Journal of Experimental Psychology:
Human Perception and Performance
2010, Vol. 36, No. 1, 195–211

0096-1523/10/$12.00 DOI: 10.1037/a0016803

The Bounds on Flexibility in Speech Perception

Matthias J. Sjerps
Max Planck Institute for Psycholinguistics

James M. McQueen
Max Planck Institute for Psycholinguistics
and Radboud University Nijmegen

Dutch listeners were exposed to the English theta sound (as in *bath*), which replaced [f] in /f/-final Dutch words or, for another group, [s] in /s/-final words. A subsequent identity-priming task showed that participants had learned to interpret theta as, respectively, /f/ or /s/. Priming effects were equally strong when the exposure sound was an ambiguous [fs]-mixture and when primes contained unambiguous fricatives. When the exposure sound was signal-correlated noise, listeners interpreted it as the spectrally similar /f/, irrespective of lexical bias during exposure. Perceptual learning about speech is thus constrained by spectral similarity between the input and established phonological categories, but within those limits, adjustments are thorough enough that even nonnative sounds can be treated fully as native sounds.

Keywords: speech perception, perceptual learning, bilingualism

Speech perception is very flexible. There are, for example, rapid perceptual adjustments in response to a speech sound that has been realized unusually (Norris, McQueen, & Cutler, 2003). This flexibility is of great value in speech comprehension in that it allows the listener to tune in to the ways talkers speak and, hence, to understand them better. In the current study, we set out to establish the bounds on the flexibility of speech perception. Specifically, we asked what the limits are on the adjustments that listeners can make to unusual sounds. Can all sounds be adjusted to, and, if not, what properties determine which adjustments can occur? We present four experiments that examined the flexibility of first-language speech perception. The results provide a picture of how perceptual flexibility can aid the listener in daily life, and, importantly, the restrictions that apply to this adaptive process.

The experiments were based on the perceptual-learning paradigm developed by Norris et al. (2003). In that study, Dutch listeners heard an ambiguous sound midway between [f] and [s], presented as the final fricative in words in either of two exposure conditions. One group of listeners heard the ambiguous [fs] sound in 20 words that normally end in [f] (e.g., *druif*, “grape”), mixed with filler words that contained no other [f]s. A second group heard the same sound in 20 words that normally end in [s] (e.g., *moeras*, “swamp”). In a subsequent phonetic categorization test, the first group categorized more sounds as /f/ on an [ɛf]–[ɛs] continuum than the second group. Additional groups who heard the ambiguous sounds in nonword contexts did not show a shift in categorization. The effect therefore appears to be due to lexical feedback. Listeners thus mold their phonemic categories to reflect the phonetic material with which they are currently confronted, using the mental lexicon to infer what the necessary adjustments are.

Norris et al. (2003) argued that this kind of adjustment, which they termed *lexically guided perceptual learning*, is used by listeners in everyday life to adapt to speakers who speak in an unusual way because of a foreign or regional accent, for example, or because of an idiosyncratic way of pronunciation. Bradlow and Bent (2008) have indeed argued that lexically guided perceptual learning supports improvements in listeners’ ability to understand speakers with a foreign accent. McQueen and Mitterer (2007) have argued that the same learning mechanism is used when listeners adjust to dialect-induced pronunciation variation, specifically in the situation where listeners watching a film use subtitles to help them adjust to an unfamiliar regional accent in a foreign language. Lexical retuning in response to talker-specific pronunciation variation has also been documented (Eisner & McQueen, 2005; Kraljic & Samuel, 2005, 2007). Although there are cases in which learning in the Norris et al. paradigm generalizes to the speech of other talkers (Kraljic & Samuel, 2005, 2006), the fact that the learning can be talker specific (even to two different talkers at the same time; Kraljic & Samuel, 2007) shows strikingly that the perceptual system is flexible enough to be able to respond to phonetic detail that is specific to individual talkers.

Flexibility alone is not enough, however. For this type of perceptual learning to be of real value in everyday listening, it should be stable over time. Eisner and McQueen (2006) showed that it is stable over at least 12 hr, irrespective of whether participants had slept in the 12 hr between exposure and test or had remained awake and presumably had interacted with multiple other talkers. Kraljic and Samuel (2005) trained participants on an ambiguous

Matthias J. Sjerps, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. James M. McQueen, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; Radboud University Nijmegen (Behavioural Science Institute and Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition), Nijmegen, The Netherlands.

This work was Matthias J. Sjerps’s thesis project for the MSc in cognitive neuroscience at the Radboud University Nijmegen, under the supervision of James M. McQueen. We thank Holger Mitterer for helpful comments on versions of this article. We also thank Jet Sueters for her assistance in running participants.

Correspondence concerning this article should be addressed to Matthias J. Sjerps, Max Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands. E-mail: matthias.sjerps@mpi.nl

[s]–[ʃ] sound in either of two training conditions ([s]- or [ʃ]-trained), and tested them by means of category identifications on a continuum ranging from [s] to [ʃ]. An effect of learning was found, constituted by a shift in identification functions for the two different training groups. These tests took place after a 25-min delay containing one of a number of different types of intervening stimuli. Only stimuli that contained speech of the same speaker with correct instances of [s] and [ʃ] significantly reduced the effect, but importantly, did not make it disappear. Speech of a different speaker containing correct instances of the fricatives did not reduce the effect, nor did intervening speech of the same or a different speaker without instances of the fricatives. The learning effect can diminish over the course of multiple test trials, however (Eisner, 2006; Stevens, 2007; van Linden & Vroomen, 2007), most likely because the test trials themselves induce further perceptual adjustments (Stevens, 2007). In the absence of test-induced unlearning, however, exposure in the lexically guided perceptual learning paradigm seems to induce adaptations to unusual speech that are robust over time.

This relative ease of achieving robust adaptations can be explained partly by the locus of these adaptations. Training-induced learning of speaker idiosyncrasies influences the perception of words that have not been used in training (McQueen, Cutler, & Norris, 2006). This finding shows that the changes take place at a prelexical level of representation, that is, at a level that establishes differences between abstract sublexical (e.g., phonemic) units. The result of the relatively low locus of such adaptations and the fact that they concern abstract phonological representations is that a few training instances lead to changes in the way in which all words containing the adjusted sounds are recognized. This is very useful for the listener because this means that the idiosyncracies of any particular speaker are easily and rapidly adapted to, not only for words that this speaker has already uttered, but also for novel words.

Lexically guided retuning is also very rapid. Exposure to only 20 words bearing the unusual sound in a 9-min lexical decision task (Norris et al., 2003) is sufficient to cause the perceptual changes. Kraljic and Samuel (2007) observed lexically guided perceptual learning with only 10 critical exposures. This underlines just how flexible the speech perception system is: Stable changes in the mapping of phonetic information (that may be unique to a given talker) onto abstract phonemic categories can arise very quickly indeed.

Given this surprising adaptiveness of the speech perception system, it is important to establish the bounds of this flexibility. One such potential limit on flexibility is familiarity. Can sounds be included in a phoneme category only if they are truly novel? The ambiguous mixtures that have been used in previous perceptual-learning experiments were selected to be highly ambiguous and were synthetically created mixtures of, for example, /s/ and /f/. Such sounds have presumably rarely been heard before by participants. In Experiment 1, we tested whether Dutch students can associate a familiar sound category with /f/ or /s/. We used the English sound /θ/ instead of an ambiguous sound halfway between /f/ and /s/. Although /θ/ is not a Dutch phoneme, Dutch university students have had at least 5 years of high school–level English education, and have thus had extensive exposure to the phonological category /θ/ in English. Sounds from this well-established category might be much less prone to be included in another category. Using a familiar nonnative phoneme is a very strong test of the limits on perceptual flexibility. Not only is the category well established, Dutch listeners have also learned that, in its normal use in English, /θ/ is distinctive from both /f/ and /s/.

