

# Language and Speech

<http://las.sagepub.com>

---

## **The Dynamic Nature of Speech Perception**

James M. McQueen, Dennis Norris and Anne Cutler

*Language and Speech* 2006; 49; 101

DOI: 10.1177/00238309060490010601

The online version of this article can be found at:

<http://las.sagepub.com/cgi/content/abstract/49/1/101>

---

Published by:



<http://www.sagepublications.com>

**Additional services and information for *Language and Speech* can be found at:**

**Email Alerts:** <http://las.sagepub.com/cgi/alerts>

**Subscriptions:** <http://las.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.co.uk/journalsPermissions.nav>

**Citations** <http://las.sagepub.com/cgi/content/refs/49/1/101>

# *The Dynamic Nature of Speech Perception*

**James M. McQueen<sup>1</sup>, Dennis Norris<sup>2</sup>,  
Anne Cutler<sup>1</sup>**

<sup>1</sup> *Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

<sup>2</sup> *MRC Cognition and Brain Sciences Unit, Cambridge, U.K.*

## **Key words**

*perceptual  
learning*

*phonetic  
categorization*

*speech  
perception*

## **Abstract**

The speech perception system must be flexible in responding to the variability in speech sounds caused by differences among speakers and by language change over the lifespan of the listener. Indeed, listeners use lexical knowledge to retune perception of novel speech (Norris, McQueen, & Cutler, 2003). In that study, Dutch listeners made lexical decisions to spoken stimuli, including words with an ambiguous fricative (between [f] and [s]), in either [f]- or [s]-biased lexical contexts. In a subsequent categorization test, the former group of listeners identified more sounds on an [ɛf]–[ɛs] continuum as [f] than the latter group. In the present experiment, listeners received the same exposure and test stimuli, but did not make lexical decisions to the exposure items. Instead, they counted them. Categorization results were indistinguishable from those obtained earlier. These adjustments in fricative perception therefore do not depend on explicit judgments during exposure. This learning effect thus reflects automatic retuning of the interpretation of acoustic-phonetic information.

## **1 Introduction**

Across the course of a lifetime, each individual's speech alters. In part this is due to changes in the voice as a function of age—thus older speakers' utterances tend to have a lower fundamental frequency (Russell, Penny, & Pemberton, 1995) and lower intensity (Hodge, Colton, & Kelley, 2001) than the utterances produced by their younger selves. But it is also due to alterations in the speech sounds which a speaker produces. Languages change across the years, as can readily be acknowledged by anyone listening to radio recordings made several decades ago. These changes do not simply result from natural attrition of older speakers and their replacement by younger speakers; as languages change, all their speakers change with them.

*Acknowledgments:* We thank Michiel Seevinck for testing the participants in this experiment.

*Address for correspondence.* James M. McQueen, Max Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands.

Tel: (+31) 24 3521373; Fax: (+31) 24 3521213; e-mail: <james.mcqueen@mpi.nl>.

This aspect of speech production has been amply attested in phonetic studies, one of the most striking examples in this body of research being Harrington, Palethorpe and Watson's (2000a, 2000b, 2005) study of the Queen's English. The Queen of England's annual Christmas broadcasts to her people over a period of some 40 years, beautifully recorded in high fidelity by the BBC, were examined by these authors and the formant structure of the monophthongal vowels (such as [æ] in *happy*, or [ɪ] in *Christmas*) and diphthongal vowels (such as [aɪ] in *my*) determined by acoustic analysis. They observed considerable change in the Queen's vowels from the 1950s to the 1980s. Moreover, the change was far from random. The direction of the change was the same direction as exemplified in the vowels of Standard Southern British English during the latter half of the 20th century, as established in other phonetic studies. The net effect was that the degree to which the Queen maintained a distance from the average speaker (her speech being more conservative, more "upper class" than that of the average speaker) was approximately the same in the earlier and in the later period. In other words, although the Queen could hardly be claimed to be a typical speaker of Standard Southern British English, the alterations in that dialect were also reflected in her speech; her speech changed along with that of her language community.

