Perception & Psychophysics
2005, 67 (2), 224-238

The specificity of perceptual learning in speech processing

FRANK EISNER and JAMES M. McQUEEN
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

We conducted four experiments to investigate the specificity of perceptual adjustments made to unusual speech sounds. Dutch listeners heard a female talker produce an ambiguous fricative [?] (between [f] and [s]) in [f]- or [s]-biased lexical contexts. Listeners with [f]-biased exposure (e.g., [wɪtlo?]; from *witlof*, "chicory"; *witlos* is meaningless) subsequently categorized more sounds on an [ɛf]–[ɛs] continuum as [f] than did listeners with [s]-biased exposure. This occurred when the continuum was based on the exposure talker's speech (Experiment 1), and when the same test fricatives appeared after vowels spoken by novel female and male talkers (Experiments 1 and 2). When the continuum was made entirely from a novel talker's speech, there was no exposure effect (Experiment 3) unless fricatives from that talker had been spliced into the exposure talker's speech during exposure (Experiment 4). We conclude that perceptual learning about idiosyncratic speech is applied at a segmental level and is, under these exposure conditions, talker specific.

This series of experiments concerns the nature of perceptual adjustments that take place in the speech recognition system in response to unusual speech production. The process of decoding speech is necessarily complex, to a great extent because, in addition to structural variation such as coarticulation, speech is characterized by a large amount of both inter- and intratalker variability. The realization of a given phoneme varies within individuals as a function of, for example, voice quality, emotional state, or speaking rate. Interindividual differences, the focus of the present series of experiments, are caused by factors such as vocal tract shape, accent, or articulatory habits (see, e.g., Klatt, 1986, 1989). The cumulative effect of all these sources of variability is that mapping from input to categories is a many-to-many problem: Not only can one phoneme have different acoustic realizations, but one acoustic pattern can elicit different phonemic percepts (Nusbaum & Magnuson, 1997; Repp & Liberman, 1987). How do listeners deal with this variability? A number of previous studies have shown that the perceptual system dynamically adjusts to speech that is initially difficult to understand. The characteristics of the mechanism that achieves perceptual constancy and the constraints under which it operates are largely unknown, however. In this study, we asked whether talker-specific adjustments are made in the speech recognition system in response to unusual productions of speech sounds, and if so, how detailed those adjustments are.

Evidence of such dynamic adjustments, as indexed by improved intelligibility after sufficient exposure, has been found with synthetic speech (Greenspan, Nusbaum, & Pisoni, 1988; see also Maye, Aslin, & Tanenhaus, 2003), noise-vocoded speech (Hervais-Adelman, Johnsrude, Davis, & Brent, 2002), and compressed speech (Dupoux & Green, 1997; Mehler et al., 1993). Such studies have revealed important constraints on perceptual learning in speech processing. Greenspan et al., for example, note that variability in the training materials is crucial for learning; repetition of a small set of stimuli did not produce improved intelligibility in their study. Hervais-Adelman et al. observed that adaptation to noise-vocoded speech was absent when listeners were presented with phonotactically legal nonword sentences. This suggests that higher level (e.g., lexical) information is required for adaptation to occur.

Perceptual learning has also been found in response to natural but accented speech input. Moving to a different dialectal environment often requires adaptation to unfamiliar input from a whole community of talkers. British English speakers who have lived in the United States, for example, learn to recognize an alveolar tap [ɾ] as an instance of [t] (Scott & Cutler, 1984). Similarly, American immigrants to Britain may have to learn that the glottal stop [ʔ] is an instance of the same phoneme. Adjustments are also made in response to talkers who speak a language with a nonnative accent. Clarke (2002, 2003) observed that after short exposure to Spanish-accented American English, listeners performed faster on a task that required matching a visual stimulus to accented auditory input than did control listeners who had had exposure to another

F.E. was supported by a doctoral grant from the Max Planck Society for the Advancement of Science. We thank Anne Cutler, Dennis Norris, Steve Goldinger, and Jim Sawusch for helpful comments on earlier versions of the manuscript. Part of this work was presented at the 2nd Dutch Endo-Neuro-Psycho (ENP) Meeting in Doorwerth, The Netherlands in June 2003 and at the Tutorials in Behavioural and Brain Sciences (TuBBS) Meeting in Grimma, Germany in July 2003. Correspondence concerning this article should be addressed to either author at Max Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands (e-mail: frank.eisner@mpi.nl or james.mcqueen@mpi.nl).

voice talking in nonaccented English. Bradlow and Bent (2003), in a training–test paradigm, investigated perceptual adjustment to Chinese-accented English. One group of listeners who had heard multiple talkers, and another group that had heard only the test talker at training, performed equivalently, and better than other training groups, on a transcription task. Listeners who had heard only a single talker at training, one who was different from the test talker, did not show improved performance. These results suggest that perceptual adaptation is useful when the same talker is encountered again and, furthermore, that adaptation in response to a single talker does not generalize to another talker. However, if there is variability in the input, as introduced by multiple talkers, the perceptual system appears to be able to extract abstract information about the accent that can be used to facilitate comprehension of other talkers with the same accent. Because of the nature of Bradlow and Bent's task, however, the type of information extracted (e.g., featural, segmental, prosodic, or rhythmic) cannot be determined. Nevertheless, as was shown in another study (Evans & Iverson, 2004), it is possible that learned characteristics of an accent can constrain the interpretation of subtle phonetic cues.

Some studies on accent normalization have used intelligibility of words or sentences as a dependent measure and hence provide few cues as to the level or detail of adjustment. Others have examined the role of phonetic detail in processing accented speech (Evans & Iverson, 2004; Scott & Cutler, 1984), but adjustments in these cases were the outcome of exposure to a whole language community, possibly over many years, and are therefore not necessarily carried out by the same mechanism as that responsible for individual talker normalization. In the present study, in contrast, we sought to evaluate the degree of detail that listeners learn about the characteristics of an individual talker's speech after short-term exposure.

There is abundant evidence that listeners make perceptual adjustments to the speech of individual talkers. Classic studies by Ladefoged and Broadbent (1957) and Ladefoged (1989) have shown that listeners evaluate a talker's vowel space and apply this computation in interpreting following vowels within the same utterance. More recently, Nygaard, Sommers, and Pisoni (1994) found that spoken word identification was improved for listeners who had previously been familiarized with the talkers' voices, compared with control listeners who had been familiarized with another set of voices (see also Nygaard & Pisoni, 1998). Their findings suggest that once a listener has adjusted to the idiosyncrasies of a particular talker's utterances, the result of this process is stored and will be used again to facilitate perception when this voice is encountered at a later point—a conclusion that is in line with the recurrent observation that listeners encode details of talkers' voices in long-term memory (Church & Schacter, 1994; Goldinger, 1996, 1998; Goldinger, Pisoni, & Logan, 1991; Martin, Mullennix, Pisoni, & Summers, 1989; Palmeri, Goldinger, & Pisoni, 1993; Pisoni, 1993).

In an earlier study, Mullennix and Pisoni (1990) investigated directly a possible influence of talker-specific information on linguistic processing. They employed a same–different classification task of either the dimension voice or the dimension phoneme (a word-initial voicing contrast). While the respective other dimension was always varied in the experimental conditions, response latencies were compared with a control condition where the other dimension was held constant. Results showed that variation of these two dimensions produced mutual interference, suggesting that they are not processed independently of each other. However, there was an asymmetry such that variation in voice caused more interference with phoneme classification than vice versa. Given this asymmetry, Mullennix and Pisoni concluded that linguistic processing is contingent on voice processing—more specifically, that talker information is extracted from the signal first and then influences phonetic processing (see also Green, Tomiak, & Kuhl, 1997; Knösche, Lattner, Maess, Schauer, & Friederici, 2002; Lattner, 2002).

Another observation on the constraints of a perceptual learning mechanism is that the initial adjustment to a talker comes at a processing cost. Mullennix, Pisoni, and Martin (1989) reported that identification and naming of a list of words in noise deteriorates and slows down when these words are produced by multiple, intermixed talkers, relative to when they are produced by the same talker. Mullennix et al. proposed that the perceptual system must engage in an adjustment process each time a novel voice is encountered. On the other hand, when there is only one talker in the set, the system is already in the right configuration at the time a word is presented, leading to better identification performance and shorter response latencies. Nusbaum and Morin (1992) reported a similar and consistent effect of multiple-talker compared with single-talker presentations in response latencies to vowels, consonants, and words.

A recent study by Norris, McQueen, and Cutler (2003) provides some insight into how a perceptual learning mechanism in speech perception might operate. This study demonstrated a lexically driven modulation of the category boundary for a consonant contrast, which was induced in an exposure phase and measured in a subsequent phonetic categorization task. In the exposure phase, listeners heard naturally produced words, some of which were edited. For one group of listeners, all instances of the fricative sound [s] were replaced by a perceptually ambiguous sound lying midway between [s] and [f]. For another group of listeners, all cases of [f] were replaced by the same ambiguous fricative sound. Results showed that the group that had heard the ambiguous sound in [s]-biased lexical contexts categorized more sounds on an [f]–[s] continuum as [s], whereas the other group categorized most sounds as [f]. In accord with what Hervais-Adelman et al. (2002) found for noise-vocoded speech, this study thus shows that a perceptual adjustment is made when an idiosyncratic production of a speech sound is placed in an appropriate lexical context.

