Cognitive Science 30 (2006) 1113–1126

Phonological Abstraction in the Mental Lexicon

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

[a]*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*
[b]*MRC Cognition and Brain Sciences Unit, Cambridge, UK*

Received 12 December 2005; received in revised form 7 April 2006; accepted 18 April 2006

Abstract

A perceptual learning experiment provides evidence that the mental lexicon cannot consist solely of detailed acoustic traces of recognition episodes. In a training lexical decision phase, listeners heard an ambiguous [f–s] fricative sound, replacing either [f] or [s] in words. In a test phase, listeners then made lexical decisions to visual targets following auditory primes. Critical materials were minimal pairs that could be a word with either [f] or [s] (cf. English *knife–nice*), none of which had been heard in training. Listeners interpreted the minimal pair words differently in the second phase according to the training received in the first phase. Therefore, lexically mediated retuning of phoneme perception not only influences categorical decisions about fricatives (Norris, McQueen, & Cutler, 2003), but also benefits recognition of words outside the training set. The observed generalization across words suggests that this retuning occurs prelexically. Therefore, lexical processing involves sublexical phonological abstraction, not only accumulation of acoustic episodes.

Keywords: Speech perception; Perceptual learning; Phonological abstraction; Episodic models; Spoken-word recognition

1. Introduction

Is knowledge stored in long-term memory as multiple episodes or as abstract prototypes? This question concerns, in the domain of language processing, the nature of the mental lexicon. According to episodic theories (Bybee, 2001; Goldinger, 1998; Hawkins, 2003; Johnson, 1997a, 1997b; Klatt, 1979, 1989; Pierrehumbert, 2001, 2002), the lexicon contains multiple detailed traces of spoken words that the listener has previously encountered. These episodic traces include, for example, acoustic details that are specific to the way a given speaker talks.

Listeners can indeed retain detailed perceptual information about individual tokens of spoken words (for a review, see Goldinger, 1998). The critical question, however, is whether these episodic representations constitute the basic substrate of the mental lexicon or should be con-

Correspondence should be addressed to James M. McQueen, Max Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands. E-mail: james.mcqueen@mpi.nl

sidered simply an adjunct to representations that are primarily abstract in nature. We argue here that evidence that listeners can show sensitivity to episodic detail should not be taken as evidence against abstract representations; further, we argue that the lexicon cannot consist solely of episodic traces.

We present data on perceptual learning in speech recognition that can only be explained by postulating abstract lexical representations. These data challenge any extreme episodic model in which there is no abstraction over the information in the speech signal prior to lexical access. According to such models, word recognition entails a comparison of the current input in all its detail with previously stored lexical episodes. In abstractionist accounts, however, word recognition is mediated by abstract prelexical representations. The speech input is mapped onto abstract phonological representations, which may, for instance, be features (Gaskell & Marslen-Wilson, 1997), phonemes (Norris, 1994), features and phonemes (McClelland & Elman, 1986), or syllables (Mehler, 1981). Lexical representations are specified in terms of those sublexical prototypes. In some abstractionist models, such as TRACE (McClelland & Elman, 1986), word representations are nodes in an interactive–activation network, and their abstract phonological content is coded in terms of the connections between those nodes and prelexical representations. In other abstractionist models, such as Shortlist (Norris, 1994), the phonological content of lexical representations is stored in the mental lexicon and is thus separate from the prelexical level. In all models of this class, however, lexical representations are phonologically abstract, and the prelexical level acts to categorize the information in the speech signal in terms of sublexical units with the purpose that that information can make contact with lexical knowledge. The fine acoustic detail in the signal, therefore, cannot be stored in the lexicon.

Extreme abstractionist and extreme episodic models lie at opposite poles of a continuum of possible models of spoken-word recognition.[1] The critical contrast between the two ends of the continuum is whether there is abstraction of the speech signal prior to lexical access. Here we use a perceptual learning paradigm to test for evidence of this abstraction. Norris, McQueen, and Cutler (2003) previously showed that listeners use lexical knowledge to retune speech–sound perception. In a two-part experiment, Dutch listeners made auditory lexical decisions to a list of words and nonwords, and then categorized ambiguous fricatives on an [ɛf] – [ɛs] continuum. During the training phase (lexical decision), participants heard a fricative that was ambiguous between [f] and [s] (henceforth referred to as [?]). One group of listeners heard [?] replacing the [f] in [f]-final words (i.e., contexts that biased interpretation of [?] toward [f] such as [wɪtlɔ?]; *witlof* is a Dutch word, *witlos* is not), plus unambiguous [s]-final words. A second group had the opposite training conditions (ambiguous [s]-final and unambiguous [f]-final words). The listeners with [f]-biased training categorized more sounds on the test continuum as [f] than those with [s]-biased training. They had retuned their fricative categories.