This process of change within the individual speaker — also separately attested for other British English speakers by Bauer (1985) and for Québec French speakers by Yaeger-Dror (1994) — is one of the reasons why languages change fairly rapidly, resulting in such effects as the noticeable archaism of older radio recordings. No psycholinguistic research has previously addressed the mechanism by which change in an individual's speech sound production might be brought about; it is clear, however, that an alteration in production which follows the trend of a general change in production across a language community must be underpinned by a change in the individual's perceptual model. That is, the speaker alters as a result of input from interlocutors within the community.

Alteration of this kind could only be helpful to language users — because it would facilitate communication. If most other speakers within a community are using more open front vowels, then the Queen will be understood more rapidly if she too uses more open front vowels. Thus there are substantial gains to be made if language users can flexibly adjust their perceptual models for speech sound categories, and consequently adjust their production categories to suit. Adjustment of the perceptual categories alone can even pay off in the short term, when any listener is confronted with a speaker of a different dialect, or a speaker with a foreign accent, or a speaker with unusual voice characteristics. Indeed, our experience tells us that adjustment to a strange speaker who is initially difficult to understand can be very rapid — within minutes, we can begin to get used to an accent and hence to understand more of what the speaker is communicating.

The perceptual learning mechanism involved in this process was investigated by Norris, McQueen, and Cutler (2003). In their study, listeners heard 20 instances of an unusual sound replacing one of the fricative sounds of their language. The perceptual boundary of the speech sound category in question was then tested, to determine whether this small amount of exposure had brought about alteration in the category definition. Norris et al. hypothesized that perceptual learning of this kind

would be driven by communicative efficacy—that is, the learning would only occur if it resulted in a payoff in facilitation of, for instance, word recognition. The way in which they tested this hypothesis was via manipulation of the exposure phase. Some listeners heard the unusual sound replacing a sound in real words. For instance, the unusual fricative replaced [f] in 20 words which would normally end in [f], such as *witlof* (Norris et al.'s experiment was in Dutch; *witlof* is the Dutch word for 'chicory'); all sounds other than [f] remained normal in the input which was presented. Norris et al. predicted that these listeners would rapidly adjust their [f] category to embrace the new sound, because it would help them understand all words containing [f] in the speaker's production. Effectively, the lexicon would guide interpretation of the unusual sound and communicative efficiency would be improved by incorporating this learned interpretation into phonetic processing. Other listeners heard the same unusual fricative replacing [f] in nonwords (such as *oeknof*, which is not a word of Dutch). Although the nonword could itself provide no information about which sound the unusual fricative should have been, the rest of the context did provide that information in that, again, all sounds other than [f] remained normal. Nevertheless, Norris et al. predicted that these listeners would not adjust their [f] category, because doing so would have yielded no benefit for lexical processing during the exposure phase. Without lexically-guided learning, phonetic processing would not rapidly adapt.

Exactly the predicted results were indeed observed. The unusual sound which Norris et al. (2003) had constructed was in fact a sound which had been established in a pretest as maximally ambiguous between Dutch [f] and [s]. The post-test categorization task involved forced-choice decision between [s] and [f]. Listeners who were exposed to the ambiguous sound replacing [f] in real words such as *witlof* (and unambiguous [s]-final words such as *naaldbos*, 'pine forest') shifted the boundary of their [f] category such that the category was more extensive than it would normally be. Other listeners who heard the ambiguous sound replacing [s] in real words such as *naaldbos* (plus unambiguous [f]-final words) shifted the boundary of their [s] category such that that category was more extensive at the expense of the [f] category. For both of these listener groups, the shift would have helped them understand other words containing the designated sound. The listeners who had heard the sound in nonwords, however, irrespective of the type or extent of the contrast offered by the remaining context, failed to show any significant shift in their fricative category boundaries in the post-test phase. Contrast alone was therefore not sufficient to induce phonetic retuning. The retuning resulted only when lexical contexts for the ambiguous sounds in the two real-word conditions provided listeners with sufficient information that adjustment would provide a useful communicative payoff in understanding the exposure speaker's speech.