Previous studies have therefore shown that the speech perception system makes adjustments to both natural speech and to speech that is in some way unusual. The adjustment requires processing capacity, and evidence has been found for one specific adjustment mechanism, which is lexically driven. Patterns extracted from these adjustments are stored, and the information is reused when speech with similar characteristics is encountered again.

A number of important questions about the constraints of perceptual learning remain unanswered, however. In the present study, we addressed two issues regarding its specificity. First, it is not clear how detailed the adjustments are: Are adjustments made at a segmental level (i.e., with respect to individual phonemes), at a lexical level (i.e., with respect to individual words), or more globally (i.e., with respect to pitch characteristics of a talker's voice)? Second, it is not clear whether the effect of perceptual learning is applied talker specifically, or whether it also affects processing of speech from other talkers. Although studies on accent learning have shown that the outcome of perceptual adjustment is beneficial for comprehension when, subsequently, talkers with the same accent are encountered (Bradlow & Bent, 2003; Scott & Cutler, 1984), it is uncertain whether such learning may be misapplied to other talkers who do not have that accent. Similarly, the outcome of individual talker normalization is clearly beneficial when one is listening to the speech of the same talker again (Nygaard et al., 1994) but may have a detrimental effect when applied to another talker. These two issues, that of the level and detail of application of learning, and that of generalization to other talkers, were investigated using the exposure–test paradigm developed by Norris et al. (2003). We chose this paradigm because it provides tight control over the learning effect; the bias in the interpretation of an ambiguous sound is determined by lexical knowledge alone, not by differences in the ambiguous sound, or any other sounds, between conditions. It is therefore well suited to test whether the learning effect is specific to a phonetic contrast alone. Furthermore, the paradigm allows testing for talker specificity of the adjustment. Listeners were exposed to edited, natural speech coming from one talker, and then tested for a perceptual learning effect with materials made from another talker's utterances.

EXPERIMENT 1

In Experiment 1, we sought to examine whether perceptual learning after exposure to a (female) talker with unusual fricative productions would generalize to a test situation where listeners are presented with a new (female) talker. Conditions where the talker at test and the talker at exposure were the same (replicating the experimental conditions of Norris et al., 2003) served as a comparison for the talker-change conditions. Two further control conditions (identical to the nonword conditions of Norris et al., 2003) were included to provide a measure of the extent to which the adjustment is lexically driven.

A pretest was conducted in order to find a fricative [f]–[s] sound that was sufficiently ambiguous to Dutch listeners. The main experiment then consisted of an exposure phase (auditory lexical decision) followed by a brief test phase (phonetic categorization). Four exposure conditions were defined by the types of words and nonwords used in the lexical decision task. In one experimental condition, all 20 instances of [s] (in word-final position) were replaced by a perceptually ambiguous sound [?], whereas all 20 [f] sounds (also in word-final position) remained natural. A second condition consisted of items in which all the [f] sounds were replaced by [?], but all [s]s were natural productions. Two control groups listened to the ambiguous sound [?] in nonword contexts, one of which additionally received naturally produced [f]-final words, the other group receiving natural [s]-final words. As in the Norris et al. (2003) study, these groups were used to control for the possibility that an effect in the experimental conditions was due to selective adaptation or contrast effects, as opposed to a lexical effect (see Norris et al., 2003, for a discussion). These four groups were then tested on an ambiguous [ɛf]–[ɛs] continuum, made from materials constructed from utterances of the talker of the exposure phase. Given that these conditions were an exact replication of the Norris et al. study, using the same words and procedure but a different talker, we expected to replicate the earlier results. The experimental exposure group that listened to ambiguous [f]-final and natural [s]-final words was expected to subsequently categorize more sounds on the [ɛf]–[ɛs] continuum as [f], whereas the other experimental exposure group was expected to categorize more sounds as [s]. The control groups were expected to give intermediate responses and not to differ from each other.

Our main interest, however, was in two further groups of participants who listened to the stimuli of the two lexically biased exposure conditions, but were then tested on an [ɛf]–[ɛs] continuum in which the vowel [ɛ] came from an utterance by a novel talker who was also female and was similar in age to the exposure talker. Since vowels are a rich source of talker identity information, this manipulation was expected to signal a change in talkers between exposure and test. If perceptual learning generalizes to another talker, a shift in category boundary as a function of exposure condition should be evident in the categorization data. That is, the categorization data for these groups should show the same pattern as those for listeners in the lexically biased exposure conditions who were tested on the exposure talker. If, however, perceptual learning does not generalize to a different talker, listeners would not be expected to apply a previously learned adjustment when they noticed a change in talkers. No difference in categorization performance would then be expected between the two novel talker test groups.

Method

Participants

A total of 105 native speakers of Dutch drawn from the MPI for Psycholinguistics participant pool took part in the experiment. Nine volunteers participated in the pretest and the remaining 96 in the

main experiment. None of them reported having any hearing disorders. All were paid for their participation.

Pretest

Stimulus construction. A number of tokens of the three syllables [ɛf], [ɛs], and [ɛx], produced by a female native speaker of Dutch, were recorded in a sound-attenuated booth onto digital audio tape (DAT). Recordings were redigitized at a 16-kHz sampling rate and 16-bit quantization on a Sun Sparc workstation and edited with Xwaves. One token each of [ɛf] and [ɛs] was selected to create an [ɛf]–[ɛs] continuum. The fricatives were excised from the vowel at a zero crossing at the onset of frication energy and edited to match the mean duration and intensity of [f] and [s] in spoken word contexts. These mean duration and intensity values (202.4 msec and 55.2 dB SPL) were derived from measurements of the experimental items recorded for the lexical decision part of the experiment (see below). The waveforms of both fricatives were cut, then linearly smoothed at offset over a 75-msec window, and finally scaled to be of equal intensity. The resulting [s] and [f] sounds were then used to make a 21-step continuum, employing an algorithm that combined each of the two sounds sample by sample in 21 graded proportions, such that Step 1 was the original [f] and Step 21 the original [s], with 19 equally spaced steps in between (McQueen, 1991). Each step was then spliced onto the vowel [ɛ], which was isolated from one of the [ɛx] syllables and which was 112 msec in duration. A vowel from a velar context was used in order to avoid transitional cues to the labiodental [f] or alveolar [s] place of articulation. (Note that Norris et al., 2003, used vowel tokens that always cued labiodental place, which resulted in a residual [f]-bias.)

Procedure. Informal listening by 4 native Dutch speakers indicated that the most ambiguous range of the [ɛf]–[ɛs] continuum was Steps 6–15. These 10 syllables were presented to the pretest listeners over closed headphones in a sound-attenuated booth. Items were pseudorandomized by concatenating 10 individually randomized lists containing one of each syllable. There was a short practice sequence in which each token was played once. Responses were made by pressing one of two buttons labeled "F" and "S," counterbalanced for handedness across the sample such that half of the participants made "F" responses and the other half "S" responses with their dominant hand. Items were presented at a rate of 2.6 sec between syllable onsets.

Results. Percentages of [f] responses to each of the 10 steps of the continuum were averaged. The continuum was judged by listeners to be most ambiguous (50% [f] responses) at the point midway between Steps 10 and 11. Hence a more fine-grained 41-step continuum was made from the endpoint stimuli using the technique described above. Step 20 corresponded to Step 10.5 on the 21-step continuum. This step [?] was then used to make the ambiguous items in the exposure phase of the main experiment and, along with Steps 12, 17, 23, and 28 (corresponding to 85%, 70%, 30%, and 15% of [f] responses, respectively), was also used in the phonetic categorization phase.

Materials and Stimulus Construction

Lexical decision. Stimuli were constructed for two experimental and two control conditions, using new recordings of the items used by Norris et al. (2003). Experimental words and nonwords, as well as filler words and nonwords, were produced by the talker of the pretest and recorded during the same session. Experimental items were 20 [f]-final Dutch words (e.g., *olijf*, "olive"), 20 [s]-final Dutch words (e.g., *radijs*, "radish"), and 20 strings that would be nonwords whether they ended in [f] or [s] (e.g., *kwirtaf*, *kwirtas*). Note that *olijs* and *radijf* are not words in Dutch. These three sets were matched in triplets for stress pattern, final vowel, and length (such that there were five items per set with one, two, three, and four syllables). The two real-word sets were also matched for frequency (13 per million for [f]-final words and 14 per million for [s]-final words). Except for the final [f] and [s] in the real-word sets, no experimental item contained any further instances of these two sounds, nor of [v] or [z]. In addition, there were 80 filler words and 100 filler nonwords, with each of these sets consisting of an equal proportion of items with one, two, three, and four syllables. None of the fillers contained the sounds [f], [s], [v], or [z]. The full set of experimental materials is listed in the study by Norris et al.

