along the continuum.

Pairwise comparisons showed that the listeners in the main experiment who had heard the natural [s]-final words and the ambiguous [f]-final words made reliably more [f] responses than the pretest listeners ($F1(1, 30) = 15.16$, $p < .001$, $MSe = 4.363$), but that there was no significant difference between the pretest listeners and the listeners in the main experiment who had heard the natural [f]-final words and the ambiguous [s]-final words ($F1(1, 29) = 2.08$, $p = .16$, $MSe = 2.823$). A final pairwise test also confirmed that even on the four less ambiguous fricative sounds, the listeners who had heard the ambiguous [f]-final and natural [s]-final words made more [f] responses than those who had heard the ambiguous [s]-final and natural [f]-final words ($F1(1, 29) = 28.94$, $p < .001$, $MSe = 3.754$). This finding indicates that the fricative training effect for these listeners was not specific to the fricative sound which was heard during lexical decision (i.e., [?], the most ambiguous stimulus); it generalized to other ambiguous [f]–[s] sounds.

3.3. Discussion

Experiment 2 replicated the main finding of Experiment 1. Listeners who were first exposed to an ambiguous fricative in lexical contexts where an [f] was expected (like [wɪtlo?], where *witlof* is a Dutch word but *witlos* is not) tended to then categorize ambiguous [f]–[s] sounds much more often as [f] than listeners who had first heard the same ambiguous sound in lexical contexts where an [s] was expected (like [na:ldbo?], where *naaldbos* is a Dutch word but *naaldbof* is not). The listeners in the first group appear to have learned to treat ambiguous [f]–[s] sounds as [f], while those in the second group appear to have learned to treat the same sounds more often as [s].

Experiment 2 also ruled out two alternative explanations for the results of Experiment 1. The group whose only exposure to [f]- and [s]-like sounds was to the natural [s]-final words had a slightly stronger tendency to label the fricative continuum as [f] than those whose only exposure to the critical sounds was to the natural [f]-final words. But this difference was not significant. This suggests that while there may be very weak selective adaptation effects due to unambiguous fricatives in this experimental situation, these effects do not explain the main training effect. Furthermore, there was no evidence in the comparison of the two groups who heard [?]-final nonwords and one or other of the two sets of natural fricative-final words that the training effect was due to the contrast between the ambiguous and unambiguous fricatives in the training phase. It appears that exposure to the ambiguous sound plus exposure to tokens of either of the unambiguous sounds during training is not sufficient to induce a shift in the identification of the fricatives at test. Instead, it appears that the ambiguous sound must appear in lexical contexts. The results of Experiment 2 thus confirm that listeners can use their lexical knowledge to adjust to the idiosyncratic fricatives of a talker.

Note that in spite of our attempts to remove the [f] bias that was present in Experiment 1, a residual bias remained in Experiment 2. Listeners in the lexical decision

phase again were unwilling to accept the [?]-final versions of most of the monosyllabic [s]-final words as tokens of the intended words. Furthermore, relative to the pretest listeners, the listeners in the four control groups had an overall bias in favor of [f] in their categorization responses, and the difference between the two experimental groups appeared to be somewhat asymmetric, with a stronger effect manifested by the experimental group whose training also led to an [f] bias at test.

As we suggested earlier, this global residual [f] bias may reflect a shift in cue weighting. Listeners in the pretest heard only one vowel, taken from an [ɛf] context. Formant-transition information in this vowel would have signaled an [f], but, since this would not be informative in the [f]–[s] decision, the pretest listeners may have tended to ignore it. In contrast, the listeners in the lexical decision phase in the main experiment heard a relatively large number of different natural utterances with at most only 10% of the items ending in [?] (i.e., in the four groups who heard any [?] sounds). This may have increased their reliance on vocalic cues, relative to the pretest listeners, making the ambiguous sound more [f]-like. This increased reliance on information in the vowel could then have continued in the (relatively short) categorization phase in the main experiment. Note that, consistent with this suggestion, the [f]-bias at test was more marked for the two control groups who heard [?]-final nonwords during lexical decision than for the two control groups who heard only natural [f]- or [s]-final words. While this bias is orthogonal to the lexically mediated training effect, it does suggest, like the lexical effect, that listeners continually adjust the way in which they interpret speech information in the light of the constraints of the particular listening situation.