An open issue remaining with respect to this finding, however, is the degree to which the perceptual learning effect is automatic. Does the retuning which takes place during the exposure phase simply follow automatically as a consequence of experience with the critical fricative-final words, or does it depend on listeners making explicit judgments about the ambiguous fricative sound? In the Norris et al. (2003) study, participants made lexical decisions on all of the stimuli in the exposure phase, including the items ending with the ambiguous fricative. They were thus forced to decide whether the ambiguous sound was an acceptable token of either [f] or [s]

(depending on exposure condition). For example, when listeners in one group heard [witlɔ?], they had to decide whether the final sound was an acceptable [f] in order to be able to respond that this item was a word (i.e., a token of the word *witlof*), or indeed in order to be able to reject the item as a nonword. It is possible that the perceptual learning effect depended on these explicit judgments. Indeed, work on perceptual learning in vision (e.g., Ahissar & Hochstein, 1993) and audition (e.g., Sussman, Winkler, Huotilainen, Ritter, & Näätänen, 2002) suggests that learning only takes place when participants attend to the relevant stimulus attributes (see Pylyshyn, 1999, for review). Furthermore, it appears that attention plays a key role when adult listeners learn new phonetic categories (e.g., when Japanese speakers learn the /r/-/l/ distinction; Francis, Baldwin, and Nusbaum, 2002; Francis & Nusbaum, 2002; Logan, Lively, & Pisoni, 1991; Strange, 1995; Tremblay, Kraus, Carrell, & McGee, 1997).

The goal of the present experiment was to test whether the perceptual learning effect observed by Norris et al. (2003) depends on explicit lexical decision responses on the ambiguous items, or whether the effect is automatic. To this end, we replicated the word conditions of the original experiment, except that we used an exposure task which did not require any explicit judgments. Listeners were asked to count the number of stimuli that they heard during the exposure phase. The stimuli (and their running orders) were identical to those in the two experimental conditions in Experiment 2 in Norris et al. (2003). The only change was in the instructions for the first phase of the experiment. The presentation order was divided into eight blocks. Listeners were asked to count the number of trials in each block, and write down the totals that they had counted in each block in short pauses between blocks. They were also asked to press a response button on every trial, as they counted. These button-press responses provided a measure of the processing demands of both the unambiguous and ambiguous fricative-final words. If the effect observed by Norris et al. depends on explicit decisions about the ambiguous fricatives, as required in the lexical decision task, the effect should be removed or at least attenuated when the exposure task is counting. In contrast, if the learning effect is an automatic consequence of exposure to the ambiguous fricative in lexically-biased contexts, it should not vary as a function of the task during the exposure phase.

## **2 Method**

### **2.1**

#### ***Participants***

Twenty-nine members of the MPI for Psycholinguistics participant pool took part. They were paid for their participation. None reported any hearing disorders, and none had participated in similar perceptual learning experiments (e.g., those reported by Norris et al., 2003).

### **2.2**

#### ***Materials and Design***

The stimuli were identical to those used in the two experimental conditions in Experiment 2 of Norris et al. (2003). That is, the same physical tokens of all items

were presented in the same orders as before. See Norris et al. for full details, including a list of the critical words and information on how the ambiguous fricative sounds were made, pretested, and spliced onto the critical words.

To summarize, there were 100 Dutch words and 100 phonotactically legal nonwords in the exposure phase, and five syllables on an [ɛf]-[ɛs] continuum during the test phase. Of the exposure words, 20 were [f] final and 20 were [s] final. The sounds [f] and [s] did not occur anywhere else in the exposure materials (and the sounds [v] and [z] never occurred). In the 20 [f]-final words, replacement of the [f] with an [s] made a Dutch nonword. Likewise, in the 20 [s]-final words, replacement of the [s] with [f] made a nonword. There were two exposure conditions. In the first, the [f] in the [f]-final words was replaced, using digital speech editing, with an ambiguous fricative [ʔ], pretested to be approximately midway between [f] and [s]. In the second condition, the [s] in the [s]-final words was replaced with the same ambiguous fricative. Thus, one condition consisted of ambiguous tokens of [f]-final words (e.g., [witlɔʔ], based on *witlof*) and natural [s]-final words (e.g., *naaldbos*), while the other had the reverse: natural [f]-final words (e.g., *witlof*) and ambiguous [s]-final words (e.g., [na:ldɔʔ]). Two different running orders were constructed consisting of all 200 exposure stimuli in pseudorandom order. Participants were randomly assigned to either exposure condition, and, within that condition, heard one of the two running orders. Participants in both exposure conditions then heard the same pseudorandom sequence of six repetitions of the five stimuli on the [ɛf]-[ɛs] test continuum, and categorized the fricative sound in each of these stimuli as either [f] or [s]. This test sequence was identical to that used in the earlier study.