There were two versions of each experimental word. One was a natural pronunciation, but in the second version the final fricative was replaced by the ambiguous sound [?] (e.g., *olij?*). To ensure that any transitional information in the final vowel did not cue [f] or [s] and was consistent across sets, ambiguous versions were made from recordings in which the final phoneme was intentionally mispronounced as the velar fricative [x] (e.g., [olɛɪf] as [olɛɪx]). This velar fricative was then excised from the preceding vowel at a zero crossing at the onset of frication, and replaced by [?]. Experimental nonwords were also created from recordings with a final velar fricative.

Phonetic categorization. For one pair of experimental exposure groups and the two control exposure groups, the items of the categorization phase were those that had been selected on the basis of the pretest, in the context of a vowel from the same talker (Talker 1). The other pair of experimental exposure groups listened to test stimuli that had been constructed by splicing these same five [f]–[s] steps onto a vowel that had been produced by a different female talker (Talker 2). This vowel [ɛ] was, as with all other spliced items in the experiment, taken from a velar context. A number of tokens of [ɛx] were produced by a female native speaker of Dutch of similar age to Talker 1. Recordings were made in a sound-attenuated booth onto DAT, then digitally transferred to a computer (48-kHz sampling rate and 16-bit quantization), downsampled to 16 kHz, and edited using Xwaves. A token of [ɛ] (171 msec in duration) was isolated at a zero crossing at the onset of frication and equated in intensity to the vowel of Talker 1 (67.5 dB SPL) before being spliced onto the five fricative steps.

Design and Procedure

Lexical decision. There were four exposure conditions, each with 100 words and 100 nonwords. In one experimental condition, there were the 20 natural [f]-final words and the ambiguous versions of the 20 [s]-final words (e.g., *olijf* and *radij?*). In addition, there were 60 filler words (15 of each of the four lengths) and 100 nonwords (25 of each length). The second experimental condition was identical, except that this list contained the natural versions of the 20 [s]-final words and the ambiguous versions of the 20 [f]-final words (e.g., *olij?* and *radijs*). Two control conditions consisted of the natural recordings of the experimental words, 20 [f]-final items in one and 20 [s]-final items in the other. The listeners in both control conditions also heard the 20 [?]-final experimental nonwords, plus 80 filler words (20 of each length) and 80 filler nonwords (20 of each length).

The 96 participants were assigned to one of six groups (16 participants per group). The two experimental exposure conditions each had two groups, differing only in the stimuli used at test—one that would hear Talker 1 in the test phase and one that would hear Talker 2. The two control exposure conditions each had one group of listeners. The stimuli were presented in a pseudorandomized running order in which experimental items did not occur on the first 12 trials but were otherwise spread equally across the course of the experiment, with at least four fillers between two experimental items. The running orders for the four conditions were identical to the extent that the appropriate experimental items always appeared in the same positions (i.e., the slot in which one experimental condition contained the natural version of a word would be filled by the ambiguous version of that word in the other experimental condition, and vice versa). The control conditions were based on the experimental conditions such that the natural versions of experimental words were in the same positions, and ambiguous versions were replaced by nonwords. To maintain an equal number of words and

nonwords, 20 filler nonwords were replaced with filler words in the control conditions.

Up to 4 participants were tested at a time in a quiet room and were presented with stimuli binaurally at a comfortable listening level over closed headphones, with an interonset interval (ISI) of 2.6 sec. They were instructed (on a computer screen) to decide as fast and as accurately as possible whether or not each item was a real Dutch word, and to respond by pressing one of two buttons labeled *Ja* ("yes") and *Nee* ("no"). The participants were further told that there would be a short second part for which they would be given instructions onscreen after the lexical decision task. They were therefore unaware, during the lexical decision phase, that they would be tested later on fricative perception. Half of the participants in each condition gave "yes" responses and the other half gave "no" responses with their dominant hand.

Phonetic categorization. The phonetic categorization task followed immediately after the exposure phase and was exactly the same for the six conditions, except that two of the four experimental groups listened to slightly different stimuli—that is, stimuli that were made with a vowel from Talker 2. Six repetitions of each of the five steps from the [ɛf]–[ɛs] continuum were presented at an ISI of 2.6 sec. The order of presentation was pseudorandomized to ensure that the five steps were spread evenly across the list and that no step would occur twice in a row. The participants were given onscreen instructions to press a button labeled "F" when they heard an [f]-like sound or a button labeled "S" for an [s]-like sound. Again, the position of the labels was counterbalanced for handedness. Unlike in the pretest, there was no practice block.

Questionnaire. The participants who listened to Talker 1 in the categorization phase were given a short questionnaire at the end of the experiment in which they were asked open questions as to whether they noticed anything unusual in the lexical decision part of the experiment, and if so, whether they were conscious of taking this into account when making their responses in either part of the experiment. The participants who listened to Talker 2 were given two different questions intended to determine whether listeners noticed the talker change. The first question asked whether any difference between the two parts had been noticed. Unless the spontaneous answer was that there had been a talker change, the second question then asked explicitly whether listeners thought that the voices in the two parts were the same or different.

Results

Lexical Decision

Performance in the lexical decision task was used as a criterion for exclusion of participants in the experimental conditions. If participants failed to label at least 50% of experimental words (ambiguous or natural versions) as existing words, they were excluded from further analyses (as in Norris et al., 2003, we excluded these participants because, first, given their unwillingness to label the experimental items as words, it is difficult to interpret their categorization data, and second, failure to label unambiguous items as words most of the time indicates poor compliance with the instructions). In the experimental groups that heard ambiguous [s]-final words, 4 participants fell below this cutoff point (2 in the same-talker condition and 2 in the different-talker condition). In one of the groups that heard ambiguous [f]-final words, 1 participant fell below the cutoff (same-talker condition).

The lexical decision data were analyzed in order to determine how acceptable the ambiguous items were compared with the natural items, and how similar (in terms of acceptability) the [?]-final [f]- and [s]-words were to each other. Mixed 2×2 analyses of variance (ANOVAs) were performed by subjects and by items on the reaction times (RTs, adjusted to measure from word offset) for "yes" responses to experimental words and (separately) on the mean percentages of "no" responses.

The factor exposure group (the two experimental conditions) was a between-subjects factor for the subjects analyses and a within-subjects factor for the items analyses, whereas the second factor final fricative (whether the original word ended in [f] or [s]) was a between-subjects factor for the items analyses but a within-subjects factor for the subjects analyses. Tests were performed separately for the same- and different-talker training groups. A summary of mean RTs for "yes" responses to experimental items is given in Table 1. Overall, listeners were faster to label the natural versions as words than to label the ambiguous versions as words (mean RTs of 188 and 240 msec, respectively), where RTs were slowest for the ambiguous [s]-final items. This difference was reflected in the analysis as a significant interaction between the factors of final fricative and exposure group in both the same-talker exposure groups [$F_1(1,27) = 6.10, p < .05$; $F_2(1,38) = 20.15, p < .001$] and the different-talker groups [$F_1(1,28) = 18.80, p < .001$; $F_2(1,38) = 21.60, p < .001$]. Neither of the main effects was significant. Our results are similar to those obtained by Norris et al. (2003).

Table 1 also shows percentages of "no" responses to experimental items. Listeners were more likely to accept

Table 1
Mean Reaction Times and Mean Percentage "No" Responses in Lexical Decision in Experiments 1–4

	Experiment 1*		Experiment 2		Experiment 3		Experiment 4	
	RT	% "No"	RT	% "No"	RT	% "No"	RT	% "No"
Natural fricatives								
[f]-final words	188	2	222	3	173	3	295	1
[s]-final words	187	7	226	3	183	9	250	12
Ambiguous fricatives								
[f]-final words	209	4	224	2	196	4	280	5
[s]-final words	272	23	265	22	207	20	288	28

Note—Mean reaction times (RTs, in msec, from word offset) are for "yes" responses only. *In Experiment 1, the data presented here are the combined results across the four experimental groups.

the natural versions as existing words than to accept the ambiguous versions: On average, they rejected 5% of the natural items and 14% of the ambiguous ones. The relatively high percentage of 23% "no" responses to ambiguous versions of [s]-final words appears to be due mainly to mono- and bisyllabic items. There were four items that 40% or more of all participants responded "no" to, all of which were mono- or bisyllabic [?]-final [s]-words. Again, this pattern of results replicates Norris et al. (2003).