4. General discussion

Our experiments have shown that listeners use their lexical knowledge to learn how to interpret ambiguous speech sounds. In Experiment 1, the speech of a Dutch talker was edited such that one group of Dutch listeners heard her produce [s]-final words in her natural way, but heard her say [f]-final words with an ambiguous final sound [?], midway between [f] and [s]. A second group of Dutch listeners heard the reverse (they heard the talker's natural tokens of the [f]-final words, but edited tokens of the [s]-final words, where the final sound was replaced with the same ambiguous fricative [?]). The first group learned to interpret the ambiguous sound as [f], while the second group learned to interpret it as [s]. During the test phase, listeners who had heard [?] in [f]-final words tended to label fricatives on an [?f]–[?s] continuum as [f], while those in the other group tended to label the fricatives as [s].

These effects were replicated in Experiment 2, using the same words and nonwords in the exposure phase, but a slightly different ambiguous sound (one more finely balanced between [f] and [s]) and a slightly different test continuum. The categorization data in this experiment showed, in addition, that perceptual learning was not limited to the ambiguous sound [?], but generalized to other ambiguous fricatives. This suggests that the learning applied to [f]-like and [s]-like sounds in general, rather than only to the specific sound used in the exposure phase, which in turn suggests that exposure altered the boundary between these two phonetic categories.

Control conditions showed that lexical contexts were required to induce perceptual learning. Listeners exposed to the ambiguous fricative at the end of nonword contexts (i.e., contexts in which neither a final [f] nor a final [s] would make a real word of Dutch) produced a pattern of categorization responses that was intermediate between the two groups of lexically biased listeners (Experiment 1). Furthermore, no reliable bias in fricative identification was observed in a comparison between listeners who heard [?]-final nonwords and unambiguous [f]-final words and listeners who heard the same [?]-final nonwords and unambiguous [s]-final words (Experiment 2). There was also no difference on fricative identification between listeners who had heard only unambiguous [f]-final words in the exposure phase and those who had heard only unambiguous [s]-final words (Experiment 2). It therefore appears that perceptual learning depended on exposure to an ambiguous speech sound in lexically biased contexts. Remarkably, learning took place when exposure to those fricatives was limited to 20 words, spread over a list of 100 words and 100 nonwords. This relatively small amount of exposure nonetheless produced a large shift in the category boundary between the two fricative sounds.

Learning, or adaptation, to novel or distorted speech over periods of only a few minutes has also been observed in the case of accented speech (Clarke, 2002), compressed speech (Mehler et al., 1993; Pallier, Sebastian-Galles, Dupoux, Christophe, & Mehler, 1998), and noise-vocoded speech (Hervais-Adelman, Johnsrude, Davis, & Brent, 2002). The study by Hervais-Adelman et al. is of particular interest because it produced evidence of lexical effects. On each trial they presented listeners with both distorted speech and an undistorted version of the same utterance. One group of listeners heard real sentences, while the other heard utterances consisting of nonwords. Only the listeners trained with real words showed signs of learning. This result complements the current study, in that it shows a lexical effect in a situation where the listener's task is to identify words rather than to perform the more metalinguistic task of phonetic categorization. The rapid learning seen in these studies stands in contrast to the very much slower process of learning to perceive the speech either of one's native language, or of a second language (Strange, 1995). Perhaps the critical difference is that all of the cases of rapid learning may reflect modulation of existing perceptual processes, rather than de novo learning of new perceptual categories. One implication of this might be that although, as we argued in the introduction, lexically guided learning could play an important role in acquisition, its main effect will be to make an otherwise very slow process faster. In the case of second language learning, lexical information will again be of assistance, but will not be able to overcome any fundamental limitation on the malleability of the mature perceptual system. The process of learning to perceive speech involves much more than lexical feedback.

In the introduction we suggested that the development and modification of phonological categories might involve both unsupervised and supervised learning. As Maye et al. (2002) have shown, the development of phonemic categories can be influenced by the statistical properties of the input in a way that is consistent with unsupervised learning, and as Behnke (1998) has shown, an unsupervised learning algorithm can successfully acquire phoneme categories from real speech input. Our data show that extant phonemic categories can also be modified by supervised learning, where the

training signal is derived from the lexicon. The critical feature of this form of learning is that information from the lexicon feeds back down to modify prelexical processing.