## 2.3

### **Procedure**

Participants were tested individually in a sound-damped booth, as in the experiments reported by Norris et al. (2003). The same kind of equipment as in the earlier study (i.e., headphones and response buttons) was used. For the exposure phase, listeners were instructed that they would hear a list of words and nonwords over the headphones, that they should ignore whether the items were real words or not, and that they should press a response button on every trial, while counting the number of trials that they heard. They were told that, after a proportion of trials, the presentation of auditory items would be suspended, and that they would have the opportunity to write down the number of tokens that they had heard in the preceding block of trials. There were eight blocks each with a different number of trials, varying between 21 and 29. Suspension of auditory stimulus presentation lasted 10s; the message “Hoeveel?” (How many?) appeared on a computer screen in front of the participant after 500ms and remained there for 8s. Participants indicated the number of stimuli that they had counted in each block on a response sheet. The disappearance of the response cue was a signal that the auditory stimuli were about to restart. Participants were asked to start counting from zero again when the list resumed. Other than the delays for the counting responses after each block, timing of stimulus presentation and running orders were the same as in the auditory lexical decision phase in Experiment 2 in Norris et al. (2003). The interonset interval between stimuli within a block was thus 2.6s. The instructions and procedure for the subsequent categorization phase were

the same as in the earlier experiment. At the end of the experiment, participants were asked to report how they performed the counting task.

### 3 Results

#### 3.1

##### Counting

We first examined performance in the exposure phase. Participants reported that they tended to ignore the lexical status of the stimuli in the exposure phase, and that they simply counted the trials as incremental numbers of events in each block. The written responses for the number of trials in each of the eight blocks were scored for accuracy. One participant failed to provide a response on three blocks, and was correct on only one of the other five. The data from this participant were excluded from all subsequent analyses. The remaining 28 participants were quite accurate in keeping count, with an average of 87% correct responses (i.e., approximately 7/8 blocks correct). Furthermore, most incorrect responses were close to being correct (i.e., they were some incorrect number between 20 and 30).

The button press responses which listeners made as they were counting the stimuli were also analyzed. Mean Reaction Times (RTs) and mean miss rates for these responses are given in Table 1. Analyses of Variance (ANOVAs) with both participant ( $F1$ ) and item ( $F2$ ) as the repeated measure, and with the factors exposure condition and final fricative (whether the critical words ended with [f] or [s]), were carried out on both the miss rate and RT data. The miss rates confirmed that participants did indeed perform the button-press task as instructed: A response was logged on almost every trial. There were no significant effects in the miss rate analysis. The RT analysis, however, revealed a pattern which was very similar to that previously observed when the task on these stimuli was lexical decision (Norris et al., 2003). Button-press responses after words with natural fricatives were faster than those after words with ambiguous fricatives. Because of the between-participant exposure manipulation, this effect took the form of an interaction between exposure condition and final fricative:  $F1(1, 26) = 444.60, p < .001$ ;  $F2(1, 38) = 191.90, p < .001$ . Neither main effect was significant by both participants and items.