The overall difference of 9% in "no" responses to natural vs. ambiguous items was significant in the same-talker groups [$F_1(1,27) = 29.28, p < .001$; $F_2(1,38) = 8.38, p < .01$] and in the different-talker groups [$F_1(1,28) = 25.93$, $p < .001$; $F_2(1,38) = 15.28, p < .001$]. The main effects were significant in both the same-talker groups [final fricative, $F_1(1,27) = 58.21, p < .001$; $F_2(1,38) = 7.73, p < .01$; exposure group, $F_1(1,27) = 14.18, p < .005$; $F_2(1,38) = 4.26, p < .05$] and the different-talker groups [final fricative, $F_1(1,28) = 50.21, p < .001$; $F_2(1,38) = 13.76, p < .005$; exposure group, $F_1(1,28) = 7.43, p < .05$; $F_2(1,38) = 7.66, p < .01$]. In short, the results from the same- and different-talker groups were very similar to each other and to the results obtained by Norris et al. (2003). Listeners labeled most of the [?]-final items as words.

Phonetic Categorization

The primary data, however, are those from the test phase. The mean percentages of [f] responses to the five continuum steps are plotted for the six groups in Figure 1. In the same-talker conditions, participants who heard the ambiguous [?] in [f]-final words during exposure labeled the continuum mostly as [f], whereas those who heard [?] in [s]-final words during exposure categorized most sounds as [s]. Averaged across steps, this constitutes a 41% difference between groups. The listeners in the two control exposure conditions gave intermediate responses. In an ANOVA on the percentage of [f] responses with step as a within-subjects factor and training condition (experimental vs. control) and fricative type (natural [f]-final vs. natural [s]-final words at exposure) as between-subjects factors, there was a significant effect of step [$F(4,228) = 28.61, p < .01$], indicating that the percentage of [f] responses varied overall across the continuum. The three-way interaction of step, training condition, and fricative type was also significant [$F(4,228) = 3.04, p < .05$]. There was also a significant interaction of training condition and fricative type [$F(1,57) = 9.03, p < .01$]. No other effects were significant.

One-way repeated measures ANOVAs on the same-talker data were performed for direct comparisons of the two experimental conditions, each of the experimental conditions with their respective control conditions, and the two control conditions. Crucially, there was a significant difference between the responses of those who listened to natural [f]-final words and ambiguous [s]-final words, and the responses of those who listened to natural [s]-final words and ambiguous [f]-final words [$F(1,27) = 13.81, p < .01$]. The comparison between the training condition group that listened to natural [f]-final words and ambiguous [s]-final words and its control group (natural [f]-final words and [?]-final nonwords) was significant [$F(1,28) = 4.97, p < .05$], whereas the difference between the training condition group that listened to natural [s]-final words and ambiguous [f]-final words and its control group (natural [s]-final words and [?]-final nonwords) was not significant [$F(1,29) = 4.06, p < .1$]. Importantly, there was no significant difference between the two control groups.

For the different-talker groups, categorization of the continuum steps shifted globally toward [s]. Orthogonal to this shift, there was a mean difference of 22% between the two groups. Data from these two groups were analyzed together with the data of the two same-talker groups that had received the same exposure conditions. We carried out a repeated measures ANOVA on the percentage of [f] responses with step as a within-subjects factor and fricative type (whether listeners heard natural [f] or [s] words at exposure) and talker change (whether or not there was a talker change in the test phase) as between-subjects factors. There was a significant effect of step [$F(4,220) = 27.80, p < .001$] and significant main effects of fricative type [$F(1,55) = 20.10, p < .01$] and talker change [$F(1,55) = 25.67, p < .01$]. Crucially, there was no interaction between these two factors [$F(1,55) = 1.86, p > .05$], suggesting that the size of the difference between the training groups did not differ as a function of whether or not there was a talker change. This was confirmed in a planned comparison of the two different-talker groups, which showed a significant difference [$F(1,28) = 6.26, p < .05$].

Questionnaire

Same-talker groups. In the experimental condition with natural [f]-final and ambiguous [s]-final words, 9 participants (64%) reported that they had heard unusual [s] sounds in some items. Typical comments were that the words had not been articulated properly or that the talker spoke "with a lisp." To the question of how this influenced their responses in the lexical decision task, the most common reply was that they were sometimes in doubt about whether or not to label these items as words (4 out of the 9). Two participants replied that they became more alert and listened more carefully when they noticed the unusual sounds. The remaining 3 did not think that their lexical decision responses were influenced by the unusual fricatives. None of the 9 participants reported being influenced by the unusual exposure sounds in their categorization responses.

Only 1 participant from the other three conditions remarked on unusual fricatives—namely, that there were "English *th*-sounds" in some of the items. This participant was in the control condition group that listened to natural [f]-final words and [?]-final nonwords and did not report being influenced by the presence of these sounds in either of the two parts of the experiment.

Different-talker groups. Three participants replied spontaneously that there were different voices in the ex-

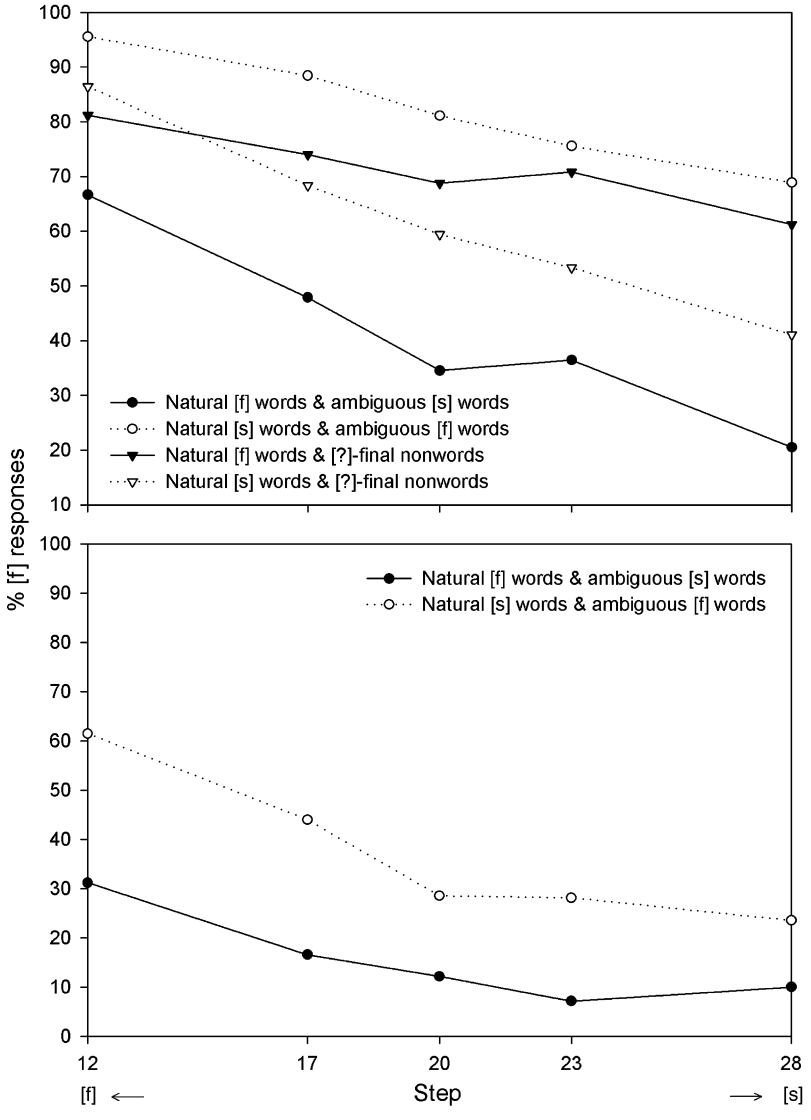

Figure 1. Experiment 1: Mean percentages of [f] responses in the six conditions plotted against each of the five continuum steps. Upper panel: Experimental and control exposure conditions with Talker 1's speech presented during exposure and categorization. Lower panel: Experimental exposure conditions with Talker 1's speech presented during exposure and Talker 2's vowel with Talker 1's fricatives during categorization.

posure and test phases. Of the remaining 27, 18 replied "different" when asked explicitly whether the voice was the same or different in the two parts, 7 replied "same," and 2 replied "don't know." Overall then, 70% of the participants who were included in the final analysis said that there was a different talker, either spontaneously or when asked explicitly.

Discussion

The perceptual learning effect reported by Norris et al. (2003) was replicated and found to persist when ambiguous fricatives, made from natural productions by the exposure talker, were presented to listeners in the context of a vowel from a novel talker.

In the exposure phase, the overall performance of the two pairs of experimental groups was very similar. Listeners labeled ambiguous versions of the experimental items as existing Dutch words most of the time. However, although the ambiguous fricative [?] was categorized equally often as [f] or [s] by the pretest listeners, participants in the main experiment seemed to treat this sound more often as [f]. This [f]-bias was also observed by Norris et al. (2003) and was reflected in a higher percentage of "no" responses to [s]-final items. This asymmetry may be explained by the constant—and therefore uninformative—vocalic context in the pretest that encouraged listeners to ignore any coarticulatory cues in the vowel, whereas in the exposure phase the ambiguous

fricatives occurred in variable vocalic contexts that apparently cued [f] more reliably than [s] (see Norris et al., 2003, for a more detailed discussion).