This demonstration of lexically supervised learning in speech perception suggests that higher-level knowledge can influence how lower-level processing changes over time. Learning has been central in the debate on the modularity of perceptual systems (Fodor, 1983; Pylyshyn, 1999; Raftopoulos, 2001). Pylyshyn (1999) distinguishes three ways in which, with training, people can use higher-level information to improve their performance on perceptual tasks. One possibility is that training might directly modify the internal workings of early perception, in such a way as to violate modularity. A second possibility, favored by Pylyshyn, is that training might enable the observer to attend to the most informative parts of the output of early perceptual processes. Clearly, this would not involve any violation of modularity. Third, Pylyshyn suggests that post-perceptual processes might, with time and repetition, become so automatized that they are indistinguishable from early perceptual processing itself. Pylyshyn refers to this process as the construction of ''compiled transducers.'' In his view, only procedures that can be achieved by table look-up can be automatized in this way. In practice, though, this caveat imposes few limitations, so that a wide range of perceptual procedures could in principle be ''compiled.'' Indeed, Pylyshyn suggests that visual letter and word recognition might be included in this category. Because a compiled transducer will behave like any other perceptual module, the possibility arises that processes may become functionally modular as a result of experience.

Although the process of learning a completely new discrimination might be explained in terms of compiled transducers, the question being addressed in the present experiments is whether lexical information can lead to a readjustment of an already learned discrimination. The compiled transducer option does not seem plausible given the speed of the learning in the present situation. Furthermore, even if a new compiled transducer were constructed (very quickly) to perform the adjusted discrimination, then either the mechanism performing the old discrimination would have to be disabled or the new transducer would have to operate alongside the old.

Might the current results then reflect an attentional process? A great deal of animal work has been directed towards identifying the neurobiological changes that underlie perceptual learning (see Gilbert, Sigman, & Crist, 2001, for review). Evidence from auditory perceptual learning studies with animal subjects shows that improvements in performance can be associated with changes in the brain regions responsible for auditory perception. There is no direct evidence, however, that changes in auditory cortex can be driven by the specific nature of the information being fed back from higher levels of processing. The most interesting feature of these results is that changes appear to be contingent on the animal making some kind of behavioral response to the relevant stimuli (e.g., Recanzone, Schreiner, & Merzenich, 1993). It is possible that the requirement to make a behavioral response may be important because it results in increased attention to those brain regions (Ahissar, Abeles, Ahissar, Haidarliu, & Vaadia, 1998). The neurobiological evidence to date might therefore reflect attentional modulation of unsupervised learning.

Similarly, research with human subjects has demonstrated the importance of attention in visual (Ahissar & Hochstein, 1993) and auditory (Sussman, Winkler,

Huotilainen, Ritter, & Näätänen, 2002) perceptual learning. Work on how adults learn new phonetic categories (i.e., distinctions between speech sounds that occur in some languages but not in the listeners' native language, such as the /r/–/l/ distinction, which is not contrastive in Japanese) has also indicated that attention plays a critical role in perceptual learning (Francis, Baldwin, & Nusbaum, 2000; Francis & Nusbaum, 2002; Logan, Lively, & Pisoni, 1991; Strange, 1995; Tremblay, Kraus, Carrell, & McGee, 1997). While there is therefore considerable behavioral and neurobiological evidence for changes in attentional focus during perceptual training, the current findings are not amenable to an explanation in these terms. Our results do not show an increase in the listeners' ability to make phonetic discriminations. Instead, our results show that there was a change in the way an ambiguous phoneme was categorized, with the direction of change determined by information that was only available from the lexicon. There is no obvious attentional account of these results. In a network with prelexical and lexical nodes, for example, the weights on the connections between these nodes could modulate the ''attention'' that each word pays to each phoneme. Training could lead word nodes to attend differentially to different aspects of the prelexical output, but any such attentional shifts would not generalize either to new words or to the categorization of the critical phonemes in isolation. The data reported here therefore appear most consistent with the first of the possibilities listed by Pylyshyn (1999). That is, lexical information really does have a direct effect on learning at the prelexical level of processing.