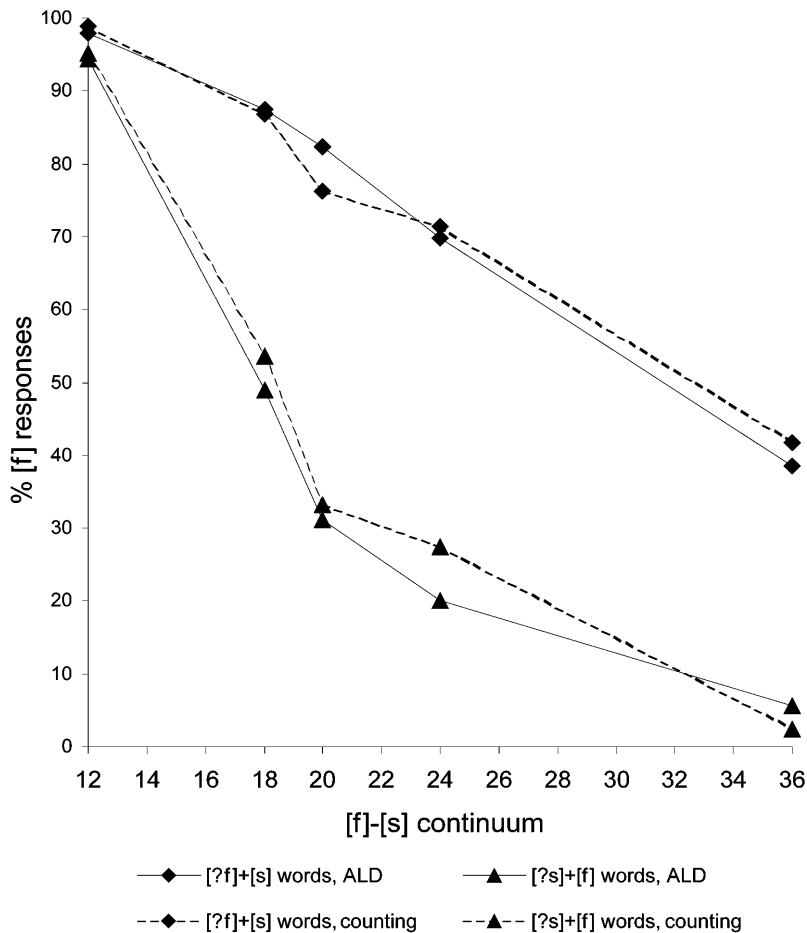
**Table 1**

Counting performance: Mean reaction times (RTs, in ms measured from word offset) and mean percentage miss rates 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	87	35	236	248
Miss rate	1	1	2	2

**Figure 1**

Total proportion of [f] responses to each of the five steps of the fricative continuum as a function of the type of exposure condition, ambiguous [f]-final and natural [s]-final words ([?f]+[s] words) or ambiguous [s]-final and natural [f]-final words ([?s]+[f] words), and as a function of the task during exposure, Auditory Lexical Decision (ALD; data from Norris, McQueen, & Cutler, 2003) or counting (the present experiment)



### 3.2

#### **Categorization**

The effect of exposure condition on phonetic categorization observed when listeners made lexical decisions to the exposure stimuli (Norris et al., 2003) was replicated when listeners counted those stimuli. Overall categorization data from the previous and present experiment are plotted in Figure 1. The functions for the two exposure tasks overlap considerably.

A by-participant ANOVA on the proportion of [f] responses for each of the sounds on the [f]-[s] continuum revealed a significant effect of exposure condition,



$F(1, 26) = 17.49, p < .001$ : Listeners who had heard the ambiguous fricative in [f]-final words during the counting task identified more of the test sounds as [f] than the listeners who had heard the ambiguous fricative in [s]-final words. The effect of continuum step was also significant,  $F(4, 104) = 54.01, p < .001$ , as was the interaction of this factor with exposure condition,  $F(4, 104) = 4.98, p < .005$ , reflecting variability in the magnitude of the exposure effect across the continuum. An ANOVA also compared categorization after counting with that after lexical decision (i.e., comparing the present data with that from the two experimental conditions in Norris et al., 2003). The effect of exposure condition ([?] in [f]- vs. [s]-final words) was again significant,  $F(1, 55) = 45.79, p < .001$ : Both in the previous and the current study exposure to [?] in [f]-biased lexical contexts led to more [f] judgments, and exposure to [?] in [s]-biased lexical contexts led to more [s] judgments. There was, however, no interaction of this effect as a function of exposure task (counting vs. lexical decision;  $F(1) < 1$ ).