The main finding of the categorization phase with the same-talker items was a replication of the perceptual learning effect reported by Norris et al. (2003). Listeners who had heard the ambiguous sound [?] in [s]-biased lexical contexts categorized the fricative continuum mostly as [s], while the group that had heard this sound in [f]-biased contexts categorized the same continuum largely as [f]. The control groups, which had been exposed to the same distribution of critical phonemes devoid of lexical context, gave intermediate responses and, as in the Norris et al. study, did not differ from each other. This suggests that, in accordance with Hervais-Adelman et al.'s (2002) findings on noise-vocoded speech, the observed effect arises as a consequence of lexical feedback and cannot be explained by a phonetic contrast effect (i.e., listeners do not appear to be able to learn, on the basis of contrast alone, that since they hear, e.g., an unambiguous [f] during the exposure phase, the ambiguous sound must be an [s]). This lexical influence is related to the Ganong effect (Ganong, 1980)—the tendency of listeners to label ambiguous sounds (including word-final fricatives; McQueen, 1991) in a lexically consistent way. As discussed extensively by Norris et al., however, the present lexical effect differs from the Ganong effect in one crucial way: It reflects a lexical influence on perceptual learning, rather than a direct influence on explicit phonemic decision making.

There were two main findings from the conditions in which, during the categorization phase, listeners heard syllables in which the vowel came from a different talker. First, categorization of the continuum shifted toward the [s] endpoint for both groups. This global effect is most likely a consequence of the acoustic properties of the vowel (Johnson, 1991; Mann & Repp, 1980; Mann & Soli, 1991). For instance, one explanation of this shift is that the lower pitch in the vowel (197 Hz for Talker 2 compared with 242 Hz for Talker 1) and/or the lower spectral center of gravity of the vowel (651 Hz for Talker 2 compared with 738 Hz for Talker 1) led listeners to expect a concentration of energy for [f] to occur in a lower frequency region. Since most of the fricatives had energy peaks that were, with respect to the preceding vowel, relatively high in frequency, these sounds were categorized largely as [s]. Borrowing from the literature on vowel normalization, listeners could be said to use *extrinsic* (Johnson, 1990; Nearey, 1989) information to adjust their interpretation of linguistic cues in the fricative.

Second, and more important, there was again a perceptual learning effect: The listeners who had heard the ambiguous fricative in [s]-biased contexts gave more [s] responses than did the other group. This lexically biased learning effect was orthogonal to the global [s]-bias. There are at least three interpretations of the learning effect. One obvious possibility is that this kind of perceptual learning generalizes to another talker. The listeners in the exposure phase made an adjustment to the [f]–[s] category boundary, and this adjustment affected processing of subsequently encountered speech, regardless of talker. An alternative explanation is that the effect persists in the different-talker conditions because the fricatives that were used here were still produced by the talker of the exposure phase. It is plausible that these stimuli were recognized by the perceptual system as being produced by the exposure talker and consequently treated as such, even though the preceding vowel indicated that the syllables were produced by a different talker. On this account, the perceptual system analyzes the incoming signal for talker identity and applies previously stored information about the talker on a phoneme-by-phoneme basis. A third account is that using a vowel from a talker of the same gender and similar age did not contain enough information for the perceptual system to treat the utterance as coming from a new talker. For example, Nusbaum and Morin (1992, Experiment 4) found evidence that the speech perception system does not necessarily carry out a new adjustment computation for a new talker if the voice of the talker is acoustically similar enough to that of the previous talker. Although the majority of participants (70%) indicated hearing a talker change, it is not clear whether this change was processed as such online. Furthermore, only 11% spontaneously pointed out a talker change when they were questioned. When the remaining listeners were asked the question explicitly, very few were confident in their replies. This account was tested in Experiment 2: If the persistent difference in categorization responses between exposure groups observed here is due to too small an acoustic difference between Talker 1 and Talker 2, using a more extreme contrast should eliminate the effect.

EXPERIMENT 2

The aim of this experiment was to test whether the perceptual learning effect that was found for the different-talker groups in Experiment 1 was due to insufficient contrast between the voices of the exposure talker and the test talker. We thus repeated the different-talker conditions of Experiment 1, but this time the test items were presented in the context of a vowel from a *male* talker.

Method

Participants

Thirty-two volunteers from the MPI for Psycholinguistics participant pool were assigned to two training conditions. None had taken part in Experiment 1, and none reported any hearing disorders. All were paid for their participation.

Materials, Stimulus Construction, and Procedure

Lexical decision. Materials in the exposure phase were the same as those used for the experimental groups in Experiment 1.

Phonetic categorization. A new set of materials was made for the categorization task in the same way as for the different-talker items in Experiment 1, but this time from recordings of a male native speaker of Dutch (Talker 3). Recording and digitization procedures were the same as for Talker 2 in Experiment 1. The vowel selected for splicing onto the five fricative steps was 152 msec in duration and equated in intensity to the vowels in the categorization

phase of Experiment 1. As before, this [ɛ] was excised from a token of the syllable [ɛx]—that is, from a velar fricative context.

Questionnaire. The participants were given the same questionnaire as the different-talker exposure groups in Experiment 1.

Procedure. The procedures were identical to those used in Experiment 1.

Results

Lexical Decision

We used the same criterion for exclusion of participants as in the previous experiment, which meant that the data from 2 participants from the group that listened to ambiguous [s]-final words and from 1 participant from the group that listened to ambiguous [f]-final words were not analyzed. The lexical decision data generally show the same pattern as in the previous experiment (see Table 1), albeit with some variability across participants' RTs. Seven mono- or bisyllabic ambiguous [s]-final words were labeled as nonwords by more than 40% of the listeners. In the RT data, no effects were significant. In the percentages of "no" responses, there were significant main effects of both final fricative [$F_1(1,29) = 37.05, p < .001$; $F_2(1,38) = 8.45, p < .01$] and exposure group [$F_1(1,29) = 24.29, p < .001$; $F_2(1,38) = 15.55, p < .001$] and a significant interaction of the two factors [$F_1(1,29) = 30.49, p < .001$; $F_2(1,38) = 14.05, p < .005$].

Phonetic Categorization

Mean percentages of [f]-responses are given in Figure 2. The average difference in responses between the two exposure groups was 25%, in the same direction as in Experiment 1. These data were analyzed again together with the two same-talker experimental exposure groups in Experiment 1. There was a significant effect of step [$F(4,224) = 32.17, p < .001$] and an interaction of step and fricative type [$F(4,224) = 2.99, p < .05$]. There was also a significant main effect of fricative type [$F(1,56) = 20.03, p < .001$] but, importantly, no interaction of fricative type and talker change: The difference between the two training groups was not modulated by a talker change. Furthermore, a pairwise comparison of the two training groups from Experiment 2 showed a significant difference [$F(1,29) = 6.45, p < .05$].

Questionnaire

None of the participants spontaneously stated that there was a different voice in the two parts of the experiment; when asked directly, however, if the talker in the two parts was the same or different, listeners unanimously replied "different."

Discussion

As in Experiment 1, listeners appear to apply a previously learned category boundary shift to fricatives that are presented in the context of a vowel from a novel talker. Unlike in Experiment 1, however, there was no main effect of a change in talker; that is, there was no global shift in categorization responses as a result of a context vowel with different acoustic properties from those of the vowel of the exposure talker. We suggested earlier that this effect occurred in Experiment 1 because the lower centroid and *f*0 of Talker 2's vowel might have caused a bias to expect the [f]–[s] boundary to be in a lower frequency range, too, consequently leading listeners to categorize the continuum largely as (high-frequency) [s]. Following this ar-

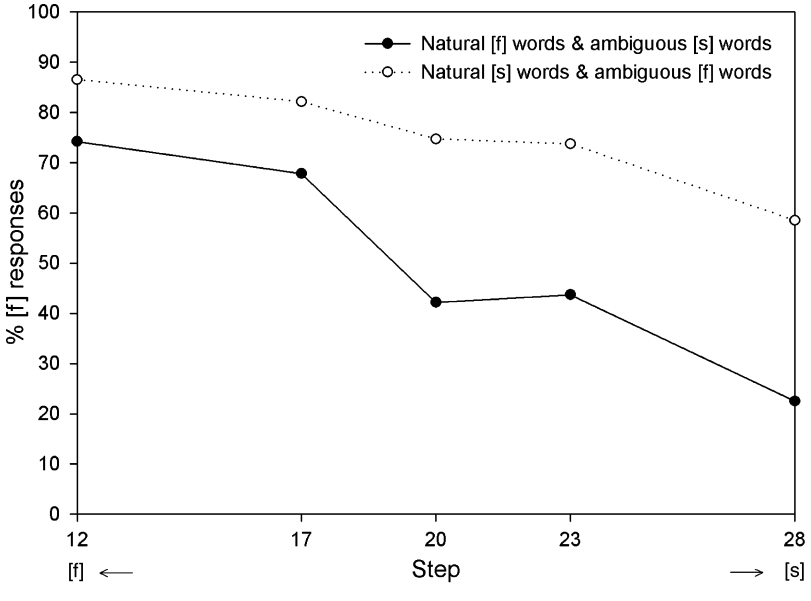

Figure 2. Experiment 2: Mean percentages of [f] responses of the two exposure groups to each of the five continuum steps: Talker 1's speech presented during exposure and Talker 3's vowel with Talker 1's fricatives during categorization.

gument, however, a similar shift would be expected in the case of the male vowel produced by Talker 3 in the present experiment since it was even lower in spectral center of gravity and *f*0 (620 Hz and 111 Hz, respectively) than the female vowel produced by Talker 2. Conceivably, there was a more powerful gender normalization process at work, which overrode an effect of the kind observed in Experiment 1.