Why should listeners retune speech–sound perception? As Norris et al. (2003) argued, the only reason would be to facilitate word recognition. They suggested that retuning the perception of individual sounds at an abstract prelexical processing level would make future recognition of other words containing those sounds faster. However, Norris et al. (2003) showed only that listeners adjusted the way they categorized sounds in a task requiring a metalinguistic judgment. It could be that the learning is restricted to tasks requiring explicit judgments and that such judgments are based on postlexical phonological representations that do not play any direct role in word recognition. Both abstractionist and episodic models can include a postlexical processing stage at which metalinguistic judgments about phonemic categories are made as in, for example, the abstractionist Merge model (Norris, McQueen, & Cutler, 2000) and the episodic models of Johnson (1997b) and Pierrehumbert (2002). Note, therefore, that we are not claiming that extreme episodic models do not or cannot have representations of phonemic categories. Our argument concerns the function that these categories serve in word recognition. In extreme episodic models, phonemic categories may exist as labels for clusters of (components of) episodic traces but serve no abstraction function during lexical access. Because both abstractionist and extreme episodic models can have postlexical phonemic categories, the learning effect in phonemic categorization reported by Norris et al. (2003), therefore, does not in itself distinguish between the two classes of models. What needs to be tested is whether the retuning influences the perception of words that contain the critical ambiguous fricative but that were not encountered in the training phase. An effect on newly encountered words would suggest that the locus of the adjustment underlying the retuning is prelexical, and that the adjustment reflects learning about abstract sublexical representations.

Therefore, we designed a new experiment that would test for lexical generalization. Specifically, we examined whether the adjustments induced by prior exposure would bias subsequent interpretation of minimal word pairs. These were Dutch words differing only in a final [f] or [s] (cf. English *knife–nice*). The first phase of the experiment replicated the Norris et al. (2003) training conditions. The second phase used cross-modal identity priming, with ambiguous primes based on minimal pairs (e.g., [doː?], from *doof* “deaf” versus *doos* “box”). Listeners heard such primes and then made visual lexical decisions to letter strings, including *doof* and *doos,* presented immediately after the primes.

The priming task allowed us to ascertain how listeners interpret [?]; lexical decisions (e.g., to [doː?]) would not tell us this. In cross-modal identity priming, responses to visual words are facilitated when the same word has just been heard (relative to after an unrelated word), but this facilitation is abolished when prime and target mismatch in one phoneme (e.g., in *fate–fake*; Marslen-Wilson, Nix, & Gaskell, 1995). If fricative learning generalizes to other words, listeners with [f]-biased training should tend to interpret [doː?] as *doof,* and so should make faster or more accurate decisions to *doof* after hearing [doː?] than after an unrelated prime. They should not show such facilitation in responses to *doos.* Listeners with the opposite training should show the opposite priming pattern. The patterns thus act as a diagnostic of how listeners interpret the words and should, in turn, tell us how they interpret the ambiguous phoneme in those words.

Lexical generalization would support abstractionist models in which training causes changes in prelexical fricative categories. If, for example, the listener has learned that the exposure speaker’s [f] is more [s]-like than normal, and if this is coded at the prelexical level, then the listener should tend to hear [doː?] as *doof.* Generalization across words about sublexical components is precisely what an extreme episodic model cannot do. A model storing only acoustically detailed lexical episodes and using only those episodes in word recognition cannot take advantage of any sublexical regularities during the word-recognition process. Storage of traces corresponding to critical training trials (e.g., [wɪtlɔ?]) will not affect the goodness of fit of [doː?] to previous episodes of *doof* and *doos.* Even after training, listeners with this kind of episodic lexicon would therefore have trouble identifying [doː?]. They could compare it with prior traces of *doof* and *doos,* but would have no grounds to prefer either interpretation.

2. Method

2.1. Participants

Forty-eight members of the Max Planck Institute participant pool were paid to participate. None reported any hearing disorders, and none had participated in similar experiments (e.g., those of Norris et al., 2003).

2.2. Materials and stimulus construction

The stimuli in the training phase were the same physical tokens, in the same order, as those used in the two experimental conditions in the training phase of Experiment 2 in Norris et al. (2003). There were 100 Dutch words and 100 phonotactically legal nonwords. Twenty of the words ended in [f] and 20 in [s]; in all 40, substitution of the [f] by [s] or vice versa would make a nonword. The sounds [f, s, v, z] did not occur elsewhere. In one condition, the final sound in the [f]-final words was replaced with an ambiguous sound, [?], which a pretest showed was midway between [f] and [s] (e.g., making [wɪtlɔ?]; *witlof* means "chicory"); the [s]-final words were natural (e.g., *naaldbos,* "pine forest"). In the second condition, the [s]-final words ended with [?], and the [f]-final words were natural (e.g., [naːldbɔ?] and *witlof*). Norris et al. (2003) listed these critical materials and specified how the stimuli were made and pretested.