## 4 Discussion

Although the listeners in this experiment were not required to perform any explicit task to identify either the words or the sounds in the exposure phase, they still modified their perceptual categorization of the ambiguous phonemes in exactly the same way as the listeners in Norris et al. (2003). They changed their categorization of the fricative sounds in the direction indicated by the lexical information in the exposure phase.

This perceptual learning seems to have taken place as a relatively automatic consequence of listening to words containing ambiguous sounds. The present findings therefore contrast with those on other aspects of perceptual learning in speech perception, where explicit training paradigms are required for reliable learning. Logan et al. (1991), for example, successfully trained Japanese listeners on the English /l/-/r/ distinction using extensive exposure (with multiple talkers) and an explicit forced-choice identification paradigm (with feedback on performance accuracy). A potentially critical difference between this situation and that tested in the present experiment is that the former involves the development of a new phonemic category distinction (the /l/-/r/ distinction is not made in Japanese) whereas the latter involves the adjustment of an existing category distinction. It may therefore be the case that learning new phonemic distinctions requires extensive exposure (10 hours in the Logan et al. study) and the requirement to make explicit judgments about the training stimuli, such that listeners learn to attend to the relevant new stimulus properties. The present results suggest, however, that retuning of existing phonetic categories requires very little exposure (less than 10 mins with a single talker) and no explicit task; such learning is thus relatively automatic.

Note, however, that listeners must have been processing the ambiguous stimuli as words (e.g., [wɪtlɔʔ] as *witlof* or [na:ldɔʔ] as *naaldbos*), otherwise there would be no possibility of lexically-guided learning. Interestingly, the listeners' counting responses were slower to words with ambiguous fricatives than to words with unambiguous fricatives. Despite being instructed simply to count the stimuli, and despite their informal report that this was all that they did, listeners were therefore influenced in some way by the phonetic characteristics of the words, and did not treat the words

merely as acoustic events. Nevertheless, the present results show that the learning effect does not depend on an explicit task which forces listeners to evaluate the fricative sounds.

These findings complement recent results on the generalizability of this perceptual learning effect. McQueen, Cutler, and Norris (in press) found that the lexically-guided perceptual learning effect alters not only listeners' behavior in the phonetic categorization task, but also the way they perceive words that they have not encountered in the exposure phase. The categorization test phase that was used here, and in Norris et al. (2003), was replaced with a cross-modal identity priming task. In that test phase, listeners heard stimuli ending in ambiguous phonemes where the stimulus formed a word regardless of whether the ambiguous phoneme was interpreted as [f] or as [s]: [do:ʔ], for example, which can be either *doof* ('deaf') or *doos* ('box'). At the acoustic offset of each of these ambiguous prime words (or after a phonologically unrelated prime), listeners were presented visually with a word corresponding to one of the two possible interpretations of the ambiguous stimulus. Listeners who had experienced the exposure condition where lexical information indicated that the ambiguous sound should be interpreted as [f] showed priming in their responses to the f-final visual words (e.g., faster responses to DOOF after [do:ʔ] than after an unrelated prime), and those who learned that the ambiguous sound should be interpreted as [s] showed priming in their responses to s-final words such as DOOS.

This result shows that lexically-guided learning influences the word recognition process itself, and not only the process responsible for making explicit decisions in a phonetic categorization task. McQueen et al. (in press) argue that the locus of the perceptual learning effect must be prelexical (i.e., at a stage of processing which precedes lexical access). If the learning took place at the lexical level, or at a postlexical level, it would not influence the processing of new words. The generalizability of the learning effect thus challenges models of speech recognition in which there is no prelexical level with abstract phonological representations (e.g., the template-matching model of Klatt, 1979, and the episodic model of Goldinger, 1998). The prelexical level provides a stage of processing where learning about phonetic categories can improve communicative efficacy. If a given listener has learned in the exposure phase that when the speaker utters the ambiguous sound [ʔ] (s)he means [f], and if this learning is implemented as an adjustment in prelexical processing, then the adjustment will help the listener to recognize an utterance such as [do:ʔ] as a token of the word *doof* (and hence not as either the word *doos* or an ambiguous token). The retuning process thus pays off through facilitation of communication with the current speaker.