Importantly, we again found an effect of previous exposure even though in this experiment the ambiguous fricatives were presented in the context of a vowel from a male talker, which was acoustically clearly different from the exposure talker's vowels. Accordingly, all of our participants reported the percept of a male talker during the test phase. Hence, the interpretation of the results of Experiment 1 in terms of insufficient difference between the two talkers used at exposure and test can be dismissed, and we are left with two possible accounts of the present results—namely, that the perceptual learning examined here is applied to different talkers, or, alternatively, that the perceptual system "recognized" the fricative sounds in the test phase as coming from the talker of the exposure phase in spite of the different-talker vowel context, and consequently applied the previously acquired modulation of the [f]–[s] category boundary.

EXPERIMENT 3

In Experiment 3, these two accounts were tested. We used the same experimental exposure conditions as in the previous experiments but presented listeners with test stimuli in which both the vowel and the ambiguous fricatives came from an unfamiliar talker. If the first account is correct and learning generalizes, we would expect a difference in the categorization responses of the two exposure groups. If, however, learning is talker specific, there should be no effect of exposure in the categorization of fricative sounds from the novel talker.

Method

Participants

Fifty-eight members of the MPI for Psycholinguistics participant pool, none of whom had participated in Experiment 1 or 2, were tested. None of them reported hearing disorders, and all were paid for their participation. Forty-eight took part in the main experiment and 10 in a pretest. More participants were tested in the main part than in the previous experiments in order to increase statistical power. Power analysis (Cohen, 1988) after we tested 16 participants in each group suggested that power was lower by an order of magnitude compared with the same-talker groups in Experiment 1 (0.086 here and 0.842 in Experiment 1), due both to decreased effect size and to increased interparticipant variability in Experiment 3.

Pretest

A pretest was conducted in order to establish five steps on a new [f]–[s] continuum that match the acoustical properties and ambiguity of the stimuli used in Experiments 1 and 2.

Stimulus construction. An [ɛf]–[ɛs] continuum based entirely on Talker 3's speech was created using the technique described in Experiment 1. The [f] and [s] endpoints were recorded by Talker 3 in the same recording session as the vowel [ɛ] (which was also the token used in Experiment 2) and redigitized in the same way. The fricative steps on this new 21-step continuum were matched in duration and intensity to the continuum used in Experiments 1 and 2.

Procedure. Informal listening suggested that the most ambiguous range of the continuum was Steps 9–18. These 10 steps were thus presented to listeners using the same procedure as in the pretest of Experiment 1.

Results. Percentages of [f] responses were again averaged for each step. Five steps for the main experiment were selected to match the fricatives used in the previous experiments as closely as possible. Since the fricatives in those experiments had average percentages of [f] responses of 85, 70, 50, 30, and 15, the steps that corresponded most closely to these percentages were also selected here. To this end, it was again necessary to create a more fine-grained 41-step continuum. The five steps that were used for the main experiment, then, were 24, 26, 28, 30, and 32.

Materials and Procedure

The stimuli for the exposure phase were those that were used in the two previous experiments. In the categorization part, the five [ɛf]–[ɛs] steps that were established in the pretest were used. The procedures for the lexical decision and the categorization tasks were as in Experiments 1 and 2, except that the participants did not fill in a questionnaire.

Results

Lexical Decision

Application of the 50% cutoff point on lexical decision performance led to exclusion of 2 participants from the group that listened to natural [f]-final and ambiguous [s]-final words at exposure, leaving 22 participants in that group and 24 in the other. Overall, the lexical decision data followed the same pattern as in the previous two experiments (see Table 1), again with some variability in the RTs. Five ambiguous [s]-final words and one natural [s]-final word were labeled as nonwords by more than 40% of the listeners. In the RT data, there was a significant interaction of final fricative and exposure group [$F_1(1,44) = 6.17, p < .05$; $F_2(1,38) = 14.28, p < .005$]. None of the main effects were significant. In the percentages of "no" responses, there were significant main effects of both final fricative [$F_1(1,44) = 61.07$, $p < .001$; $F_2(1,38) = 6.84, p < .05$] and exposure group [$F_1(1,44) = 10.59, p < .005$; $F_2(1,38) = 5.36, p < .05$] and a significant interaction between the two factors [$F_1(1,44) = 20.89, p < .001$; $F_2(1,38) = 9.67, p < .005$].

Phonetic Categorization

The mean percentages of [f] responses to the five fricative sounds are plotted in Figure 3. There is a small mean difference of 7% between the exposure groups, going in the same direction as in previous experiments. We again conducted a 2 × 2 ANOVA to compare this effect with the categorization data of the same-talker experimental exposure conditions in Experiment 1.

There were significant effects of step [$F(4,284) = 30.88, p < .001$] and fricative type [$F(1,71) = 8.79, p < .005$] but not of talker change. Crucially, there was a significant interaction of the two latter factors [$F(1,71) = 4.17, p < .05$]; that is, there was a difference in the mag-

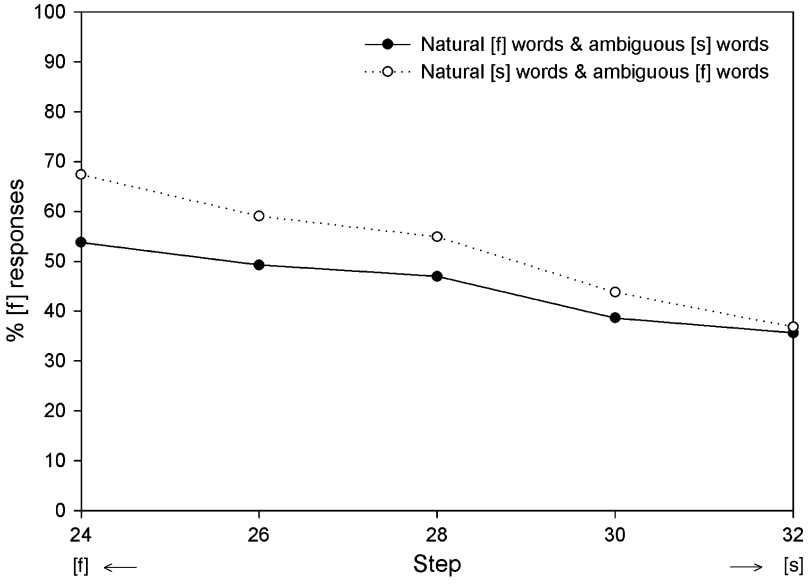

Figure 3. Experiment 3: Mean percentages of [f] responses of the two exposure groups to each of the five continuum steps: Talker 1's speech presented during exposure and Talker 3's vowel and fricatives during categorization.

nitude of the perceptual learning effect between Experiments 1 and 3. We then conducted a planned comparison of the two exposure groups of Experiment 3 only. There was no significant effect of fricative type ($p = .49$); that is, there was no effect of exposure on categorization responses in this experiment.

Discussion

Unlike in Experiments 1 and 2, where listeners were presented with ambiguous fricatives produced by the talker that they had heard in the exposure phase, here we found no effect of exposure when listeners were tested on fricative sounds produced by a novel talker. This result suggests that adjustments to atypical speech are reapplied in a talker-specific manner and do not generalize to processing of utterances from other talkers. Furthermore, the presence of the effect in Experiments 1 and 2 suggests that adjustments affect a specific phonetic contrast, and they are reapplied regardless of the context in which the test sounds appear. Since this conclusion is based on a null effect in Experiment 3, however, it was followed up in Experiment 4.

EXPERIMENT 4

The aim of Experiment 4 was to show that perceptual learning, under appropriate exposure conditions, can be applied to the fricative continuum of Talker 3 that was used in the previous experiment. At the same time, we wanted to have another test of the specificity of perceptual learning of a particular phonetic contrast. Would perceptual learning about Talker 3's fricatives occur in the context of words produced by Talker 1?