The test-phase materials were based on 20 minimal pairs of monosyllabic Dutch words (e.g., *doof–doos*). Mean frequencies of occurrence in CELEX (Baayen, Piepenbrock, & Gulikers, 1995) of the [f]- and [s]-final items were 35 and 32 per million, respectively. Both the f- and s-final versions served as visual targets. Spoken forms, ending with the ambiguous fricative that was used in training, served as primes. Each ambiguous prime was paired with a phonologically unrelated monosyllabic prime word (e.g., [doː?] was paired with *krop,* "head" (of lettuce); see the Appendix). There were another 40 word targets and 100 nonword targets, containing no f or s. There were also 120 [?]-final filler primes (60 based on [f]-final words: e.g., *raaf,* "raven" and *motief,* "motif"; and 60 based on [s]-final words: e.g., *kaas,* "cheese" and *moeras,* "swamp"; *raas, kaaf, moties,* and *moeraf* are meaningless) and 60 unambiguous prime words. The unambiguous sounds [f, s, v, z] did not occur in the primes, which varied in length from one to three syllables.

The talker who produced the training phase stimuli (a female native speaker of Dutch) spoke the primes. Recordings were made and edited in the same way as in Norris et al. (2003). All fricative-final words were recorded with a final [f] (i.e., [f]-final words were recorded as such, but [s]-final words were mispronounced with a final [f]). The ambiguous primes were made by splicing the ambiguous fricative onto each prime final vowel. This mirrored the earlier procedure and ensured that any cues to place of articulation in the vowels always signaled labiodental place. A natural token of each unambiguous prime was selected.

2.3. Design and procedure

The two training phase conditions were each paired with four versions of the test phase. These four versions differed only in counterbalancing. Each contained all 40 critical targets (i.e., both members of all 20 minimal pairs). Ten f-final and 10 s-final targets were paired with ambiguous primes, and the other 10 targets of each type were paired with unrelated primes. Each version was also split into two halves, such that only one member of each minimal pair appeared in each half. For any pair, if one target appeared after its ambiguous prime in one half (e.g., *doo?–doof*), the other appeared after its unrelated prime in the other half (e.g., *krop–doos*). Stimuli were thus rotated over four versions, with, in any given half of any version, five trials in each of the four conditions.

The four test versions also differed with respect to the [?]-final filler primes. Listeners who heard [?] in [f]-final words in the training phase continued to hear [?] in [f]-final primes in the test phase, and those who had heard [?] in [s]-final words during training continued to hear [?] in [s]-final primes in the test phase. In each version, 60 filler targets were coupled with ambiguous versions of either [f]- or [s]-final primes, depending on training. Twenty filler word targets and 20 filler nonword targets were paired with unrelated ambiguous primes (e.g., word targets: *raa?–trein* or *kaa?–trein,* depending on training, where *trein* means “train”; nonword targets: *motie?–weuk* or *moera?–weuk*). The other 20 nonword targets in each version (necessarily different across training conditions) were paired with phonologically related ambiguous primes (e.g., *bla?–blap*, after training with [?] in [f]-words; and *kla?–klang*, after training with [?] in [s]-words).

Therefore, there were eight versions of the test phase (4 versions of the test stimuli, each with 2 different sets of filler trials, conditional on training). The remaining filler trials with unambiguous primes (10 identical, 10 related and 20 unrelated word targets, and 20 related nonword targets) were always the same. Therefore, in all versions, there were 80 word targets and 80 nonword targets, and ambiguous primes were just as likely to be followed by a related target (word or nonword) as by an unrelated target (word or nonword). In trials where primes were phonologically related to targets, there were equal numbers of trials where the target was identical to the prime and where it differed only on the final phoneme (both for ambiguous and unambiguous primes).

One pseudorandom running order was constructed for all versions of the test phase, with never more than three words or nonwords in a row and with the critical trials well spaced. Where possible, target order was identical across versions; otherwise, matched targets across conditions (e.g., *doof* and *doos*) appeared in the same position. As in Norris et al. (2003), there were two different running orders in the training phase, each of which was paired with all test phase versions. Therefore, there were 16 stimulus lists in total, each presented to three participants.

Participants were tested in groups of up to three in separate carrels in a quiet room. Auditory primes were presented over closed headphones. Visual targets were presented in lower-case Arial letters on a computer screen at the primes’ acoustic offset. Written instructions for auditory lexical decision were provided for the training phase (see Norris et al., 2003). Training phase procedure was the same as in that study. New written instructions were given after training. Participants were told that they would continue to hear words over the headphones, but that they would now also see letter strings. It was stressed that their task was now to decide, as fast and as accurately as possible, whether the *visual* stimuli were real words. Responses continued to be made via two buttons labeled “JA” (yes) and “NEE” (no); yes responses were made with the dominant hand.

3. Results

One participant failed to follow the test phase instructions, so was excluded from all analyses.