Taken together, all of these studies of perceptual learning indicate that listeners can rapidly modify their perception of utterances produced by a speaker whose speech has unusual characteristics. As Norris et al. (2003) argue, lexical involvement in this retuning process suggests that there is some kind of supervised learning in the speech perception system, with lexical knowledge providing a training signal through feedback of information to the prelexical level. It is important to distinguish this type of feedback—off-line feedback for perceptual learning, which does not alter the immediate processing of the current input—from immediate feedback which acts to alter the prelexical representation of the speech input as that input is being heard (see Norris et al. for further discussion). Critically, only feedback for perceptual

learning is beneficial to the listener. Lexically-guided learning has considerable adaptive value in helping listeners to understand a specific individual who speaks in an unusual fashion.

This kind of learning would not be adaptive, however, if an encounter with a single unusual speaker modified one's perceptual categories for all other speakers. Eisner and McQueen (2005) addressed this issue in a modified version of the Norris et al. (2003) paradigm by testing whether exposure to a female speaker would generalize to the processing of test utterances originating from a male speaker. They found no evidence for generalization unless exactly the same fricative sound was used in both exposure and test. The learning is thus highly specific, so much so that it appears that one can fool the perceptual system into making adjustments by splicing fricative sounds originating from the female speaker used in the exposure phase into the speech of a male speaker for use in the test phase. But there is no generalization of learning to another speaker when the information about speaker identity in the fricative sounds differs between the exposure and test phases. This is exactly what one would expect. Listeners should not generalize from a single speaker to the entire language community. The implication of this result is that listeners must be able to maintain more than one set of perceptual categories at the same time. That is, the modified perceptual categories that can be applied to a further encounter with the speaker heard in the exposure phase must be maintained alongside the normal categories that continue to be applied to other speakers.

This perceptual flexibility has obvious advantages outside the laboratory setting. Most listeners encounter speakers with a range of widely differing regional accents. Some of these accents can only be understood with considerable experience. For example, speakers of British English require experience with American English in order to learn that an alveolar flap [ɾ] in intervocalic position is an instance of the phoneme /t/ (Scott & Cutler, 1984). However, there would be no benefit in learning how to interpret a new accent if that interfered with the perception of the majority of speakers that one normally listened to.

The fact that listeners must be able to maintain alternative perceptual categories reinforces the message of the McQueen et al. (in press) study. If listeners modify their prelexical representations, each new accent can be accommodated by adding a small number of new phonological representations, and/or by adding knowledge about where the boundaries between existing categories lie in the new accent. These modifications can then be used to assist processing whenever listeners encounter words spoken in that accent, even if they have never heard those words in that accent before, or even if they have never heard a particular word ever before (i.e., a completely new word, perhaps specific to that dialect). If, however, listeners have no such abstract prelexical phonological representations, as implied by episodic theories (Goldinger, 1998), each new accent will require a completely new set of lexical representations. That is, each new accent effectively requires a new lexicon of word forms. Listeners would therefore only be able to learn an accent by learning about each word individually and would not be able to generalize learning to the processing of novel words on the basis of exposure to a limited subset of words.

More research is required to delineate further the nature of the lexically-guided perceptual learning effect. The exposure conditions tested so far with the Norris et al. (2003) paradigm have used only one speaker. It is possible that exposure to multiple speakers would induce learning that would generalize to novel speakers. This would thus be the laboratory analog of learning about an accent group, rather than about the idiosyncracies of an individual speaker. Bradlow and Bent (2003) have indeed shown, using a different experimental paradigm, that exposure to multiple talkers of Chinese-accented English led to improvement in recognition of words spoken by a novel speaker of Chinese-accented English. It will therefore be important to establish when learning in the Norris et al. paradigm is speaker-specific and when it generalizes over speakers. One factor determining specificity may be the phonetic contrast under manipulation. Kraljic and Samuel (in press) show some generalizability of the lexically-guided perceptual learning effect across speakers when stop consonants were manipulated (in contrast to the fricatives tested here and by Eisner & McQueen, 2005). One explanation for this difference may be that fricative sounds carry more speaker-specific information than stops do.