How then might the lexicon supervise learning at the prelexical level? In practice, almost any supervised learning algorithm could perform this task. All that is required is for the system to be able to use lexical information to change the way in which ambiguous phonemes are categorized. This could be achieved by moving the position of the category prototype, by adding the ambiguous phoneme as a new exemplar of the appropriate category (e.g., Nosofsky's, 1984, 1986, GCM model), or changing the category boundaries (e.g., Ashby & Lee, 1991). Perhaps the most familiar supervised learning procedure is the back-propagation algorithm of Rumelhart et al. (1986). An error-correcting training signal could be derived from the degree of discrepancy between the phonological representation induced by the stimulus, and the phonological representation expected from the lexical representation. In the specific case tested here, when there is a discrepancy between the fricative information in the speech signal and stored knowledge about a word-final [f] or [s] in that context, the prelexical representation corresponding to the lexically consistent sound would be more weakly activated than would normally be expected. The weights on the connections providing the input to that fricative could then be adjusted so as to reduce this discrepancy. The recognition system would thus adjust its response, over time, to [f]- and [s]-like frication noises, and hence change its representation of the categories [f] and [s].

This back-propagation account is not the only way to implement an influence of lexical knowledge on prelexical learning. An alternative would be to use Hebbian learning (Hebb, 1949), in which particular connections can be strengthened or weakened by experience. Consider a feedforward Hebbian network with levels corresponding to features, phonemes and words. The training phase of experiments like

those presented here would activate representations of the phonemes /f/ and /s/ along with the most strongly activated word. Hebbian learning would strengthen the association between the word representation and the lexically consistent phoneme representation because both would be activated together. That particular pattern of activation over phoneme nodes would therefore come to activate that word more strongly. However, any changes in these associations would be specific to the trained word. There would be no generalized effect on prelexical processing, because the weights between features and phonemes would be unchanged. The problem with a feedforward Hebbian network is that changes in the activation of words will never change the activations of phonemes, so the lexical information would not be able to modify the mapping between input features and phonemic categories. Any given input would continue to activate the same phonemes regardless of any changes in the weights between phonemes and words.

For a Hebbian network to be able to account for the training effect observed in the present experiments, it would have to be possible for lexical information to modify the connections between features and phonemes. That is, prelexical learning would have to be supervised by the lexicon. This could be achieved by adding feedback connections from words to phonemes, just as in TRACE (McClelland & Elman, 1986). If activation fed back from the most strongly activated word to its constituent phonemes there would be coactivation of phonemes and features. By Hebbian learning the relevant feature-phoneme connections could then be modified. In the case of a word containing an ambiguous phoneme, the input would activate two phonemes equally, but lexical activation from the best matching word would feed back to the phoneme nodes for that word's constituent sounds. The lexically consistent phoneme would therefore be more strongly activated than the alternative phoneme, and the connections between the input nodes and the lexically consistent phoneme would accordingly become stronger. As a result, later presentations of the same ambiguous input would be more likely to activate the phoneme that had been indicated by the lexical information. Lexical feedback would thus have adjusted the phoneme category boundaries. McClelland and Elman (1986, pp. 74–75) indeed suggest that Hebbian learning could be used to retune phoneme detectors in TRACE.

Both the account with an error-correcting training signal (e.g., in a back-propagation network) and the account using Hebbian learning are consistent with the present data. Note, however, that the feedback connections involved in the Hebbian learning account are exactly the sort of connections used to explain the influence of the lexicon on phonemic decision-making in an interactive model like TRACE. In assessing the value of feedback in on-line perception, Norris et al. (2000) pointed out that the main effect of on-line feedback is to make decisions at early levels of analysis line up with decisions already made at later levels. Lexical feedback forces the phoneme level to interpret its input according to the decision already made at the lexical level. Once a word has become more strongly activated than others at the lexical level, feedback to the phoneme level will activate the phonemes that ought to be in that word. This feedback is of no benefit to the on-line recognition of the word. Lining up interpretations made at different levels, however, is exactly what is required to allow Hebbian learning to operate over more than one layer of units. This raises the interesting

possibility that, although activation feedback might not be able to make any useful contribution to the on-line recognition process itself, it might exist as a side effect of a mechanism whose primary function is for learning. But the use of activation feedback to drive learning comes with a cost; there is a risk that on-line feedback might sometimes cause hallucinations (Massaro, 1989; Norris et al., 2000).