3.1. Auditory lexical decision

Table 1 shows mean reaction times (RTs) and error rates in the training phase. These data closely replicate Norris et al. (2003). The participants tested here were, on average, slower but more accurate (compare Table 2 in Norris et al., 2003). As in the earlier experiment, participants were much faster (218 msec, on average) to say "yes" to unambiguous words than to ambiguous words. Analyses of variance (ANOVAs) with either participants (*F1*) or items (*F2*) as the repeated measure were carried out on the RT data. There were two factors: training condition ([f]- or [s]-biased; between-subject but within items) and final fricative ([f] or [s]; within-subjects but between items). Because of the between-subject design, the strong ambiguity effect appeared as an interaction between these factors: *F1*(1, 44) = 552.08, $p < .001$; *F2*(1, 38) = 573.53, $p < .001$. Neither main effect was significant at the $p < .05$ level by both participants and items.

Similarly, as in the earlier study, listeners judged most (i.e., about 90%) of the ambiguous items to be words. There were more "no" responses to ambiguous items than to unambiguous items, however, as revealed by an interaction of training condition and final fricative: *F1*(1, 44) = 10.99, $p < .005$; *F2*(1, 38) = 5.02, $p < .05$. As before, there was also a main effect of fricative, with more "no" responses to [s]-final items than to [f]-final items: *F1*(1, 44) = 27.13, $p < .001$; *F2*(1, 38) = 7.33, $p < .05$. The effect of training condition on errors was statistically significant only by participants: *F1*(1, 44) = 6.19, $p < .05$; *F2*(1, 38) = 3.69, $p > .05$.

3.2. Cross-modal identity priming

Table 2 shows mean RTs and error rates in the test phase. Fig. 1 plots priming effects (the difference in RT or error rate to targets after ambiguous vs. unrelated primes). ANOVAs with participants or items as repeated measures revealed a three-way interaction of prime type (ambiguous vs. unrelated), target type (f-final vs. s-final words), and training condition ([?] in [f]- vs. [s]-final words in training) both in RTs: *F1*(1, 39) = 15.07, $p < .001$; *F2*(1, 38) = 8.06, $p <$

Table 1
Auditory lexical decision performance for the natural and ambiguous versions of the [f]- and [s]-final words

	Natural Fricatives		Ambiguous Fricatives	
	[f]-final words[a]	[s]-final words[b]	[f]-final words[c]	[s]-final words[d]
Responses				
Mean RT "yes"	119	160	354	360
Mean % "no"	2	5	3	14

Note. Mean correct reaction times (RTs; in milliseconds measured from word offset) and percentage error rates are shown.

[a]For example, *witlof.* [b]For example, *naaldbos.* [c]For example, *witlo?.* [d]For example, naaldbo?.

Table 2
Visual lexical decision performance for the f- and s-final words in each priming condition as a function of auditory lexical decision training conditions

	f-Final Target		s-Final Target	
Training Conditions	Ambiguous Prime[a]	Unrelated Prime[b]	Ambiguous Prime[c]	Unrelated Prime[d]
Mean RT				
[?] in [f]-final words	609	695	689	663
[?] in [s]-final words	623	657	582	628
Mean % error				
[?] in [f]-final words	1	7	11	4
[?] in [s]-final words	8	7	6	6

Note. Mean correct reaction times (RTs; in milliseconds measured from word onset) and percentage error rates are shown.

[a]For example, *doo?–doof.* [b]For example, *krop–doof.* [c]For example, *doo?–doos.* [d]For example, *krop–doos.*

.01; and in errors: *F1*(1, 39) = 9.87, *p* < .005; *F2*(1, 38) = 9.92, *p* < .005. These three-way interactions show that the two groups of listeners responded differentially to f- and s-final words preceded by ambiguous [?]-final primes. These critical interactions are examined in more detail later. The only other effects that were significant by both *F1* and *F2* were: the interaction of prime type and target type, RT: *F1*(1, 39) = 9.85, *p* < .005; *F2*(1, 38) = 8.14, *p* < .01; errors: *F1*(1, 39) = 9.13, *p* < .005; *F2*(1, 38) = 6.43, *p* < .05; the interaction of target type and training condition, RT: *F1*(1, 39) = 9.69, *p* < .005; *F2*(1, 38) = 13.11, *p* < .001; errors: *F1*(1, 39) = 8.44, *p* < .01; *F2*(1, 38) = 5.53, *p* < .05; and, only in RTs, the main effect of prime type: *F1*(1, 39) = 25.96, *p* < .001; *F2*(1, 38) = 7.86, *p* < .01. Note that the differences between the two training groups in the unrelated conditions likely reflect between-subject variability. The factor, First/ Second Half, of the experiment was not involved in any effects that were significant by both *F1* and *F2.* The overall pattern was thus stable over the course of the experiment. The First/Second Half factor was dropped from subsequent analyses.