The results to date with this paradigm nevertheless suggest that listeners come equipped to deal with variability in speech sounds—whether that variability arises from differences among speakers or within a given speaker over time. We have argued that this is because of the distinction between prelexical and lexical levels in the speech recognition system, and because of the flexibility of that system. A dynamic prelexical level allows efficient coding of the acoustic-phonetic differences that characterize individuals or accent groups, and allows that knowledge to be used to improve subsequent recognition of novel words spoken by that individual or group. We have also suggested that this payoff in communicative efficacy may be one of the driving forces in language change. If a listener (majestic or otherwise) adjusts her perceptual models of a phonetic category through experience with other speakers in a community, and if these new models are stable enough (i.e., if they are true of the community as a whole, rather than of particular individuals), then this may result in changes in how that listener herself produces tokens of that category. The flexibility in the speech perception system over the lifespan may thus be one factor which determines how the operation of the speech production system changes over time.

Furthermore, the present results suggest that these changes in the perceptual system are largely automatic; they do not depend on explicit judgments during exposure to unusual speech. The view that emerges, therefore, is of a speech recognition system which is inherently dynamic. It is not the case that, after initial stages of language acquisition, the speech perception system in an individual is stable for the rest of that person's life. Instead, speech perception appears to change as the speech around that person changes.

## References

- AHISSAR, E., & HOCHSTEIN, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences U.S.A.*, **95**, 869–875.
- BAUER, L. (1985). Tracking phonetic change in the received pronunciation of British English. *Journal of Phonetics*, **13**, 61–81.

- BRADLOW, A. R., & BENT, T. (2003). Listener adaptation to foreign-accented speech. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Conference of Phonetic Sciences* (2881–2884). Adelaide: Causal Productions (CDROM).
- EISNER, F., & McQUEEN, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception and Psychophysics*, **67**, 224–238.
- FRANCIS, A. L., BALDWIN, K., & NUSBAUM, H. C. (2002). Effects of training on attention to acoustic cues. *Perception and 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.
- GOLDINGER, S. D. (1998). Echoes of echoes?: An episodic theory of lexical access. *Psychological Review*, **105**, 251–279.
- HARRINGTON, J., PALETHORPE, S., & WATSON, C. I. (2000a). Does the Queen speak the Queen's English? *Nature*, **408**, 927–928.
- HARRINGTON, J., PALETHORPE, S., & WATSON, C. (2000b). Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen's Christmas broadcasts. *Journal of the International Phonetic Association*, **30**, 63–78.
- HARRINGTON, J., PALETHORPE, S., & WATSON, C. (2005). Deepening or lessening the divide between diphthongs: An analysis of the Queen's annual Christmas broadcasts. In W. J. Hardcastle & J. M. Beck (Eds.), *A figure of speech: A festschrift for John Laver* (pp. 227–261). Mahwah, NJ: Lawrence Erlbaum Associates.
- HODGE, F. S., COLTON, R. H., & KELLEY, R. T. (2001). Vocal intensity characteristics in normal and elderly speakers. *Journal of Voice*, **15**, 503–511.
- KLATT, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, **7**, 279–312.
- KRALJIC, T., & SAMUEL, A. G. (in press). How general is perceptual learning for speech? *Psychonomic Bulletin and Review*.
- 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.
- McQUEEN, J. M., CUTLER, A., & NORRIS, D. (in press). Phonological abstraction in the mental lexicon. *Cognitive Science*.
- NORRIS, D., McQUEEN, J. M., & CUTLER, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, **47**, 204–238.
- PYLYSHYN, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, **22**, 341–423.
- RUSSELL, A., PENNY, L., & PEMBERTON, C. (1995). Speaking fundamental-frequency changes over time in women—a longitudinal study. *Journal of Speech and Hearing Research*, **38**, 101–109.
- SCOTT, D. R., & CUTLER, A. (1984). Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior*, **23**, 450–466.
- 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.
- YAEGER-DROR, M. (1994). Phonetic evidence for sound change in Québec French. In P. A. Keating (Ed.), *Phonological structure and phonetic form: Papers in laboratory phonology III* (pp. 267–292). Cambridge: Cambridge University Press.