Following McQueen, Cutler, and Norris (2006), we tested the learning effect by means of a cross-modal identity-priming paradigm. Participants heard spoken prime words, followed immediately by letter strings to which they made lexical decisions. A robust finding in cross-modal identity priming is that an identical prime leads to facilitation, whereas priming with a word that mismatches with one phoneme does not (Marslen-Wilson, Nix, & Gaskell, 1995) and sometimes even leads to inhibition (van Alphen & McQueen, 2006). McQueen, Cutler, and Norris (2006) capitalized on Dutch minimal pairs of the kind *doof/doos* (both are Dutch words; *doof* means "deaf" and *doos* means "box")." Participants who had learned to interpret the ambiguous sound [fs] as, for instance, /s/ (during training where the [fs] sound appeared in /s/-final words) responded faster to the target *doos* when they heard the prime *doo*[fs] than when they heard an unrelated prime. On the other hand, participants who had learned during training to interpret [fs] as /f/ produced the opposite effect at testing: faster responses to *doof* after hearing the prime *doo*[fs] than after hearing an unrelated prime. This method thus acts as a way to infer how participants interpret the ambiguous sound without their making overt judgments on the sound. In the current experiments, we used the same training and testing design, but for the first experiment, we used [θ] instead of the ambiguous [fs] sound (see Table 1).

There were a number of important considerations that led to the decision to use the McQueen, Cutler, and Norris (2006) design. First, listeners were not asked to make direct judgments on the ambiguous sound. The experiment was thus not a meta-linguistic judgment task. This makes it unlikely that performance reflects task-specific strategies on the basis of the sound. Second, and more important, because we used a naturally produced phoneme [θ] as the critical learning item, it would have been difficult to create a digitally mixed [ɛf]–[ɛθ]–[ɛs] test continuum similar to those used in other lexically guided retuning experiments. Third, in line with McQueen, Cutler, and Norris (2006), cross-modal priming provides a measure of how phonologically abstract perceptual retuning influences online word recognition.

The current experiment also addressed another issue. Research looking at lexically guided retuning by means of ambiguous acoustic items has so far used digital mixtures of two natural sounds. Two possible mechanisms could potentially underlie the adjustments that are made after being trained with such digital mixtures of sounds. The first is that learning involves the inclusion of those ambiguous sounds into the category of instances for which they have been trained. The alternative possibility, however, is that listeners simply learn to filter out the contextually incongruent phoneme, that is, to suppress the noise in the ambiguous sound that is inconsistent with the target phoneme. For example, in the condition where participants learned that [fs] represents /f/, participants might have learned to suppress the [s] part of the signal, as if it were simply noise that disrupts perception of the remaining [f] material in the ambiguous sound. If this filtering interpretation were true, the learning effect would only be possible because the ambiguous sounds were artificially made from two clear unambiguous phonemes. Because ambiguous but natural phonemes in the real world are rarely, if ever, constituted by two simultaneous

Table 1
Design Overview, With Examples of Critical Stimuli for Each Experiment

Training phase								Testing phase	
Experiment 1 [θ]		Experiment 2 [fs]		Experiment 3 natural		Experiment 4 [#]			
/f/-group	/s/-group	/f/-group	/s/-group	/f/-group	/s/-group	/f/-group	/s/-group	Auditory primes	Visual targets
drui[θ]	druif	drui[fs]	druif	druif	druif	drui[#]	druif	krop	doos
moeras	moera[θ]	moeras	moera[fs]	moeras	moeras	moeras	moera[#]	doo[?]	doof
								wijn	poef
								poe[?]	poes

Note. Critical stimuli in the training phase were auditory targets. The critical sounds in this phase were [θ], an [fs] mixture, natural [f] or [s], or signal-correlated white noise ([#]) in Experiments 1–4, respectively. These sounds appeared in /f/-final words for the /f/ group or in /s/-final words for the /s/ group. In the testing phase, [?] represents these experiment-specific sounds, which appeared in auditory primes followed by both f- and s-final visual targets. Phonologically unrelated primes were also followed by both types of target.

phonemes, failing to find a learning effect with [θ] would leave open the possibility that lexically guided perceptual learning effects found in the past were an artificial consequence of the stimulus construction technique. On the other hand, finding a learning effect with [θ] would exclude such a filtering interpretation, and would support the idea that adjustments to an existing phoneme category can involve the inclusion of completely new sounds.

The main hypothesis tested in Experiment 1, however, was that although Dutch listeners have acquired a solid sound category for /θ/ in their second language, the flexibility of the perceptual system in dealing with native phonemes will override the claim that the second language has made on the part of acoustic–phonetic space that [θ] occupies. If so, participants who learn during the training phase to interpret [θ] as an instance of /f/ would be facilitated in their responses to *doof* targets after hearing *doo*[θ], whereas participants who learn to interpret [θ] as /s/ would be facilitated on *doos* targets after hearing *doo*[θ].

Experiment 1

Method

Participants. Thirty-eight participants of the Max Planck Institute subject pool were paid to take part in the experiment. None reported any hearing disorders, and all had normal or corrected-to-normal vision. Only 2 had participated in similar experiments, but that was more than a year prior to the current experiment. The average amount of English education for this participant pool is 7–8 years (Broersma & Cutler, 2008).

Materials and stimulus construction.

Training Materials. The stimuli in the training phase were the same items as those that were used in the experiments of Norris et al. (2003) and McQueen, Cutler, and Norris (2006). This set of items consisted of 100 words and 100 nonwords that were phonotactically legal in Dutch. The set of 100 words consisted of 40 training items (see the Appendix) and 60 filler words. The 40 training items consisted of 20 words ending in /f/ (e.g., *druif*, "grape") and 20 words ending in /s/ (e.g., *moeras*, "swamp"). The fillers words did not contain any of the phonemes /f, s, v, z/. The two groups of 20 words consisted of 5 words of one syllable, 5 of two syllables, 5 of three syllables, and 5 of four syllables. Word frequency was matched for these two sets of words (means: /f/ words, 13 per million; /s/ words, 14 per million). Replacing /f/ by /s/ and vice versa in the 40 training items would create nonwords. In one condition of the experiment, the [f] in the /f/-final words was replaced with [θ]. In the other condition, the [s] in the /s/-final words was replaced with [θ] (see Table 1). Participants were thus either presented with normal [s] and [θ] or with [θ] and normal [f].

Digital recordings of the training stimuli were made by a female native speaker of Dutch in a sound-treated booth, sampling at 44 kHz. All 40 target items were recorded in their natural version and in a version ending in the voiceless velar fricative [x] instead of their natural final fricative (e.g., both *druif* ([drœyf]) and *druig* ([drœyx]). Following Eisner and McQueen (2005), this was done to prevent a coarticulatory [f] or [s] bias in the critical materials. When creating the [θ]-final items, the [x]-final versions were used, from which the [x] was removed (cutting at a positive-going zero-crossing at the start of frication) and replaced with an instance of [θ], which was selected by means of two pretests.

Pretest 1. To select an instance of [θ] that was both ambiguous and heard mostly as neither [f] nor [s], two pretests were run. In the first, Dutch listeners were asked to identify the three fricatives [f], [s], and [θ], presented as final phonemes in nine Dutch nonwords. The goal was to find instances of [θ] that were most distinctively not [f] or [s].

Method. Ten members of the Max Planck Institute subject pool were selected on the basis of the same criteria as for the main experiments. Digital recordings were made in a sound-treated booth of the same speaker used in Experiment 1. Recordings were made of instances of final [θ] in 14 English words (e.g., *bath*) and instances of final [s] and [f] in 14 Dutch words (e.g., *das*, "scarf"; *laf*, "cowardly"), each spoken 3 times. Of these items, 10 of each phoneme were selected by informal judgment to be good instances of that particular phoneme. The frication noises of the instances of [s], [f], and [θ] were excised from the recordings, cutting in each case at a positive-going zero-crossing at the onset of frication energy and at a zero-crossing close to the end of frication. Nine Dutch nonwords ending in [x] were also recorded. The frication noise of the final fricative was removed from the offset of each word by cutting at a positive-going zero-crossing at the onset of frication. All of the selected instances of [θ], [s], and [f] were then spliced onto the 9 nonword onsets, resulting in 270 items. Five different randomizations of all 270 items were created, and these 5 versions were paired by a mirror version of itself, resulting in 10 different orderings, 1 for each participant. Listeners received written instructions, and were asked to indicate, on a three-button box (with the labels *th*, *s*, and *f*), what the final phoneme of each nonword was.