The pattern of priming effects was therefore modulated as a function of both target type and training conditions. Pairwise comparisons showed that there were priming effects only where predicted. Participants given [?] in [f]-final words in the training phase were faster to respond to f-final words after ambiguous related primes than after unrelated primes: *F1*(1, 23) = 30.31, *p* < .001; *F2*(1, 19) = 9.89, *p* < .01. They were also slower to respond to s-final words after ambiguous than after unrelated primes (significant only by participants): *F1*(1, 23) = 4.84, *p* < .05; *F2*(1, 19) = 2.24, *p* > .1. However, participants given [?] in [s]-final words in the training phase were faster to respond to s-final words after ambiguous related primes than after unrelated primes: *F1*(1, 22) = 10.71, *p* < .005; *F2*(1, 19) = 14.94, *p* < .005. They were also faster to respond to f-final words after ambiguous than after unrelated primes, but not significantly: *F1*(1, 22) = 2.43, *p* > .05; *F2*(1, 19) = 3.25, *p* > .05.

The participants given [?] in [f]-final words in the training phase also showed priming effects in their errors. They were more accurate in responses to f-final targets after ambiguous than after unrelated primes: *F1*(1, 23) = 7.38, *p* < .05; *F2*(1, 19) = 4.93, *p* < .05. The inhibitory

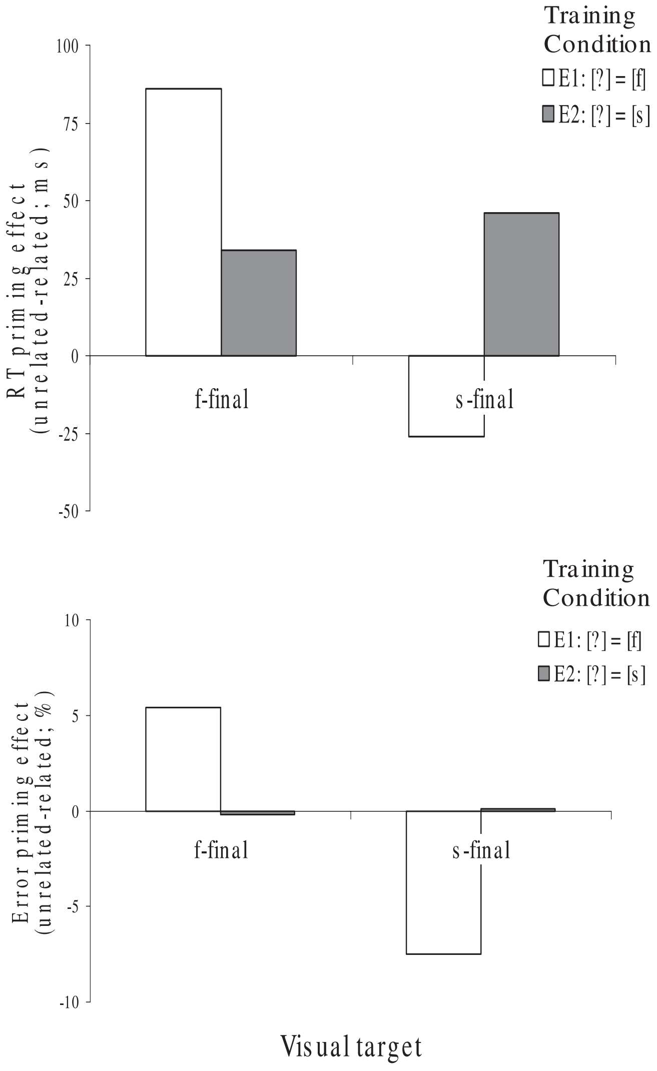

Fig. 1. Test phase: Priming effects (the difference between target responses after ambiguous primes and those after unrelated primes) for reaction times (RTs; upper panel) and error rates (lower panel) to f-final words (e.g., *doof*) and s-final words (e.g., *doos*) for participants who were trained on the ambiguous fricative in [f]-final words (E1: [?] = [f]) and for participants who were trained on the ambiguous fricative in [s]-final words (E2: [?] = [s]).

trend seen in RTs on s-final targets was significant: There were more errors on s-final words after ambiguous than after unrelated primes: *F1*(1, 23) = 15.15, $p < .001$; *F2*(1, 19) = 9.68, $p < .01$. The participants given [?] in [s]-final words showed no significant priming effects in their errors (all *F*s < 1).

The pattern of priming effects was therefore stronger for the listeners given [?] in [f]-final words during training than for those given [?] in [s]-final words. A similar [f]-bias was also observed in the training phase, and in Norris et al. (2003). As Norris et al. (2003) argued, this bias is presumably due to a shift in fricative cue weighting between the pretest and the main experiments. All [s]-final words were made from natural [f]-final utterances, to control for place of articulation cues in the vowel. These cues would have played a stronger role in lexical decision and priming than in the pretest because the vocalic information in the pretest was constant and therefore not informative for the [f]-[s] decision; but it was potentially distinctive in word recognition.