One might want to argue that the evidence of lexical involvement in phonemic decision-making (Connine et al., 1997; Cutler et al., 1987; Ganong, 1980; Marslen-Wilson & Warren, 1994; McQueen, 1991; McQueen et al., 1999; Newman et al., 1997; Pitt & Samuel, 1993; Rubin et al., 1976; Samuel, 1981, 1987, 1996) supports the Hebbian learning account. As we pointed out in the introduction, however, all of these results can be explained in a feedforward model (Merge, Norris et al., 2000). Furthermore, such lexical effects can be dissociated from lower-level adjustments due to inferred rate of speech (Miller & Dexter, 1988; Miller et al., 1984) or those which compensate for coarticulation (Pitt & McQueen, 1998). Lexical involvement in phonemic decision-making thus appears to reflect a stage of processing (e.g., the decision level in Merge) distinct from the prelexical stage at which short-term adjustments of the perceptual system to the speech signal are made.

We suggest that the adjustments in phonetic category boundaries due to lexical influences on perceptual learning observed in the present study take place at the same processing level (i.e., at a stage of processing where they can be of benefit to subsequent word recognition). Note that the short-term adjustments (i.e., those due to speech rate and coarticulation) require no invocation of learning mechanisms for their explanation, and are best explained in terms of on-line perceptual processing. If, however, all of these adjustments take place at the prelexical level, then, since lexical biases in phonemic decision-making can be dissociated from the effects of prelexical mechanisms, it seems likely that the on-line lexical influence on decision-making is distinct from the lexical influence on perceptual learning. In other words, if both types of lexical effect were due to the same mechanism (i.e., the feedback in the Hebbian account), one would not expect lexical and prelexical effects to dissociate.

We therefore suggest that there may be two types of lexical influence in speech perception: a lexical bias on phonemic decision-making that does not involve any form of feedback, and lexical feedback for perceptual learning. The results on perceptual learning, however, also have potential implications for the interpretation of studies of phonemic decision-making. Samuel (1997) reported data from a series of experiments using the phoneme restoration and selective adaptation paradigms. Listeners heard words in which a given phoneme (/b/ or /d/) had been replaced by noise; these words produced a selective adaptation effect similar to that which would be expected had the original phoneme actually been present. There was a shift in the identification of stimuli on a /bɪ/–/dɪ/ continuum relative to the preadaptation baseline. There was no such effect when the phonemes were replaced by silence. Samuel argued that the noise-replaced phonemes were being perceptually restored, and that these restored phonemes were producing selective adaptation, just as if the actual phonemes had been presented.

Samuel's (1997) experiments involved just five words, each of which contained only /b/ or, in the other condition, only /d/. Consider what happens when a listener hears a set of words that originally contained the phoneme /b/, but where the /b/ has

been replaced by noise. The listener hears noise but lexical information indicates that the noise should be considered to be an instance of the phoneme /b/. These are therefore conditions in which lexically guided perceptual learning can take place. Such learning will alter the listener's perceptual representation of that phoneme. This raises two possibilities. One is that listeners learn to treat the noise as being a possible instance of /b/, and that therefore the noise causes adaptation, in the same way that /b/ would. Note that the adaptation effect caused by restored phonemes was very small. Alternatively, because the /b/ category moves away from its normal position (though presumably does not become more like a /d/), an ambiguous stimulus half way between /d/ and /b/ will be most likely to be classified as a /d/, as it will now be closer to a /d/ than a /b/. Either of these possibilities would explain how a perceptually restored phoneme might produce the adaptation effects reported by Samuel.

In a further study, Samuel (2001) reported similar results with ambiguous word-final fricatives. The critical conditions in his study thus resemble the lexical training conditions in Experiments 1 and 2. Listeners were presented with adapting stimuli containing an ambiguous phoneme half way between /s/ and /ʃ/. In one condition there was a lexical bias toward perceiving this phoneme as /s/ (i.e., the contexts were the words *bronchitis*, *embarrass*, *malpractice*, and *tremendous*), and in the other there was a bias toward perceiving it as /ʃ/ (i.e., the words *abolish*, *demolish*, *diminish*, and *replenish*). Listeners showed a selective adaptation effect in the same direction as would be expected if the lexical information had altered the on-line interpretation of the phoneme. However, once again, the use of repeated presentations of distorted examples of a phoneme leads to the possibility of perceptual learning. In this experiment listeners should learn to classify the ambiguous phoneme as an instance of the lexically determined phoneme.