Results. Table 2 shows overall percentages of responses to the three different fricatives and the scores for the individual instances of [θ]. Overall, [θ] was the most ambiguous fricative of the three. This is mainly caused by its frequent confusion with [f]. The three critical instances of [θ] selected for the second pretest are marked with a superscript a ([a]). They were selected because they had low proportions of confusions with [f] and high [θ] scores.

Pretest 2. This pretest was a lexical decision experiment in which the three [θ] sounds selected in Pretest 1 replaced 18 word-final instances of [f] and, in another group of listeners, 18 word-final instances of [s]. The aim was to determine which of the [θ] sounds was the most ambiguous, that is, could be interpreted as both /s/ and /f/.

Method. Twelve new members of the Max Planck Institute subject pool were recruited, using the same criteria as before. This pretest was an adaptation of the training phase of the main experiment. Physically, the same fillers and training items were used, minus 2 training-word onsets in each training condition. These items are marked with an asterisk in the Appendix. This resulted in 18 training-word onsets per condition. The selected instances of [θ] were spliced onto every word onset. The two groups of 18 words each consisted of 3 words of one syllable and 5 words each with two, three, or four syllables. Participants were presented with every target onset only once. The target onsets were thus divided into three groups that were balanced for both length and final vowel. Across participants, each of the three versions of [θ] was presented with every target onset an equal number of times.

As in the main experiment, listeners heard [θ] in either /f/- or /s/-final words. A participant in the condition where [θ] replaced [f] was presented with 18 words ending in natural [s] and 18 words that normally end in [f], but where the [f] was replaced with one of the three instances of [θ]. A participant in the condition where [θ] replaced [s] was presented with 18 words ending in natural [f] and 18 words that normally end in [s] for which the [s] was replaced with one of the three instances of [θ]. Three lists were constructed for each of the two conditions, thereby rotating the three instances of [θ] over the three balanced groups of onsets. The rest of the experimental procedure and materials were the same as in the training phase of the main experiment.

Table 2
Pretest 1: Confusion Matrix for Nonword Final Fricatives [f], [s], and [θ]

Stimulus	Response /f/	/s/	/θ/	No response
[f] overall	72	0	24	4
[s] overall	0	94	3	3
[θ] overall	43	5	47	6
[θ]				
1	59	1	36	4
2	46	3	47	4
3	43	3	48	6
4	61	3	31	4
5[a]	40	6	52	2
6[a]	26	8	60	7
7	41	8	47	4
8[a]	31	7	56	7
9	48	4	41	7
10	36	6	49	10

Note. The table displays averaged percentages overall (first three rows), and individual percentages for each of the instances of [θ].
[a] Selected items.

Results. Table 3 displays the percentage of "yes" responses made to words ending in the three [θ]s, and the mean reaction times (RTs) of these responses. The data are split by participants who were presented with [θ] in /s/-final words and participants who heard [θ] in /f/-final words. The table displays only the summaries of responses that were made more than 100 ms after fricative onset. Selection of the item was based on ambiguity; θ-8 was the most ambiguous item in terms of percentage of yes responses, as reflected in the smallest difference score and in terms of RTs. It was therefore selected and used to make the training and test materials for the main experiment.

Testing materials. There were 20 Dutch minimal pairs of the type *doof*/*doos*, [dof]/[dos], "deaf"/"box," as listed in the Appendix. The pairs consist of monosyllabic words. The mean frequencies of occurrence for the /f/ and /s/ versions of the minimal pairs were 35 and 32 per million, respectively. Visual versions of the minimal pairs served as targets (e.g., *doof* and *doos*). These visual versions were paired with unrelated primes (e.g., [krɔp], "head of a lettuce") or related [θ]-final primes derived from the minimal pairs (e.g., [doθ]).

Another 40 filler word targets and 80 filler nonword targets were selected. These targets did not contain any of the phonemes /f, v, s, z/, and varied from one to three syllables in length. The filler word targets were paired with one of three possible types of prime: (a) 20 existing Dutch words ending in [f] (in the /f/-trained condition only) where the [f] was replaced by [θ] (e.g., auditory [xɪθ], based on *gleuf*, "groove," paired with visual *krab*, "crab"), or 20 Dutch words ending in [s] in the /s/-trained condition (e.g., auditory [prɔθ], based on *reis*, "journey," paired with visual *krab*); (b) 10 identical primes (e.g., auditory *loep* paired with visual *loep*, "magnifying glass"); and (c) 10 phonologically related word primes (e.g., auditory *lijm*, "glue," paired with visual *lijn*, "line").

The filler nonword targets were paired with one of four possible primes: (a) 20 phonologically related Dutch words ending in [f] or [s] (depending on training) where the final fricative was replaced with [θ] (e.g., auditory [xɪθ], based on *gif*, "poison," paired with visual *gip* for participants in the /f/-trained condition, or, for participants in the /s/-trained condition, auditory items such as [pauθ], based on *paus*, "pope," paired with visual *paup*); (b) 20 phonologically unrelated word primes ending in [f] or [s] where the fricative was replaced by [θ] (e.g., auditory [prɔθ], based on *prof*, "professor," paired with visual *twouk*, or, for participants in the /s/-trained condition, auditory [nøθ], based on *neus*, "nose," paired with visual *bolg*); (c) 20 phonologically related word primes (e.g., auditory *kleed*, "rug," paired with visual *kleem*); and (d) 20 unrelated word primes (e.g., auditory *robijn*, "ruby," paired with visual *nong*).

The same speaker who had recorded the training materials also spoke the primes for the cross-modal priming phase during the

Table 3
Pretest 2: Overall Percentages of Yes Responses and Mean Reaction Times (RTs, in Milliseconds From Word Offset) for Words Ending in Three Selected Instances of [θ]

Item	Measure	[θ] in /f/-final words	[θ] in /s/-final words	Difference
[θ]-5	% yes	100	81	19
	RTs	236	338	102
[θ]-6	% yes	100	82	18
	RTs	328	459	131
[θ]-8[a]	% yes	100	86	14
	RTs	300	380	80

[a] Selected item.

same recording session. In fact, the items for the different parts of the experiment were intermixed during the recording. A [θ]-final priming version of this minimal pair was created by recording, for example, [dox] and then replacing the final velar fricative [x] by the [θ] sound that was selected via the pretests.

Design and Procedure

Training. A pseudorandom running order was prepared such that the order of the items was in principle the same over the two conditions. The only difference was that on any given experimental word trial, either a natural version of the word could be presented or the [θ]-final version of that word. Participants never heard more than four words or four nonwords in a row. The natural and [θ]-final items were evenly spread through the trials. The first 12 trials did not contain any of the critical trials. Two additional running orders were prepared by also running the two conditions in the opposite order from the 12th trial onwards.

Listeners were individually tested in a sound-dampened experimental carrel. They received written instructions, telling them to judge for every word whether they thought it was an existing Dutch word. Participants were instructed to respond by pressing *Ja* (yes) or *Nee* (no) on a button box. Yes responses were made with the dominant hand. Participants were instructed to respond both as fast and as accurately as possible. Participants were not informed about the ambiguous nature of some of the phonemes. Stimuli were presented over Sennheiser HD 280-13 headphones at a comfortable listening level.

Trials were presented with an intertrial interval of 2 s from word offset. Raw RTs were measured from word onset. In processing the data, word duration was subtracted from the raw RTs to obtain a word-offset measure.

Testing. All 40 critical items (i.e., both versions of the 20 minimal pairs such as *doof/doos*) were presented once to every participant. These were preceded by either an unrelated or a [θ]-final prime (see Table 1). The experiment was constructed such that the two members of a minimal pair never occurred in the same half of the experiment. For example, if *dooθ–doof* occurred in the first half of the experiment, *krop–doos* would occur in the second half of the experiment.

Two versions of the test phase were required to counterbalance the [θ]-final and the unrelated priming conditions across participants. To control for effects of presentation order, we created two orders of each of these where the first half of the critical items were presented in the second half of the experiment and vice versa. This resulted in a total of 4 different versions for the testing phase. The training also consisted of 4 versions (/f/- and /s/-trained, and both of these in reverse order). This resulted in a total of 16 different versions of the experiment. Note that the training condition determined which filler primes were used in the test condition. The fillers in the /f/-trained conditions were the /f/-final words where [f] was replaced by [θ] (e.g., [xɪθ], based on *gleuf*), and the fillers for participants in the /s/-trained condition were the /s/-final words where [s] was replaced by [θ] (e.g., [prɔθ], based on *reis*). These items served to disguise the critical items and to reinforce the learning effect during the testing phase.