4. Discussion

Perceptual learning about speech sounds generalizes over words. Dutch listeners exposed to an ambiguous fricative in [f]-biased lexical contexts such as [wɪtlɔ?] and to unambiguous [s]-final words such as *naaldbos* learn to interpret that ambiguous sound as [f], whereas listeners with the opposite training conditions (e.g., [naːldbɔ?] and *witlof*) learn to interpret [?] as [s] (Norris et al., 2003). We have shown here that these conditions also influence recognition of fricative-final words that listeners have not heard during training. Listeners who heard ambiguous [f]-final words during training produced reliable facilitatory identity priming in both RTs and errors on f-final words (e.g., [doː?]–*doof*), and inhibitory priming in the error rates on s-final words (e.g., [doː?]–*doos*). However, those who first heard ambiguous [s]-final words produced reliable facilitatory priming on s-final words (in RTs only). Therefore, ambiguous words such as [doː?] tended to be interpreted as *doof* by the former group and as *doos* by the latter group.

These results offer further support for the claim (Norris et al., 2003) that training on ambiguous fricatives in lexically biased contexts leads to adjustments to the prelexical representations of fricatives. If the effect observed in categorization by Norris et al. (2003) were due to adjustments to postlexical phonemic categories (in either an abstractionist model such as Merge: Norris et al., 2000; or an episodic model such as that proposed by Pierrehumbert, 2002), one would not expect those adjustments to influence word recognition. Furthermore, the effects in the priming task are not likely to be due to postperceptual processes. Because the prime–target interval was 0 msec, there was no time for expectancy biases on target decisions to emerge (Neely, 1977), or for the operation of postperceptual strategies (e.g., adjusting interpretation of the ambiguous primes explicitly based on the training conditions). Instead, the learning effect appears to reflect automatic, prelexical adjustments. In line with this automaticity claim, McQueen, Norris, and Cutler (2006) and Eisner and McQueen (2006) showed that the learning effect does not depend on explicit judgments about the fricatives during training.

Only models of spoken-word recognition in which there is a prelexical level of processing that codes abstract phonetic information (i.e., that somehow represents the category distinction between [f] and [s]) can account for these results. Such models are also supported by other recent data showing generalization across the vocabulary of other kinds of phonetic learning (Davis, Hervais-Adelman, Taylor, Carlyon, & Johnsrude, 2005; Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Maye, Aslin, & Tanenhaus, 2003). This abstractionist view receives further support from other types of evidence including results from word and nonword identification (Nearey, 1990, 2001), phonological priming (Radeau, Morais, & Seguí 1995; Slowiaczek, McQueen, Soltano, & Lynch, 2000), and subliminal speech priming (Kouider & Dupoux, 2005); and results from studies on learning novel phoneme sequencing constraints (Onishi, Chambers, & Fisher, 2002) and on second-language listening (e.g., Cutler, Weber, & Otake, 2006; Pallier, Colomé, & Sebastián-Gallés, 2001; Weber & Cutler, 2004). Recent results using the present lexically guided learning paradigm suggest that prelexical representations are both abstract and flexible. In addition to these data, evidence that learning can generalize to similar sounds (Kraljic & Samuel, 2006) and to the full range of sounds used in the test continuum (Norris et al., 2003) suggests that the retuning involves phonologically abstract representations. The flexibility of these representations is shown by the fact that perceptual learning can be talker-specific (Eisner & McQueen, 2005), and that this depends on the degree to which the manipulated phonemes encode talker-specific information (Kraljic & Samuel, 2005, 2006).

These results challenge extreme episodic views of the mental lexicon. In models in which there is no coding of the speech signal in terms of abstract sublexical representations prior to lexical access, there is no way in which exposure to ambiguous sounds in lexically biased contexts could influence recognition of newly encountered words containing those same sounds. In Klatt's (1979) Lexical Access From Spectra (LAFS) model, for example, spectrograms of the current speech signal are mapped directly onto a lexicon of spectral templates. Klatt (1979) also proposed a phonetic analysis device, SCRIBER, which allows new vocabulary to be learned when LAFS fails. Our results on generalization suggest, however, that phonetic analysis is a necessary part of word recognition, such that adjustments to that analysis can influence recognition of words that the present talker has not been heard to speak before but that are part of the vocabulary.