It might appear that any perceptual learning in this study would lead listeners to interpret the ambiguous phoneme in line with the lexical bias. The ambiguous phoneme should then be classified as an instance of the lexically biased phoneme. But listeners behaved as though the lexically biased phoneme had caused selective adaptation—that is, they were less likely to categorize the ambiguous phoneme as an instance of the lexically biased phoneme. But there are a number of procedural differences between our experiments and those of Samuel (2001). Most significantly, Samuel's experiments used only a small set of words (four), and had repeated adaptation and test phases. These procedural differences could have acted to alter any effect of perceptual learning. For example, the repeated test phases could reset the perceptual learning by providing anchor points at both ends of the test continuum. The small number of repeated test items might prevent generalization of perceptual learning to new contexts. Furthermore, the ambiguous phoneme was never contrasted with an unambiguous phoneme from the same continuum (i.e., there was never an unambiguous /s/ or /ʃ/). This could influence the effect since learning could depend on whether or not there is a contrast between an ambiguous phoneme and an unambiguous one (given that the effect is lexical, c.f. the contrast control in our Experiment 2). Finally, since there was no control for adaptation by an unambiguous phoneme, we do not know how large the adaptation effect is relative to the standard effect. In the Samuel (1997) study, the adaptation effect for restored phonemes was much smaller than the

adaptation effect for intact phonemes. (In fact, we have been unable to replicate the adaptation effect for restored phonemes; McQueen, Cutler, & Norris, 1999.)

Any or all of these differences in procedure could have influenced the nature or extent of perceptual learning in the Samuel (2001) study, and thus have produced a different pattern of results from those in the present study. It is also possible, however, that the same kind of perceptual learning did take place in the two studies, but, because of the selective adaptation procedure in the Samuel (2001) study, the direction of the effect was the opposite to that observed here. In short, two experiments apparently providing evidence for on-line lexical feedback (Samuel, 1997, 2001) use a procedure that admits the possibility of perceptual learning during the course of the experiment. It is thus possible that these experiments show evidence of feedback for learning rather than of on-line feedback.

The central conclusion from the two experiments reported here is that, over time, information available only from a lexical level of analysis can lead listeners to retune their prelexical categories. That is, there must be some feedback of information from the perceptual processes responsible for lexical analysis to earlier processes responsible for prelexical analysis. The lexicon provides a training signal that can be used to alter the way prelexical information is interpreted. However, this feedback is qualitatively different from the on-line lexical feedback that has been the subject of so much debate in the speech perception literature. Feedback of a training signal could take place without any on-line feedback. This contrast between two qualitatively different forms of feedback has implications for the debate over the modularity of speech perception, in which this distinction between on-line feedback and the longer term process of perceptual learning has not always been acknowledged.

Although our results bear primarily on lexical feedback in perceptual learning, they also have implications for the debate over on-line feedback. The existence of training feedback allows an interpretation of some studies of lexical effects in terms of lexically driven perceptual learning. A further implication of our data concerns the motivation for postulating on-line feedback. As Norris et al. (2000) argued, on-line feedback offers no immediate benefit for speech perception. However, lexical feedback that helps perceptual learning can be of great value, both for adult listeners, and for infants acquiring language.

As discussed above, there are two options for implementing this form of lexical feedback. One would be to use a completely separate feedback path for the training signal. This path would carry the training signal, but would not feed back any activation exercising an immediate influence on on-line processing. Alternatively, the feedback path might need to be constructed from the same neural components as the feedforward pathways. In this case, feedback of a training signal might produce the possibility of constantly available on-line activation feedback. That is, on-line feedback might have a purpose after all—not to help on-line recognition, but rather to help learning. If there is on-line feedback, then, it may have arisen as an epiphenomenon of a perceptual learning mechanism. Powerful theoretical arguments exist for the benefits of lexical feedback in the learning of phonetic categories. The findings presented here show that such feedback does indeed occur. The overall pattern of data on feedback from lexical to prelexical processing is, in our view, consistent

with a "rational analysis" (Anderson, 1990) of the task of spoken word recognition. Rational analysis would imply that ideally the speech recognition system should only employ feedback when it is beneficial. That is, there should be feedback for learning, but not for on-line perception.