In every version, the testing phase consisted of 80 word targets and 80 nonword targets. For all the instances where a prime ended in the fricative [θ], the following target had an equal chance of being phonologically related or unrelated and an equal chance of being a word or a nonword. In the cases where primes where phonologically related to the targets, there was an equal chance of the target being identical or that it mismatched on the last phoneme. A pseudorandom running order was constructed for each version of the testing phase where a participant would never be presented with more than three word or nonword targets in a row. Furthermore, the critical trials were evenly distributed over the running order.

The testing phase immediately followed the training phase. Auditory primes were again presented over the headphones. Participants saw the visual targets on a computer screen situated about 50 cm in front of them. Visual targets were presented in white lowercase Arial letters on a black background, at the acoustic offset of the auditory primes. Participants were instructed to indicate as fast and as accurately as possible whether the words they saw on the screen were Dutch words. Responses were made with the same *Ja* (yes) and *Nee* (no) buttons as were used in the training phase. Yes responses were again made with the dominant hand. RTs were measured from target onset.

Results

Training by lexical decision. All participants who rejected more than 50% of the θ-final items in the training phase were excluded from further analysis (as in Norris et al., 2003). On the basis of this criterion, 6 participants were excluded. These participants were all in the /s/-training condition. Thirty-two participants were thus included in the analysis, 16 in each training condition. Table 4 shows the mean RTs and errors in the training phase (averages are based on the subject analysis).

Analyses of variance (ANOVAs) of the RTs were carried out using participants (*F*1) or items (*F*2) as repeated measures. Two factors were included: training condition (whether participants were trained to interpret [θ] as /f/ or as /s/) and final fricative (whether the item normally ends in [f] or [s]). An interaction between these two factors would reflect the fact that [θ] constituted an imperfect representative of either, or both, /f/ and /s/. This interaction was not found, $F1 < 1$; $F2(1, 38) = 2.40$, $p < .5$, $\eta_p^2 = .059$. A main effect of final fricative was found with subjects as the repeated measure, $F1(1, 30) = 8.75$, $p < .01$, $\eta_p^2 = .226$, but not by items, $F2(1, 38) = 2.24$, $p < .2$, $\eta_p^2 = .056$. A main effect

Table 4
Experiments 1, 2, 3, and 4: Auditory Lexical Decision Performance in the Training Phase (Mean Reaction Times [RTs] in Milliseconds From Word Offset and Mean Percentage of "No" Responses) for Natural, [θ]-, [fs]-, or [#]-Final Critical Items

		Natural fricatives		[θ], [fs], or [#]	
Experiment	Measure	/f/-final words[a]	/s/-final words[b]	/f/-final words[b]	/s/-final words[a]
1, [θ]	Mean RT "yes"	288	294	243	323
	Mean % "no"	3	6	2	22
2, [fs]	Mean RT "yes"	293	363	397	338
	Mean % "no"	1	5	5	9
3, natural [f] and [s]	Mean RT "yes"	288	263		
	Mean % "no"	2	3		
4, [#]	Mean RT "yes"	368	250	230	339
	Mean % "no"	3	7	3	75

[a] Participants in the /s/-trained group in Experiments 1, 2, and 4. [b] Participants in the /f/-trained group in Experiments 1, 2, and 4.

of training condition was found only by items, $F1 < 1$; $F2(1, 38) = 27.60$, $p < .001$, $\eta_p^2 = .421$.

Analysis of the errors did reveal an interaction between final fricative and training condition, $F1(1, 30) = 25.65$, $p < .001$, $\eta_p^2 = .461$; $F2(1, 38) = 12.87$, $p < .005$, $\eta_p^2 = .253$. This effect is mainly caused by the high rejection rate for /s/ words in the [θ]-final condition (when [θ] replaced [s]). The main effect of final fricative was also significant, $F1(1, 30) = 59.26$, $p < .001$, $\eta_p^2 = .664$; $F2(1, 38) = 9.51$, $p < .005$, $\eta_p^2 = .200$, reflecting more rejections for the /s/-final items than for /f/-final items. A main effect was also found for training condition, with more no responses by participants in the /s/ condition than in the /f/ condition, $F1(1, 30) = 16.58$, $p < .001$, $\eta_p^2 = .356$; $F2(1, 38) = 15.01$, $p < .001$, $\eta_p^2 = .283$. As also became apparent in the two pretests, [θ] was ambiguous, but this ambiguity was not symmetric: [θ] was a more acceptable representative of /f/ than of /s/.

Testing by cross-modal identity priming. Table 5 displays the mean RTs and errors in the test phase (based on the subject analysis), subdivided by the factors target type (visual target was f final or s final), prime type (auditory prime was unrelated or [θ] final), and training condition (during the training phase [θ] replaced [s] or [θ] replaced [f]). Figure 1 displays these data as priming effects (the difference in RTs after a [θ]-final prime vs. after an unrelated prime) for all experiments, with the data for Experiment 1 in the leftmost part. ANOVAs revealed a three-way interaction in RTs in the subject analysis, $F1(1, 24) = 11.48$, $p < .005$, $\eta_p^2 = .323$, but not in the items analysis, $F2(1, 38) = 2.42$, $p = .128$, $\eta_p^2 = .06$. A three-way interaction in errors was found in both analyses, $F1(1, 24) = 6.42$, $p < .05$, $\eta_p^2 = .211$; $F2(1, 38) = 5.77$, $p < .05$, $\eta_p^2 = .132$. These interactions are the critical results because they reveal that learning occurred: They show that presenting a [θ] as replacement for [f] instead of [s] during training results in different behavior when participants are subsequently presented with f- and s-final targets preceded by θ-final or unrelated primes.

Other effects that were significant by both $F1$ and $F2$ were as follows. In the RT analysis, there was an interaction of training condition and target type, $F1(1, 24) = 16.21$, $p < .001$, $\eta_p^2 = .403$; $F2(1, 38) = 17.22$, $p < .001$, $\eta_p^2 = .312$. This effect is a reflection of the critical three-way interaction: The priming effects for f-trained participants in the f targets and the s-trained participants in the s targets were the largest effects. Also in the RT analysis, there was an interaction between prime type and the first versus second half of the experiment, $F1(1, 24) = 5.60$, $p < .05$, $\eta_p^2 = .189$; $F2(1, 38) = 4.97$, $p < .05$, $\eta_p^2 = .116$: Participants became faster to respond to targets after unrelated primes in the second half of the experiment. Finally, there was a main effect of prime type, $F1(1, 24) = 26.91$, $p < .001$, $\eta_p^2 = .529$; $F2(1, 39) = 26.78$, $p < .001$, $\eta_p^2 = .413$: Participants were faster to respond to targets after [θ]-final primes than after unrelated primes. No other effects were significant in the error analysis.