Our results also challenge more recent extreme episodic theories that share the assumption with LAFS that there is no phonological abstraction prior to lexical access (Bybee, 2001; Goldinger, 1998; Hawkins, 2003; Johnson, 1997a, 1997b; Pierrehumbert, 2001, 2002). In the Pierrehumbert (2002) perception model, for example, there is a prelexical stage of processing (the Fast Phonological Preprocessor [FPP]), but the output of the FPP is not phonologically abstract (an image of the output as "a grainy spectrogram" is suggested; p. 123). Goldinger also proposed an extreme episodic model in which lexical entries are memory traces of each token of each word that has been heard, complete with all perceptual details. Because this model has been implemented (it is based on Hintzman's, 1986, MINERVA 2 model), it is possible to use it to test our claim that extreme episodic models cannot account for these data. Simulations show that, if anything, the Goldinger model predicts the reverse of the pattern observed in the human data (Cutler, Eisner, McQueen, & Norris, in press).

It is important to note that Goldinger (1998) did not explicitly rule out the possibility that the input to the model could include abstract phonological representations as well as detailed acoustic traces. However, if abstract representations were included, this would mean that it would no longer be an extreme episodic model. To be consistent with our data, the Goldinger model would have to be transformed into one in which the central burden of word recognition was carried by abstract representations. Perceptual learning would have to take place exactly as proposed here: Learning would occur by retuning the mapping between the input and abstract representations. Our data cannot address whether abstract representations are stored only once with each lexical representation or stored separately with each episodic trace. It is not clear, however, what purpose would be served by including abstract representations in lexical episodes instead of storing only a single abstract representation for each distinct alternative pronunciation of a word. In a model where episodes also consist of abstract representations, every episode of a word would contain exactly the same abstract knowledge.

The central problem with storing words as episodes is that the mapping between input and word is performed primarily by the episodes corresponding to the specific word. Although the form of the representation retrieved by an episodic model is influenced by the whole ensemble of traces in the lexicon, there is no way of changing the mapping of specific phonemes within a word without altering all of the episodes involved in recognizing that word. That is, the only way to capture these data would be to alter the composition of a significant proportion of the episodes (not just the words, but the actual episodes of those words) in the lexicon. As the simulations in Cutler et al. (in press) showed, simply adding new episodes is not sufficient to produce the correct retuning of the system. Without additional mechanisms, no amount of exposure to the ambiguous fricative in contexts such as [wɪtlɔ?]) will affect the relative goodness of fit of [doː?] to previously stored episodes of *doof* and *doos.* In models where the mapping between input and the lexicon is achieved via abstract representations, however, all that is required is to alter the mapping between the input and the single perceptual category that needs to be retuned.

Word recognition, therefore, cannot be based solely on comparison of lexical episodes. However, as we noted at the outset, there is considerable evidence in support of episodic theories. Talker identity influences participants' performance across a range of experimental tasks (Goldinger, 1998), and there are many frequency and gradedness effects in speech production (e.g., *t/d* deletion in English occurs more often in high- than in low-frequency words; Bybee, 2000; Pierrehumbert, 2002). These data are a challenge to abstractionist models, but only extreme ones in which all acoustic detail is filtered out during speech recognition and forgotten. An abstract lexicon for language comprehension and production could be combined with separate storage of talker detail; similarly, frequency effects could arise in an abstractionist model if, for example, multiple pronunciation variants were stored, if processing were probabilistic, or both. However, the data on episodic effects do suggest that extreme abstractionist models are incorrect. The present data show that extreme episodic models are also incorrect. No matter how much of word recognition proves to be episodic, these data show that there must be abstraction in the lexical access process. Hybrid abstractionist–episodic models, therefore, hold considerable promise. It will be essential in developing such models to specify which components are episodic and which are abstract.

There are several benefits of abstraction using flexible prelexical representations. (These arguments are presented in detail in Scharenborg, Norris, ten Bosch, & McQueen, 2005.) What we have shown here is that abstraction makes listening more efficient when one encounters a talker speaking in an unusual way. Listeners used their experience in the training phase to benefit subsequent word recognition. Prelexical adjustments thus removed what would otherwise have been a lexical ambiguity (e.g., *doof* vs. *doos*) in the test talker's speech. This kind of learning, therefore, helps communication in later encounters with the same talker. These arguments, and our results, suggest that an adequate model of spoken-word recognition must include flexible and abstract prelexical representations and, hence, also abstract lexical representations.

Notes

1. Less extreme possible models along this continuum include a model in which multiple acoustic traces are stored prelexically, but the content of lexical entries is phonologically abstract. If lexical representations are abstract, however, prelexical representations have to be abstract too, with the same vocabulary of representation at both levels; otherwise, there could be no contact between prelexical and lexical processing. One version of this model, therefore, is one in which, at the prelexical level, phonemes are recognized episodically. With respect to the data to be presented here, this class of model is indistinguishable from purely abstractionist models in which multiple traces are stored neither lexically nor prelexically. Another intermediate possible model is one in which abstract prototypes are stored at the prelexical level, but in which multiple traces are stored in the lexicon. The content of lexical representations would have to be phonologically abstract and would thus lack specification of full acoustic detail, but there would be multiple exemplars of words (or pronunciation variants) rather than single lexical prototypes. This class of model, because it requires prelexical abstraction, is thus also not an extreme episodic model, as defined earlier. In fact, it is not clear that a model like this should properly be termed episodic. All of the stored episodes for each pronunciation variant would contain copies of exactly the same abstract representation. Without any distinction between episodes, this model would just be a very inefficient implementation of a purely abstract model.