Acknowledgments

Appendix A

Fricative-final items used in Experiments 1 and 2

[f]-final words	[s]-final words	[f]- & [s]-final nonwords
rif (reef)	nis (niche)	blif/s
druif (grape)	muis (mouse)	truif/s
braaf (honest)	baas (boss)	knaaf/s
proef (test)	roes (intoxication)	gloef/s
lijf (body)	krijs (scream)	knijf/s
witlof (chicory)	naaldbos (pine forest)	oeknof/s
aanhef (beginning)	hakmes (hatchet)	granklef/s
olijf (olive)	radijs (radish)	rachtdijf/s
karaf (carafe)	karkas (carcase)	kwirtaf/s
octaaf (octave)	relaas (account)	porsiaaf/s
achterneef (great nephew)	pimpelmees (bluetit)	dontiwef/s
middenrif (midriff)	hagedis (lizard)	wongadrif/s
onderlijf (lower body)	paradijs (paradise)	kotilmijf/s
ongeloof (disbelief)	grandioos (grandiose)	ramegoof/s
biograaf (biographer)	geitekaas (goat's cheese)	nikuraaf/s
landbouwbedrijf (agriculture)	ingenieus (ingenious)	onminkelnijf/s
indicatief (indicative)	anekdotisch (anecdotal)	corninbatief/s
locomotief (locomotive)	problematisch (problematic)	daponimief/s
choreograaf (choreographer)	champagneglas (champagne glass)	larodiaaf/s
kwalitatief (qualitative)	democratisch (democratic)	dimorantief/s

References

Ahissar, E., Abeles, M., Ahissar, M., Haidarliu, S., & Vaadia, E. (1998). Hebbian-like functional plasticity in the auditory cortex of the behaving monkey. *Neuropharmacology, 37*, 633–655.

Ahissar, E., & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences USA, 95*, 869–875.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Ashby, F. G., & Lee, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General, 120*, 150–172.

Bauer, L. (1985). Tracking phonetic change in the received pronunciation of British English. *Journal of Phonetics, 13*, 61–81.

Behnke, K. (1998). *The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach (MPI Series in Psycholinguistics 5)*. Wageningen: Ponsen & Looijen.

Clarke, C. M. (2002). Perceptual adjustment to foreign-accented English with short-term exposure. *Proceedings of the 7th international conference on spoken language processing* (Vol. 1, pp. 253–256). Denver, Colorado.

Connine, C. M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language, 37*, 463–480.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology, 19*, 141–177.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology, 4*, 99–109.

Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language, 27*, 143–165.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper, & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 27–85). Hillsdale, NJ: Erlbaum.

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & Psychophysics, 62*, 1668–1680.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 349–366.

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

Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron, 31*, 681–697.

Harrington, J., Palethorpe, S., & Watson, C. I. (2000). Does the Queen speak the Queen's English? *Nature, 408*, 927–928.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Hervais-Adelman, A., Johnsrude, I. S., Davis, M. H., & Brent, L. (2002). Adaptation to noise-vocoded speech in normal listeners: Perceptual learning depends on lexical feedback. Poster presented at the *BSA Short Papers Meeting on Experimental Studies of Hearing and Deafness, University of Sheffield*, Sept 16th–17th 2002.

Kohonen, T. (1988). The ''neural'' phonetic typewriter. *Computer, 21*, 11–22.

Labov, W. (1994). *Principles of linguistic change*. Cambridge, MA: Blackwell.

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

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*, 874–886.

Lotto, A. J., Kluender, K. R., & Green, K. P. (1996). Spectral discontinuities and the vowel length effect. *Perception & Psychophysics, 58*, 1005–1014.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]–[s] distinction. *Perception & Psychophysics, 28*, 213–228.

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America, 69*, 548–558.

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review, 101*, 653–675.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.

Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology, 21*, 398–421.

Massaro, D. W. (1997). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA: MIT Press.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101–B111.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1–86.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*, 375–407.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 433–443.