The pattern of effects was further investigated in planned pairwise comparisons of the [θ]-final and unrelated prime conditions, collapsing over the two halves of the experiment. The results of these comparisons are displayed in Table 6 (see also the confidence intervals in Figure 1). In each condition, there was significant priming only where predicted. Participants who were trained to interpret [θ] as /f/ responded faster to the f-final

Table 5
Experiment 1: Visual Lexical Decision Performance (Mean Correct Reaction Times [RTs] in Milliseconds From Target Onset and Mean Error Rates) for f- and s-Final Targets in Each Priming Condition for Both Training Conditions

		f-final target		s-final target	
Measure	Training condition	[θ]-final related prime	Unrelated prime	[θ]-final related prime	Unrelated prime
Mean RT	[θ] in /f/-final words	588	688	698	709
	[θ] in /s/-final words	685	714	638	707
Mean % error	[θ] in /f/-final words	3	7	6	3
	[θ] in /s/-final words	6	4	1	3

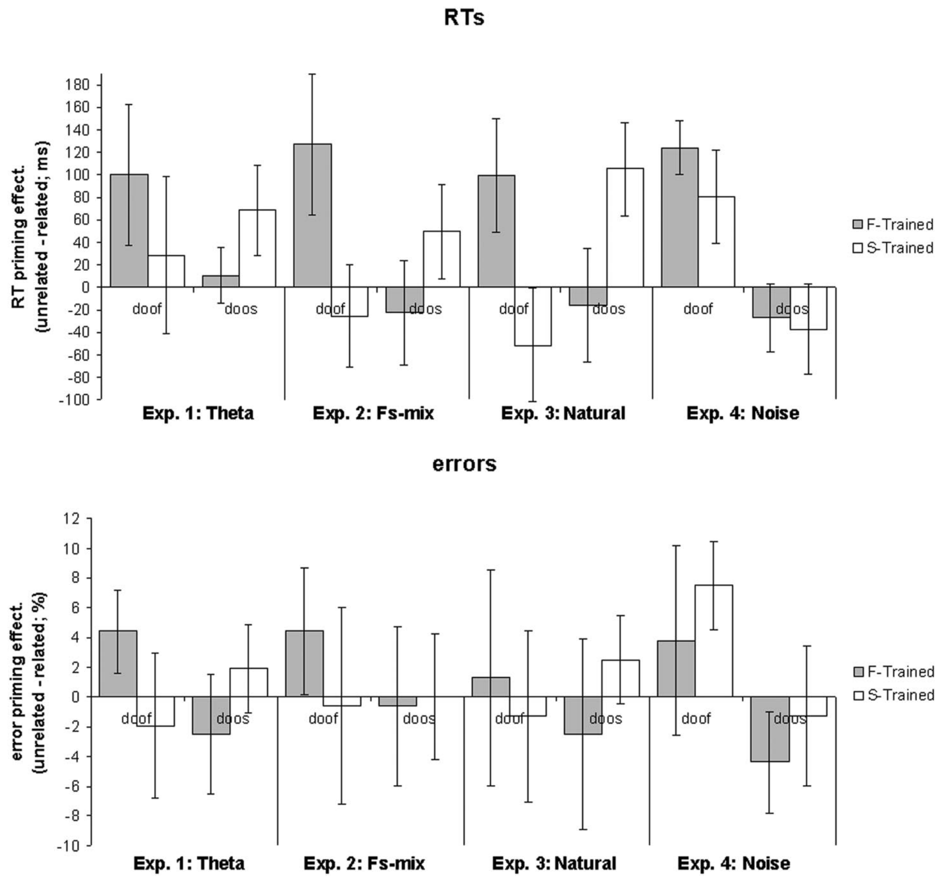

Figure 1. Experiments 1, 2, 3, and 4. Priming effects (target responses after unrelated primes minus target responses after related, [θ]-, [fs]-, natural-, or noise ([#])-final primes) with 95% confidence intervals. The upper panel displays priming effects for reaction times (RTs) and the lower panel displays priming effects for error rates. Priming effects were obtained for responses to visual f-final words (e.g., *doof*) and s-final words (e.g., *doos*) for participants who were trained on the critical sound in /f/-final words (F-Trained) and for participants who were trained on the critical sound in /s/-final words (S-Trained). Positive priming effects in RTs or errors mean, respectively, faster responses or fewer errors than in the unrelated condition.

visual targets (e.g., *doof*) when they had just been primed with a [θ]-final prime such as [doθ] (as compared with an unrelated prime). This priming effect was also reflected in errors for the subject analysis, but not for the items analysis. These participants also did not respond differently to s-final visual targets as a function of prime type. Participants who were trained to interpret [θ] as /s/, however, did not respond faster, or more accurately, to f-final visual targets when they had just heard a [θ]-final prime, but responded faster to the s-final visual targets (e.g., *doos*) when they had just been primed with a [θ]-final prime. There was, however, no effect of accuracy in this condition.

An additional aspect of these data is that the /s/-trained participants tended to show less priming than the /f/-trained participants in their congruent-target conditions (see Figure 1). Moreover, the /s/-trained participants tended to show more priming than the /f/-trained participants in their incongruent-target conditions. This pattern reflects the fact that there seems to be a slight bias to interpret [θ] as /f/ rather than /s/. This trend toward an f bias has been observed in a number of other perceptual learning studies using /f/ and /s/ (McQueen, Cutler, & Norris, 2006; Norris et al., 2003). McQueen, Cutler, and Norris (2006) and Norris et al. (2003) suggested that this bias could have been induced by the fact that ambiguous items had been spliced onto onsets that were recorded as [f] final. Although the current pattern seems to be somewhat smaller here than in the study of McQueen, Cutler, and Norris, removing the coarticulatory bias (by using [x] during the recording as the last phoneme of the materials used to make the [θ]-final primes) did not completely remove this effect. Furthermore, the results of the training part of this experiment already show that [θ] was more likely to be interpreted as /f/ than as /s/. A likely explanation is a higher spectral similarity of [θ] to [f] rather than [s]. Table 7 displays spectral measures of the sounds that were used in these series of experiments (obtained with PRAAT software, Version 5.0.35). Center of gravity refers to the average frequency of the sound. To exemplify, Experiment 4 used a white noise nonspeech sound [#]; as the sounds are all sampled at 22050 Hz, the average frequency of white noise should be close to 5513 Hz (half the nyquist frequency), which it is. Variance refers to the spread of frequencies around the center of gravity (with a value of

Table 6
Experiment 1: Pairwise Comparisons of Priming Effects for f- and s-Final Targets in Each Training Condition

Target/training condition	Measure	Subject analysis *df*	*F*1	*p*	η_p^2	Items analysis *df*	*F*2	*p*	η_p^2
f-final									
[θ] in /f/-final words	Reaction time	1, 15	11.47	.004	.433	1, 19	8.56	.009	.311
	Errors	1, 15	11.67	.004	.438	1, 19	3.20	.090	.144
[θ] in /s/-final words	Reaction time	1, 15	0.77	.40	.049	1, 19	0.70	.41	.036
	Errors	1, 15	0.68	.42	.043	1, 19	0.52	.48	.026
s-final									
[θ] in /f/-final words	Reaction time	1, 15	0.90	.36	.057	1, 19	0.25	.62	.013
	Errors	1, 15	1.67	.22	.100	1, 19	2.92	.10	.133
[θ] in /s/-final words	Reaction time	1, 15	13.41	.002	.472	1, 19	4.58	.046	.194
	Errors	1, 15	1.90	.19	.112	1, 19	1.31	.27	.064

zero for a sine-wave and a high value for white noise). Skewness depends on the relative asymmetry of the frequency distribution (positive skewness reflects relatively more energy for lower frequencies), and kurtosis reflects peakedness of the frequency distribution, with low values for a flat distribution (see the white noise example). Table 7 shows that our [θ] was on three of the four measures (variance, skewness, kurtosis) spectrally closer to a selected [f] than to an [s] sound (these two sounds were in fact used to create an ambiguous [fs] mixture for Experiment 2). The higher spectral similarity of [θ] to [f] probably caused the trend toward an asymmetric learning pattern.

Discussion

A learning effect was found with the English fricative [θ] replacing either of the two Dutch fricative sounds [f] or [s]. Participants who had been trained with [θ] in /f/-biasing contexts interpreted [θ] more as /f/, and participants who had heard [θ] in /s/-biasing contexts interpreted [θ] as representing /s/. In its normal use, however, [θ] is perceived as distinct from [f] and [s] by Dutch listeners (Cutler, Weber, Smits, & Cooper, 2004). Furthermore, Pretest 1 showed that listeners were to some extent able to distinguish the sounds in a Dutch context. The [θ] that we used was identified as /θ/ 56% of the time. Despite this initial discriminability, Experiment 1 showed that appropriate learning could cause the sound to be interpreted as a member of either the /f/ or /s/ category. This finding shows that lexically guided retuning in a native language can, with relatively little exposure, override a nonnative claim on part of the acoustic–phonetic space.

Table 7
Spectral Measures for the Sounds Used in the Experiments ([θ], [fs], and [#]) and the [f] and [s] Sounds Used To Create the [fs] Mixture

Sound	Measure: Center of gravity (Hz)	Variance (MHz)	Skewness	Kurtosis
[f]	4168	4.538	0.898	1.024
[s]	4972	1.284	2.435	7.391
[θ]	8083	5.771	−0.988	0.317
[fs]	4889	1.662	1.528	5.528
[#]	5560	10.616	−0.010	−1.213

Note. Measures were taken from a 40-ms window in the middle of frication of the sound after downsampling to 22050 Hz.