Acknowledgments

This work was first presented at the 9th conference of the Dutch Psychonomics Society (McQueen, Cutler, & Norris, 2003).

We thank Loulou Edelman, Marloes van der Goot, Elske Schoenmakers, Keren Shatzman, and Marloes Weijers for experimental assistance.

References

Baayen, H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* [CD-ROM]. Philadelphia: Linguistic Data Consortium.

Bybee, J. L. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage based models of language* (pp. 65–85). Stanford, CA: CSLI Publications.

Bybee, J. L. (2001). *Phonology and language use.* Cambridge, UK: Cambridge University Press.

Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (in press). Coping with speaker-related variation via abstract phonemic categories. *Papers in Laboratory Phonology, 10.*

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34,* 269–284.

Davis, M. H., Hervais-Adelman, A., Taylor, K., Carlyon, R. P., & Johnsrude, I. S. (2005, June). *Transfer of perceptual learning of vocoded speech: Evidence for abstract pre-lexical representations.* Poster presented at the ISCA workshop on Plasticity in Speech Perception, London.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General, 134,* 222–241.

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

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

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12,* 613–656.

Goldinger, S. D. (1998). Echoes of echoes?: An episodic theory of lexical access. *Psychological Review, 105,* 251–279.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics, 31,* 373–405.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Johnson, K. (1997a). The auditory/perceptual basis for speech segmentation. *Ohio State University Working Papers in Linguistics, 50,* 101–113.

Johnson, K. (1997b). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego, CA: Academic.

Klatt, D. H. (1979). Speech perception: A model of acoustic–phonetic analysis and lexical access. *Journal of Phonetics, 7,* 279–312.

Klatt, D. H. (1989). Review of selected models of speech perception. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169–226). Cambridge, MA: MIT Press.

Kouider, S., & Dupoux, E. (2005). Subliminal speech priming. *Psychological Science, 16,* 617–625.

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

Kraljic, T., & Samuel, A. G. (in press). How general is perceptual learning for speech. *Psychonomic Bulletin & Review, 13,* 262–268.

Marslen-Wilson, W., Nix, A., & Gaskell, G. (1995). Phonological variation in lexical access: Abstractness, inference and English place assimilation. *Language and Cognitive Processes, 10,* 285–308.

Maye, J., Aslin, R., & Tanenhaus, M. (2003, March). *In search of the Weckud Wetch: Online adaptation to speaker accent.* CUNY Conference on Sentence Processing, Cambridge, MA.

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

McQueen, J.M., Cutler, A., & Norris, D. (2003, December). *Perceptual learning in speech generalises over words.* Paper presented at the 9th Wintercongres of the Nederlandse Vereniging voor Psychonomie, Egmond aan Zee, The Netherlands.

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

Mehler, J. (1981). The role of syllables in speech processing: Infant and adult data. *Philosophical Transactions of the Royal Society, Series B, 295,* 333–352.

Nearey, T. M. (1990). The segment as a unit of speech perception. *Journal of Phonetics, 18,* 347–373.

Nearey, T. M. (2001). Phoneme-like units and speech perception. *Language and Cognitive Processes, 16,* 673–681.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 226–254.

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.

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

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition, 83,* B13–B23.

Pallier, C., Colomé A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science, 12,* 445–449.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–157). Amsterdam: Benjamins.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 101–139). Berlin: de Gruyter.

Radeau, M., Morais, J., & Seguí J. (1995). Phonological priming between monosyllabic spoken words. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1297–1311.

Scharenborg, O., Norris, D., ten Bosch, L., & McQueen, J. M. (2005). How should a speech recognizer work? *Cognitive Science, 29,* 867–918.

Slowiaczek, L. M., McQueen, J. M., Soltano, E. G., & Lynch, M. (2000). Phonological representations in prelexical speech processing: Evidence from form-based priming. *Journal of Memory and Language, 43,* 530–560.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50,* 1–25.

Appendix

Ambiguous/unrelated primes: doo?/krop, brie?/poen, hoe?/wrang, le?/ruim, kui?/bron, ka?/wijd, lo?/taak, mu?/drang, gaa?/beuk, gro?/kraan, roo?/pak, lie?/pret, klui?/bang, poe?/wijn, be?/krijt, gra?/boot, bo?/map, wij?/bruin, hal?/droom, loo?/gat.

Each pair of primes was paired with two targets: the f- and s-final versions of the ambiguous primes (e.g., *doo?/krop* with *doof* and *doos*).