We therefore used speech editing to splice an ambiguous fricative [?], based on Talker 3's speech, and unambiguous tokens of [f] and [s], into the critical fricative-final materials from the exposure phase, as spoken by Talker 1. Thus, in one version of the materials, the [?] in the [f]-final materials (e.g., *olij?*) spoken by Talker 1 was replaced with the sound used as the most ambiguous step in the Experiment 3 test continuum, based on Talker 3's speech, and the [s] in the [s]-final words (e.g., *radijs*) was a natural [s] spoken by Talker 3. In the other version of the materials, the [f] in the [f]-final words came from Talker 3, and the ambiguous sound in the [s]-final words was based on his speech.

If an adjustment of the [f]–[s] boundary on Talker 3's fricatives can be induced by this situation, the null effect in Experiment 3 can be attributed to talker specificity. Furthermore, if learning about Talker 3's fricatives can be induced in the context of speech produced by another talker, this would provide additional support for the account that the perceptual learning mechanism operates regardless of context and can affect a specific phonetic contrast.

Method

Participants

Thirty-nine members of the MPI for Psycholinguistics participant pool took part. None reported hearing disorders, none had participated in the previous experiments, and all were paid to participate. There were 19 participants in the group receiving [f]-biased exposure and 20 in the group receiving [s]-biased exposure.

Materials and Procedure

Lexical decision. Again, two lexically biased exposure conditions were identical to those of Experiment 3 in all respects, except

for the two following manipulations: The critical ambiguous fricative [?] used for the creation of ambiguous [f]- and [s]-final words was now taken from the Talker 3 continuum (specifically, the fricative sound that had been established as the most ambiguous sound in the pretest of Experiment 3; Step 28). Unlike in previous experiments, the natural [f]- and [s]-final words were spliced as well. For these items, the final fricatives were excised at zero-crossings and replaced by the appropriate natural endpoint of Talker 3's continuum (Step 1 for [f] and Step 41 for [s]).

Phonetic categorization. Procedures and stimuli for the test phase were identical to those of Experiment 3.

Results

Lexical Decision

On the basis of the exclusion criterion used in previous experiments, 4 participants from the group that listened to natural [f]-final and ambiguous [s]-final words at exposure were not entered into further data analyses. The lexical decision data show that participants tended to respond more slowly than in the previous experiments and to label more experimental items as nonwords (see Table 1). Five ambiguous [s]-final words and two natural [s]-final words were labeled as nonwords by more than 40% of listeners. On average, however, 93% of the natural versions and 84% of the ambiguous versions were accepted as words although they had been constructed by concatenating speech from different talkers.

In the percentages of "no" responses, there was a significant main effect of final fricative [$F_1(1,33) = 64.56$, $p < .001$; $F_2(1,38) = 6.33$, $p < .05$]: [s]-final items were labeled as nonwords more often than were [f]-final items. There was no significant main effect of exposure group and no significant interaction. In the RT data, neither of the main effects, nor the interaction, was significant.

Phonetic Categorization

The results of the test phase showed a mean difference of 39% in the percentages of [f] responses between the two exposure groups (see Figure 4). As in Experiments 1 and 2, the pattern was such that the group that had listened to [?] in [f]-final words gave more [f] responses to the test stimuli.

We first compared this bias effect with the effect in the same-talker conditions of Experiment 1 in a 2 (type of natural fricative at exposure) × 2 (experiment) ANOVA. There was a significant effect of step [$F(4,240) = 30.92$, $p < .001$] and of fricative type [$F(1,60) = 29.07$, $p < .001$] but no main effect of experiment. Importantly, there was no interaction between these two factors: The bias effects in the present experiment and in the same-talker conditions of Experiment 1 are of similar magnitude.

Second, we repeated this ANOVA with the categorization data from Experiment 3. Again, the only significant main effects were step [$F(4,308) = 36.51$, $p < .001$] and fricative type [$F(1,77) = 9.30$, $p < .005$]. Crucially, however, the interaction between these factors was significant [$F(1,77) = 4.25$, $p < .05$]. This interaction was then followed up with a pairwise comparison of only the two exposure groups in Experiment 4 [$F(1,33) = 15.38$, $p < .001$].

Discussion

The listeners in this experiment applied an adjustment to Talker 3's fricatives that was learned when an ambiguous fricative produced by Talker 3 was placed in the context of words produced by Talker 1. The learning effect here was statistically indistinguishable from the one in the same-talker conditions in Experiment 1, but dif-

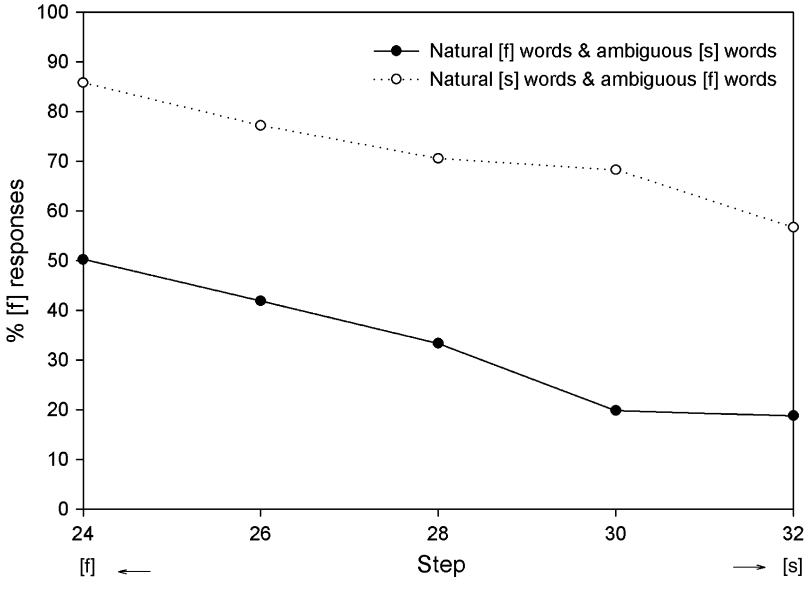

Figure 4. Experiment 4: Mean percentages of [f] responses of the two exposure groups to each of the five continuum steps: Talker 3's fricatives in Talker 1's speech presented during exposure and Talker 3's vowel and fricatives during categorization.

ferent from that in Experiment 3, where the fricative sounds at exposure and test came from a different talker. We can therefore conclude that the null effect in Experiment 3 was a consequence of the experimental setup and that it occurred because the perceptual adjustment investigated here does not generalize across talkers.

GENERAL DISCUSSION

The results of this perceptual learning study show that an adjustment made by the perceptual system in response to one talker's unusual production of speech sounds is stored and reapplied to speech of the same talker, but does not affect processing of speech from other talkers.

Perceptual learning after exposure to an ambiguous fricative sound [?] was evident when this and other sounds on an [f]–[s] continuum were presented in the context of a vowel [ɛ] produced by the talker about whose speech learning had occurred, as well as in the context of vowels produced by other talkers. When presented with test syllables made with vowels from other talkers, listeners perceived a talker change, but the fricatives were treated in a similar way to when they appeared in syllables made entirely from the speech of the exposure talker. With an [ɛf]–[ɛs] test continuum made entirely from utterances of a novel talker, however, we found no evidence of application of previous learning, unless the fricative sounds learned during the exposure phase had themselves originated from the test talker (i.e., the test talker was in fact not entirely new to the listeners).

The perceptual learning effect clearly seems to be lexically mediated (Norris et al., 2003). Evidence for this conclusion comes from two control conditions, in which listeners received the same distribution of critical sounds as did the experimental listeners, but in which, unlike in the experimental conditions, ambiguous sounds did not occur in lexical contexts. Since listeners in these control conditions did not show evidence of a category boundary shift, the difference in categorization responses in the experimental conditions cannot simply be a contrast effect. Rather, the modulation of the category boundary appears to be the result of a feedback signal from the lexicon. When an incoming ambiguous sound can be disambiguated by lexical information, feedback from the lexicon to a prelexical level results in an adjustment of the phonetic category boundary that can, in turn, affect perception of future instances of similar ambiguous sounds.

From the perspective of talker normalization, this effect is in line with previous research on normalization of an individual's speech (Nygaard & Pisoni, 1998; Nygaard et al., 1994). Listeners make adjustments to idiosyncratic speech production, and the outcome of these computations appears to be stored for later use (Mullennix et al., 1989; Nusbaum & Morin, 1992). One question that was examined here was whether this kind of learning may affect the processing of, or be misapplied to, the speech of other talkers. Given that in Experiment 3 we found no effect of exposure on categorization of ambiguous syllables produced by a novel talker, the answer to this is negative. However, this may turn out to be true only under single-talker conditions. Bradlow and Bent (2003) have shown that listeners are able to apply the outcome of a perceptual adjustment to a novel talker when multiple talkers at exposure share the same idiosyncrasy (in their case, Chinese-accented English). Furthermore, Lively, Logan, and Pisoni (1993) found that talker variability plays an important role in the acquisition of a new phonetic contrast, rather than modification of an existing one. Taken together, these two studies suggest that talker variability facilitates the development and modification of abstract representations of speech. When there are multiple talkers at exposure, the system may be better able to discern acoustic patterns that talkers have in common from those that are idiosyncratic. In the case of single-talker exposure, however, it is less clear which properties of the input signal are characteristic of a phonetic contrast and which are characteristic of the individual talker's vocal tract shape or articulatory habits. The perceptual system would thus be well-advised not to generalize learning to other voices too readily, because such adjustments do not necessarily benefit the processing of other talkers' speech.