Apart from its training function, the training part of the experiment showed that [θ] was *judged* to be an acceptable representative of /f/ and /s/ most of the time; the results (see Table 4) show that participants interpreted the [θ]-final words as legal Dutch words. However, the testing phase of the experiment revealed that, after training, [θ] could also be *perceived* as representing /f/ or /s/, as its influence on target decisions shows that it engaged in the online process of prime word recognition, without any overt judgments needing to be made about the [θ] sound. This means that [θ] did not only resemble [f] or [s] enough to result in the metalinguistic judgment of sufficient similarity to the appropriate native fricative, presentation of [θ] also resulted in perception of the phoneme that it was learned to represent. That is, there was perceptual evidence of learning in the absence of a meta-linguistic judgment. Moreover, interpretation of [θ] generalized to words that were not in the training part of the experiment. This finding suggests, in line with McQueen, Cutler, and Norris (2006), that adjustments to phonemes are made at a prelexical level.

Another related outcome of Experiment 1 is that it makes a filtering interpretation of lexically guided perceptual learning unlikely. As discussed in the introduction, perceptual learning effects found in the past might have been due to the fact that the ambiguous items that were used in those studies consisted of digital mixtures of two natural sounds. Listeners might have learned to filter the incongruent sound out of the signal instead of including a new sound in a phonemic category. The critical item used here, however, was not a mixture of sounds. A simple filtering hypothesis, therefore, cannot explain either these results or, by extension, previous results based on digitally mixed ambiguous sounds. The results of Experiment 1 thus also further support the idea that this kind of learning is a process that aids listeners in dealing with speaker idiosyncrasies in real life.

The results of Experiment 1 can be taken as a demonstration of considerable flexibility in our perceptual system. However, although the learning effect reported in McQueen, Cutler, and Norris (2006) was replicated here, it is not yet clear whether the two effects are similar in nature. It might be that the participants' familiarity with English, and thus many years of learning that [θ] is distinct from [f] or [s], obstructed the learning process. This may have resulted in a smaller priming effect than was previously observed. Experiment 2 was designed to provide a direct comparison of these two learning effects. It was a replication of Experiment 1, but with the crucial difference that an ambiguous mixture of [f] and [s] was used instead of [θ]. This experiment thus also provides a direct replication of the McQueen, Cutler, and Norris (2006) study using a different speaker.

Experiment 2

Method

Participants. Thirty-three members of the Max Planck Institute subject pool were paid to participate. All participants were selected on the basis of the restrictions given earlier. No participants had participated in a similar experiment before.

Materials, design, and procedure. The materials, design, and procedure of this experiment were exactly the same as in Experiment 1. The only critical difference was that every occurrence of the [θ] sound was replaced with an ambiguous [fs] mix (see Table 1). This sound was selected by means of Pretest 3. Like Pretest 2, this pretest used the lexical decision task. Note that Pretests 2 and 3 are a new way of testing items for their ambiguity. In previous perceptual learning experiments (e.g., Norris et al., 2003), ambiguous items were selected by having listeners label individual, meaningless syllables. Pretesting by means of a lexical decision experiment is an improvement because it provides the exact test of ambiguity that is needed for the training phase of the main experiment.

Pretest 3.

Method. Twelve new participants from the same population as for the previous experiments took part. The procedure in Norris et al. (2003) was used. The same speaker as in Experiment 1 recorded (during the same session) a number of [f] and [s] sounds. These sounds were produced in five different vowel contexts (e.g., *aflas*, *efles*). An informal listening procedure led to the selection of a specific [af]–[as] pair as the best pair of fricatives. The frication parts of these syllables were excised by cutting them off at a positive-going zero-crossing at the start of frication. The sounds were edited to match them for duration and amplitude. Next, a digital mixing procedure was applied (McQueen, 1991; Repp, 1981). This mixing procedure creates a continuum of sounds ranging from [f] to [s] by gradually increasing the amplitude of one sound while decreasing the amplitude of the other. The two endpoints of the continuum represent the two clear fricatives, and the midrange consists of mixtures of the two sounds with attenuated amplitudes. A 21-step continuum was created. The most ambiguous region of the continuum was selected by informal listening. This region was judged to involve Steps 5, 6, and 7. To select the most ambiguous sound, we used the same lexical decision design as in Pretest 2. The only difference was that instead of splicing three versions of [θ] onto the critical word onsets, the three preselected [fs] mixtures were spliced onto these onsets.

Results. The overall percentages of yes responses, along with their mean RTs, made to the three fricative sounds are displayed in Table 8. Only responses that were made at least 100 ms after fricative onset were considered. There were a higher number of yes responses in the [f]-final group than in the [s]-final group; [fs]-6 had the lowest difference score on percentage of yes responses and had the overall highest percentage of yes responses. In terms of RTs, [fs]-7 was the most ambiguous item, and [fs]-6 was the second most ambiguous one; [fs]-6 was selected to be used in the main experiment for which it was spliced onto the training and testing sounds in the same way that [θ] had been spliced onto the materials for construction of Experiment 1.

Table 8
Pretest 3: Overall Percentages of Yes Responses and Mean Reaction Times (RTs, in Milliseconds From Word Offset) for Words Ending in Three Selected Instances of [fs]

Item	Measure	[fs] in /f/-final words	[fs] in /s/-final words	Difference
[fs]-5	% yes	97	86	11
	RTs	256	206	51
[fs]-6[a]	% yes	100	94	6
	RTs	226	267	41
[fs]-7	% yes	97	82	15
	RTs	144	173	30

[a] Selected item.

Results

Training by lexical decision. As in Experiment 1, all participants who rejected more than 50% of the [fs]-final items in this training phase were excluded from further analysis. This criterion led to the exclusion of 1 participant from the /s/-trained condition. Analyses were thus based on 16 participants in each group. Table 4 shows a summary of the lexical decision performance.

ANOVAs like those in Experiment 1 were carried out. An interaction between training condition and final fricative was found in the RT analysis, $F1(1, 30) = 6.83$, $p < .05$, $\eta_p^2 = .185$; $F2(1, 38) = 13.11$, $p < .005$, $\eta_p^2 = .256$. A main effect for training condition was found only by items, $F1(1, 30) = 1.24$, $p < .5$, $\eta_p^2 = .040$; $F2(1, 38) = 12.36$, $p < .005$, $\eta_p^2 = .245$: Participants in the /f/-trained group were relatively slow. Analysis of errors revealed an interaction between final fricative and training condition by participants, but not by items, $F1(1, 30) = 7.75$, $p < .01$, $\eta_p^2 = .205$; $F2(1, 38) = 3.67$, $p < .1$, $\eta_p^2 = .088$. This was because participants in the /s/-trained group accepted most natural /f/-final words but rejected 9% of the [fs]-final words, unlike participants in the /f/-trained group, whose responses were much more balanced. A main effect was found for final fricative by participants, $F1(1, 30) = 7.75$, $p < .01$, $\eta_p^2 = .205$; $F2(1, 38) = 2.09$, $p < .5$, $\eta_p^2 = .052$, reflecting the higher percentage of no responses for /s/-final words. No other effects were significant.

The mixed [fs] sound was accepted as representing both /f/ and (in other participants) /s/. The ambiguity was again not completely symmetric, in accordance with the results of Pretest 3.

Testing by cross-modal identity priming. Table 9 displays the mean RTs and errors in the testing phase. Figure 1 displays the mean priming effects. ANOVAs equivalent to those in Experiment 1 revealed a three-way interaction in RTs, $F1(1, 24) = 24.51$, $p < .001$, $\eta_p^2 = .505$; $F2(1, 38) = 27.16$, $p < .001$, $\eta_p^2 = .417$. This three-way interaction pattern was not present in the errors, $F1(1, 24) = 1.59$, $p < .5$, $\eta_p^2 = .062$; $F2(1, 38) = 1.29$, $p < .5$, $\eta_p^2 = .033$.