McQueen, J. M., Cutler, A., & Norris, D. (1999). Lexical activation produces impotent phonemic percepts. Presented at the 138th meeting of the Acoustical Society of America, Columbus, November 1999 (Abstract in *Journal of the Acoustical Society of America, 106*, 2296).

McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1363–1389.

McQueen, J. M., Norris, D., & Cutler, A. (2001a). Can lexical knowledge modulate prelexical representations over time? In R. Smits, J. Kingston, T. M. Nearey, & R. Zondervan (Eds.), *Proceedings of the SPRAAC Workshop, Nijmegen, The Netherlands* (pp. 9–14). Nijmegen: MPI for Psycholinguistics.

McQueen, J. M., Norris, D., & Cutler, A. (2001b). Lexical re-tuning of phonological categories. Presented at the 42nd Annual Meeting of the Psychonomic Society, Orlando, November 2001. *Abstracts of the Psychonomic Society, 6*, 37.

Mehler, J., Sebastian, N., Altmann, G., Dupoux, E., Christophe, A., & Pallier, C. (1993). Understanding compressed sentences—the role of rhythm and meaning. *Annals of the New York Academy of Sciences, 682*, 272–282.

Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 369–378.

Miller, J. L., Green, K., & Schermer, T. M. (1984). A distinction between the effects of sentential speaking rate and semantic congruity on word identification. *Perception & Psychophysics, 36*, 329–337.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25*, 457–465.

Newman, R. S., Sawusch, J. R., & Luce, P. A. (1997). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 873–889.

Norris, D. (1990). A dynamic net model of human speech recognition. In G. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 87–104). Cambridge, MA: MIT Press.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189–234.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*, 299–325.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning Memory and Cognition, 10*, 104–114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–57.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review, 85*, 172–191.

Pallier, C., Sebastian-Galles, N., Dupoux, E., Christophe, A., & Mehler, J. (1998). Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory & Cognition, 26*, 844–851.

Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language, 39*, 347–370.

Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 699–725.

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences, 22*, 341–423.

Raftopoulos, A. (2001). Is perception informationally encapsulated? The issue of the theory-ladenness of perception. *Cognitive Science, 25*, 423–451.

Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *Journal of Neuroscience, 13*, 87–103.

Repp, B. H. (1981). Perceptual equivalence of two kinds of ambiguous speech stimuli. *Bulletin of the Psychonomic Society, 18*, 12–14.

Rubin, P., Turvey, M. T., & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in non-words. *Perception & Psychophysics, 19*, 394–398.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318–362). Cambridge MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89*, 60–94.

Rumelhart, D. E., & Zipser, D. (1986). Feature discovery by competitive learning. In D. E. Rumelhart, & J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, (Vol. 1, pp. 151–193). Cambridge MA: MIT Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110*, 474–494.

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

Samuel, A. G. (1987). Lexical uniqueness effects on phonemic restoration. *Journal of Memory and Language, 26*, 36–56.

Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General, 125*, 28–51.

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology, 32*, 97–127.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science, 12*, 348–351.

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

Scott, D. R., & Cutler, A. (1984). Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior, 23*, 450–466.

Stevenson, D. C. (1979). *Categorical perception and selective adaptation phenomena in speech*. Unpublished Ph.D. dissertation, University of Alberta.

Strange, W. (Ed.). (1995). *Speech perception and linguistic experience: Issues in cross-language research.* Baltimore: York Press.

Sussman, E., Winkler, I., Huotilainen, M., Ritter, W., & Näätänen, R. (2002). Top–down effects can modify the initially stimulus-driven auditory organization. *Cognitive Brain Research, 13*, 393–405.

Tremblay, K., Kraus, N., Carrell, T. D., & McGee, T. (1997). Central auditory system plasticity: Generalization to novel stimuli following listening training. *Journal of the Acoustical Society of America, 102*, 3762–3773.

Werker, J. F., & Tees, R. C. (1999). Experiential influences on infant speech processing: Towards a new synthesis. *Annual Review of Psychology, 50*, 509–535.

Yaeger-Dror, M. (1994). Phonetic evidence for sound change in Quebec French. In P. A. Keating (Ed.), *Phonological structure and phonetic form: Papers in laboratory phonology III* (pp. 267–292). Cambridge: Cambridge University Press.