Talker specificity in application of perceptual learning is in accord with the results of Mullennix and Pisoni (1990) and Green et al. (1997), who found, at a phonemic level, evidence for a processing dependency between voice information and linguistic information in which linguistic processing is contingent on voice processing. In the present experiments, we found evidence of such a processing dependency in the application of a previously learned category boundary modulation. More specifically, our results suggest that application of learned adjustments to a talker is mandatory when that talker's voice is encountered again (i.e., even when that talker's speech sounds occur in the context of another talker's vowels).

A second question we asked concerned the phonetic specificity of perceptual adjustments. The mechanism by which this learning is applied to the incoming speech signal appears to be remarkably sensitive and robust. Given the null effect in Experiment 3, the effect in the talker-change conditions of Experiments 1 and 2 can only be due to the fact that in these experiments, fricatives based on the exposure talker's speech were presented. Whereas the syllables as a whole were perceived as coming from a novel talker, the perceptual mechanism that reapplies stored adjustments appears to operate on a subsyllabic level, irrespective of context. The speech signal thus appears to be monitored continuously for talker identity and for potentially useful information about talkers with a resolution at least at the level of individual segments. Additional evidence for a segmental locus of the learning effect was found in Experiment 4. In this experiment, a modulation of the [f]–[s] category boundary was made in response to fricatives that were based on the test talkers' speech but had been spliced into the exposure talkers' utterances; there was simply no other infor-

mation available about the test talker during exposure, apart from his [f]–[s] productions.

The present findings thus suggest that the perceptual learning mechanism investigated here affects representations of fricative sounds at a segmental, prelexical level. Further evidence that these adjustments are prelexical comes from a related cross-modal priming experiment (McQueen, Cutler, & Norris, 2003), which used exposure conditions similar to the present experiments. Listeners in that study showed identity priming effects for ambiguous items as a function of exposure condition (ambiguous items such as [doː?] primed either *doof*, "deaf," or *doos*, "box"). An adjustment made at a prelexical stage of processing therefore appears to have biased the interpretation of subsequently heard ambiguous sounds, which in turn affected activation of words that had not been heard at exposure.

No current model of word recognition can accommodate perceptual learning at a segmental level. Models that have units of perception only at the lexical level (Klatt, 1979, 1989) can explain adjustments to individual talkers but not specificity of these adjustments at the segmental level. Other models (e.g., McClelland & Elman, 1986; Norris, 1994; Stevens, 2002) that propose abstract phonetic categories prior to lexical access, on the other hand, do not yet have a mechanism of handling talker (or any other kind of) variability in the process of mapping the incoming speech signal to these categories. They can, however, be supplemented by models specifically aimed at handling variability in the input (e.g., Johnson, 1997; Kruschke, 1992; Nearey, 1989; Smits, 2001a, 2001b). This will hopefully lead to models of word recognition with increasingly fine-grained and dynamic input representations.

With respect to the relation of talker identity information and phoneme recognition, our results support a processing model in which linguistic and talker identity information are processed in parallel, and in which talker identity information constrains the interpretation of linguistic cues (cf. models of vowel normalization, e.g., Hirahara & Kato, 1992). Talker identity information, in this sense, comprises what is intrinsic in the signal and processed on line, as well as previously acquired and stored information. According to this view of talker normalization, the perceptual system achieves perceptual constancy by exploiting the sources of variability in the speech signal to constrain the interpretation of that inherently ambiguous signal. We therefore do not endorse talker normalization in its narrow sense as a process in which any indexical information is stripped off the signal prior to access to linguistic units of representation (see, Pisoni, 1997, for discussion).

The results from these experiments extend previous research that has shown that listeners adjust individual phoneme boundaries in response to unusual speech (Ladefoged, 1989; Norris et al., 2003; Scott & Cutler, 1984) and that listeners make talker-specific adjustments (Mullennix et al., 1989; Nusbaum & Morin, 1992; Nygaard et al., 1994) by showing that perceptual adjustments to speech can be highly specific. These adjustments appear to be specific both with respect to segmental information (the adjustments can be specific to a single phonetic contrast) and with respect to information about talker identity (the adjustments can be about one particular talker). We have argued that these findings can best be explained in a model of speech processing in which fricative information is represented at a prelexical stage. These prelexical representations are then modulated by feedback from the lexicon in a talker-specific manner.

REFERENCES

Bradlow, A. R., & Bent, T. (2003, August). Listener adaptation to foreign-accented speech. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 2881-2884). Barcelona.

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 521-533.

Clarke, C. M. (2002, September). Perceptual adjustment to foreign-accented English with short-term exposure. In *Proceedings of the 7th International Conference on Spoken Language Processing* (pp. 253-256). Denver.

Clarke, C. M. (2003). *Processing time effects of short-term exposure to foreign-accented English*. Unpublished doctoral dissertation, University of Arizona, Tucson.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception & Performance*, **23**, 914-927.

Evans, B. G., & Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America*, **115**, 352-361.

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

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 1166-1183.

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

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 152-162.

Green, K. P., Tomiak, G. R., & Kuhl, P. K. (1997). The encoding of rate and talker information during phonetic perception. *Perception & Psychophysics*, **59**, 675-692.

Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (1988). Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 421-433.

Hervais-Adelman, A., Johnsrude, I. S., Davis, M. H., & Brent, L. (2002, September). *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.

Hirahara, T., & Kato, H. (1992). The effect of *F*0 on vowel identification. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic structure* (pp. 89-112). Tokyo: Ohmsha.

Johnson, K. (1990). The role of perceived speaker identity in *F*0 normalization of vowels. *Journal of the Acoustical Society of America*, **88**, 642-654.

Johnson, K. (1991). Differential effects of speaker and vowel variability on fricative perception. *Language & Speech*, **34**, 265-279.

Johnson, K. (1997). 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: Academic Press.

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

Klatt, D. H. (1986). The problem of variability in speech recognition and in models of speech perception. In J. S. Perkell & D. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 301-324). Hillsdale, NJ: Erlbaum.

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.

Knösche, T. R., Lattner, S., Maess, B., Schauer, M., & Friederici, A. D. (2002). Early parallel processing of auditory word and voice information. *NeuroImage*, **17**, 1493-1503.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.

Ladefoged, P. (1989). A note on "information conveyed by vowels." *Journal of the Acoustical Society of America*, **85**, 2223-2224.

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

Lattner, S. (2002). *Neurophysiologische Untersuchungen zur auditorischen Verarbeitung von Stimminformation* [Neurophysiological investigations into auditory processing of voice information] (Max Planck Series in Cognitive Neuroscience, Vol. 29). Leipzig: Max-Planck-Institut fur Neuropsychologische Forschung.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, **94**, 1242-1255.

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.], & Soli, S. D. (1991). Perceptual order and the effect of vocalic context on fricative perception. *Perception & Psychophysics*, **49**, 399-411.

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 676-684.

Maye, J., Aslin, R., & Tanenhaus, M. (2003, March). In search of the weckud wetch: Online adaptation to speaker accent. In *Proceedings of the 16th Annual CUNY Conference on Human Sentence Processing* (pp. 153). Cambridge, MA.

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

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

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

Mehler, J., Sebastian, N., Altmann, G., Dupoux, E., Christophe, A., & Pallier, C. (1993). Understanding compressed sentences: The role of rhythm and meaning. In P. Tallal, A. M. Galaburda, R. R. Llinás, & C. von Euler (Eds.), *Temporal information processing in the nervous system: Special reference to dyslexia and dysphasia* (Annals of the New York Academy of Sciences, Vol. 682, pp. 272-282). New York: New York Academy of Sciences.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, **47**, 379-390.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, **85**, 365-378.

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

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

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

Nusbaum, H. [C.], & Magnuson, J. (1997). Talker normalization: Phonetic constancy as a cognitive process. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 109-129). San Diego: Academic Press.

Nusbaum, H. C., & Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic structure* (pp. 113-134). Tokyo: Ohmsha.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, **60**, 355-376.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, **5**, 42-46.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 309-328.

Pisoni, D. B. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech Communication*, **13**, 109-125.

Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9-30). San Diego: Academic Press.

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

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

Smits, R. (2001a). Evidence for hierarchical categorization of coarticulated phonemes. *Journal of Experimental Psychology: Human Perception & Performance*, **27**, 1145-1162.

Smits, R. (2001b). Hierarchical categorization of coarticulated phonemes: A theoretical analysis. *Perception & Psychophysics*, **63**, 1109-1139.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, **111**, 1872-1891.

(Manuscript received October 7, 2003;
revision accepted for publication May 7, 2004.)