Table 9
Experiment 2: Visual Lexical Decision Performance (Mean Correct Reaction Times [RTs] in Milliseconds From Target Onset and Mean Error Rates) for f- and s-Final Targets in Each Priming Condition, for Both Training Conditions

	f-final target		s-final target	
Measure/training condition	[fs]-final related prime	Unrelated prime	[fs]-final related prime	Unrelated prime
Mean RT				
[fs] in /f/-final words	662	789	784	762
[fs] in /s/-final words	693	667	588	637
Mean % error				
[fs] in /f/-final words	4	8	6	5
[fs] in /s/-final words	6	5	3	3

Other effects that were significant by both $F1$ and $F2$ were as follows. Only in RTs there was an interaction of training condition and target type, $F1(1, 24) = 27.01$, $p < .001$, $\eta_p^2 = .530$; $F2(1, 38) = 22.45$, $p < .001$, $\eta_p^2 = .371$; this is again a reflection of the critical interaction collapsed over prime type. Also only in RTs, there were main effects of prime type, $F1(1, 24) = 7.21$, $p < .05$, $\eta_p^2 = .231$; $F2(1, 38) = 4.92$, $p < .05$, $\eta_p^2 = .115$ (participants responded to targets faster after ambiguous than after unrelated primes), and training condition, $F1(1, 24) = 7.40$, $p < .05$, $\eta_p^2 = .235$; $F2(1, 38) = 84.75$ $p < .001$, $\eta_p^2 = .690$ (participants in the /s/-condition were faster overall). No effects reached significance in the error analysis. The factor of first versus second half of the experiment was not involved in any of the interactions that were significant by both $F1$ and $F2$. The effect thus remained stable over the course of the experiment. The factor of first versus second half of the experiment was dropped from further analyses.

The three-way effect was again further investigated by looking at the planned pairwise comparisons of the ambiguous and unrelated prime conditions (displayed in Table 10, and as confidence intervals in Figure 1). These pairwise comparisons again revealed a significant priming effect only where predicted. Participants who had been trained to interpret [fs] as /f/ responded faster to the f-final visual targets (e.g., *doof*) when they had been primed with an ambiguous prime than with an unrelated prime. This effect was found in RTs and in errors for participants, but not for items. These participants did not respond more slowly, or more inaccurately, to s-final visual targets. Participants who were trained to interpret [fs] as /s/ did not respond more slowly, or more inaccurately, to f-final visual targets when they had just been primed with an ambiguous prime. But they did respond more quickly to the s-final visual targets (e.g., *doos*), although not more accurately.

A final analysis compared the results of Experiments 1 and 2, testing in particular for a four-way interaction between prime type, target type, training condition, and experiment. This four-way interaction tests whether the nature of the critical three-way interaction varied across experiments (thereby as a function of the preselected critical sound). The interaction was absent in RTs, $F1(1, 48) = 2.56$, $p < .5$, $\eta_p^2 = .051$; $F2 < 1$, and errors ($F1$ and $F2 < 1$). There is thus no evidence to assume that the learning effects obtained with the two critical sounds differ in size or nature.

Discussion

A learning effect was found using an ambiguous [fs] sound. This finding provides a direct replication of the results found by McQueen, Cutler, and Norris (2006). Critically, no difference was found between this effect and that in Experiment 1. That is, the priming effect obtained with the ambiguous [fs] sound was indistinguishable from that obtained with a second-language phoneme that is normally distinguished from [f] and [s]. In other words, the established status of the sound category [θ] does not seem to have

Table 10
Experiment 2: Pairwise Comparisons of Priming Effects for f- and s-Final Targets in Each Training Condition

		Subject analysis				Items analysis			
Target/training condition	Measure	*df*	*F*1	*p*	η_p^2	*df*	*F*2	*p*	η_p^2
f-final									
[fs] in /f/-final words	Reaction time	1, 15	19.07	.001	.560	1, 19	12.90	.002	.404
	Errors	1, 15	4.62	.048	.236	1, 19	2.06	.167	.098
[fs] in /s/-final words	Reaction time	1, 15	1.40	.26	.085	1, 19	0.93	.35	.047
	Errors	1, 15	0.04	.84	.003	1, 19	0.11	.748	.006
s-final									
[fs] in /f/-final words	Reaction time	1, 15	1.03	.33	.064	1, 19	0.33	.57	.017
	Errors	1, 15	0.06	.81	.004	1, 19	0.06	.80	.003
[fs] in /s/-final words	Reaction time	1, 15	6.30	.024	.296	1, 19	5.53	.030	.225
	Errors	1, 15	0.00	1.0	.000	1, 19	.00	1.0	.000

influenced our listeners' ability to learn that this sound is /f/ or /s/. [θ] was adapted to as quickly and as thoroughly as the type of ambiguous sound that has been used in most other perceptual learning experiments of this type.

But just how thorough is this learning? If one comes across an individual with an odd pronunciation in daily life, how complete is learning about that pronunciation? It may be that the new variant comes to be treated as equivalent to more prototypical variants, or it may be that it is less acceptable than prototypical tokens. To answer this question, an adaptation of the previous cross-modal priming experiments was designed. Instead of measuring the identity-priming effect of ambiguous primes, Experiment 3 measured the identity-priming effect of unambiguous primes, such as, for example, the effect of the natural prime *doof* on targets such as *doof* or *doos*. This experiment again consisted of two conditions, where listeners in one condition were primed with [f]-final words such as *doof*, and listeners in the other condition heard [s]-final primes such as *doos* before deciding on the targets *doof* or *doos* (see Table 1). In other words, this experiment measured the cross-modal priming effect that is obtained with versions of [s] or [f] that are normal exemplars of their phoneme categories, like those that are used in everyday life. Comparing this natural-fricative experiment with the previous experiments will reveal how thorough the adaptations to newly acquired items really are.

Experiment 3

Method

Participants. Thirty-two further members of the Max Planck Institute subject pool were paid to participate. As in the previous experiments, none reported any hearing disorders, and all had normal or corrected-to-normal vision. Again, most participants had never participated in similar experiments, and for those who had (4 participants), it had been more than a year previously.

Materials and Stimulus Construction

Training materials. The training part of the experiment consisted of the same physical items as in the previous experiments except that no ambiguous sounds were used. Where Experiments 1 and 2 had two different training conditions, one biasing the critical sound to be interpreted as /f/ and the other biased to interpret the critical sound as /s/, this experiment has only one training version for all participants. This contained only the natural versions of the 40 experimental items (e.g., *moeras* and *druif*). This phase thus did not have a training effect, but was present so that the participants underwent a similar procedure to those in Experiments 1 and 2.

Testing materials. All items that ended in an ambiguous sound in the previous experiments were replaced with tokens ending in an unambiguous natural fricative. This fricative was an [f] for one group of participants (e.g., they heard *doof*) and an [s] for the other group of participants (e.g., they heard *doos*). This change involved both the critical experimental items and the fillers. The unambiguous versions of the critical minimal pairs and the fillers were recorded during the same session and always directly before or after the velar version that was used for cross-splicing in the previous experiments. Unlike the previous two experiments, the test materials did not involve cross-spliced versions of the /f/- and /s/-final words. As a result, this experiment provides an estimate of the amount of priming arising from completely natural words.

Design and procedure. The design and procedure were the same as for the previous experiments, with the exception that one group heard only [f]-final related primes instead of [θ]- or [fs]-final primes, and the other group heard only [s]-final related primes.

Results

Training by lexical decision. None of the participants had to be rejected as a result of the 50% criterion. Table 4 displays the results for the lexical decision phase. One-way ANOVAs were carried out on the RTs and errors, with final fricative as the independent variable. No significant differences between the two groups of words ([f]- or [s]-final) were found.

Testing by cross-modal identity priming. Table 11 displays the mean RTs and percentage of errors for the different conditions in the testing phase of the experiment. The factors were target type (visual target was f- or s-final), testing condition (participants were presented with either [s]- or [f]-final experimental primes), and prime type (participants were primed with either an unrelated prime or with a prime ending with [f] or [s]). Figure 1 displays the priming effects for the four prime–target combinations. ANOVAs with both subjects and items as repeated measures revealed a strong three-way interaction in RTs, $F1(1, 24) = 43.74$, $p < .001$, $\eta_p^2 = .646$; $F2(1, 38) = 70.02$, $p < .001$, $\eta_p^2 = .648$, but not in errors, $F1(1, 24) = 3.00$, $p < .10$, $\eta_p^2 = .111$; $F2(1, 38) = 3.26$, $p < .1$, $\eta_p^2 = .079$. The interaction in RTs reflects the predicted cross-modal identity priming effect (participants respond faster when primed with the same word) in both the /f/ and /s/ conditions.

The following effects also reached significance by both *F*1 and *F*2. In RTs, there was an interaction of testing condition by target type, $F1(1, 24) = 35.15$, $p < .001$, $\eta_p^2 = .594$; $F2(1, 38) = 86.80$, $p < .001$, $\eta_p^2 = .695$, again a reflection of the critical three-way interaction. Also in RTs, there were main effects for prime type, $F1(1, 24) = 7.77$, $p < .01$, $\eta_p^2 = .245$; $F2(1, 38) = 10.92$, $p < .005$, $\eta_p^2 = .223$ (participants were overall faster to respond to targets after [s]- or [f]-final primes than to targets after unrelated

Table 11
Experiment 3: Visual Lexical Decision Performance (Mean Correct Reaction Times [RTs] in Milliseconds From Target Onset and Mean Error Rates) for f- and s-Final Targets in Each Priming Condition for Both Testing Conditions

	f-final target		s-final target	
Measure/testing condition	[f]- or [s]-final related prime	Unrelated prime	[f]- or [s]-final related prime	Unrelated prime
Mean RT				
[f]-final primes	602	701	709	693
[s]-final primes	736	685	560	666
Mean % error				
[f]-final primes	4	6	7	4
[s]-final primes	8	7	1	3

primes), and for the first versus second half of the experiment, $F1(1, 24) = 6.97$, $p < .05$, $\eta_p^2 = .225$; $F2(1, 38) = 9.19$, $p < .005$, $\eta_p^2 = .195$, because responses overall became faster in the second half of the experiment. One effect reached significance in errors by both $F1$ and $F2$. This involved, again, an interaction of testing condition by target type, $F1(1, 24) = 7.00$, $p < .05$, $\eta_p^2 = .225$; $F2(1, 38) = 5.62$, $p < .05$, $\eta_p^2 = .129$. There were no interactions of other factors with the factor of first versus second half of the experiment. This factor was dropped from subsequent analyses.

Planned pairwise comparisons revealed priming in the expected directions, displayed in Table 12, and by means of confidence intervals in Figure 1. Participants who were primed with words ending in [f] (e.g., *doof*) were faster (but not more accurate) saying yes to the identity targets (e.g., *doof*) than after the unrelated primes (e.g., *krop*). But these participants were not significantly slower (or less accurate) saying yes to the mismatching targets (e.g., *doos*). In contrast, participants who were primed with words ending on [s] (*doos*) were slower (but no less accurate) saying yes to the mismatching targets (*doof*) than after the unrelated primes. Finally, these participants were faster (but not more accurate) saying yes to the identity targets.

An analysis comparing Experiments 1 and 3 investigated whether the priming effect was altered as a function of whether the effect was obtained with a natural prime or with a prime containing a newly learned [θ], representing /f/ or /s/. This analysis revealed an interaction for RTs by participants, $F1(1, 48) = 6.41$, $p < .05$, $\eta_p^2 = .118$, but not for RTs with item as a repeated measure, $F2(1, 76) = 2.77$, $p = .1$, $\eta_p^2 = .035$, and for neither of the error analyses ($F1$ and $F2 < 1$). There was thus no robust evidence for a difference in processing between the items in the two experiments. A similar four-way analysis comparing Experiments 2 and 3 did not reveal an interaction in RTs, $F1 < 1$; $F2(1, 76) = 1.11$, $p < .5$, $\eta_p^2 = .014$, or errors ($F1$ and $F2 < 1$). There is thus no evidence that the priming effect obtained with a newly learned ambiguous [fs] sound differs from the priming effect obtained with natural fricatives.

Discussion

An expected cross-modal identity priming effect was obtained with natural versions of the word final fricatives [f] and [s]. The critical four-way test, investigating how thorough the newly learned items were processed, revealed an interaction between Experiment 1 ([θ]) and Experiment 3 (natural fricatives) on one measure (RTs on $F1$). This finding indicates a tendency of the newly learned item [θ] to be processed differently than natural instances of the fricatives that [θ] replaced. However, this interaction was absent on the other three measures. Although the interpretation of this unreliable tendency for a difference is not clear, the lack of an effect on three measures suggests that there is no substantial difference between the identity-priming effect obtained with the reallocated second-language sound and fricatives that have been used throughout life.

The test for thoroughness of processing for the [fs]-mix sound revealed no four-way interaction, leading to a more straightforward interpretation. It appears that there are no differences in the processing of a newly acquired ambiguous sound and sounds that constitute natural exemplars of their phonemic categories. Odd versions of a native phoneme can thus be adapted to quickly and thoroughly.

But how flexible is the speech perception system then? Given that both [θ] and [fs] were so readily and almost completely accepted as tokens of [f] or [s], it might appear that there is no limit on what sounds can be accepted as new instances of native phonemic categories. For example, what about the unclaimed territory in acoustic–phonetic space that a nonspeech sound occupies? That is, can a nonspeech sound be accepted as a native phoneme, even though it was not produced by a vocal tract? This question was tested in Experiment 4 with signal-correlated noise, which is not a speech sound, but which does have the amplitude, duration, and spectral range of a speech sound. A signal-correlated noise version of [θ] was used (see Table 1).

Signal-correlated noise produces a flat spectrum within the amplitude envelope of the source sound (Schroeder, 1968). A study by Jongman, Wayland, and Wong (2000), looking at acoustic characteristics of English fricatives, showed that a fricative sound with a flat spectrum, expressed in spectral variance, is closer to the average [f] than it is to [s]. This is confirmed by the spectral measures displayed in Table 7. Therefore, it may be that the nonspeech sound ([#]) will prove to be a better representative of /f/ than of /s/. Related to this prediction is the finding that spectral similarities can play an important role in generalization of perceptual learning. Although learning with fricatives tends not to generalize to other speakers (suggesting that idiosyncratic articulations of fricatives are stored in a speaker-specific manner; Eisner

Table 12
Experiment 3: Pairwise Comparisons of Priming Effects for f- and s-Final Targets in Each Testing Condition

		Subject analysis				Items analysis			
Target/testing condition	Measure	*df*	*F*1	*p*	η_p^2	*df*	*F*2	*p*	η_p^2
f-final									
[f]-final primes	Reaction time	1, 15	17.57	.001	.539	1, 19	53.00	.001	.736
	Errors	1, 15	0.14	.72	.009	1, 19	0.49	.49	.025
[s]-final primes	Reaction time	1, 15	4.75	.046	.240	1, 19	7.61	.013	.286
	Errors	1, 15	0.21	.65	.014	1, 19	0.19	.67	.010
s-final									
[f]-final primes	Reaction time	1, 15	0.46	.51	.030	1, 19	0.88	.36	.044
	Errors	1, 15	0.71	.41	.045	1, 19	1.36	.26	.067
[s]-final primes	Reaction time	1, 15	29.24	.001	.661	1, 19	29.83	.001	.611
	Errors	1, 15	3.00	.10	.167	1, 19	2.11	.16	.100

& McQueen, 2005; Kraljic & Samuel, 2006, 2007), generalization across speakers using fricatives can be obtained when the training item of one speaker has a spectral mean that is close to that of the testing items of another speaker (Kraljic & Samuel, 2005). When the spectral difference is large, however, learning does not generalize to another speaker.

If the nonspeech sound can be perceived as an instance of existing phonemes, Experiment 4 will show that a sound does not have to be produced by a human vocal tract to be perceived as speech. Furthermore, if the asymmetric spectral similarity of [#] to [f] and [s] is expressed in a different learning pattern, Experiment 4 will show that learning requires such similarity. That is, it will show that there are limits on the flexibility of the speech perception system.

Experiment 4

Method

Participants. Thirty-two new participants were recruited. These participants were selected to meet the experimental requirements stated earlier. Two of the participants had taken part in a similar experiment, but more than a year previously.

Materials, design, and procedure. The ambiguous sound for this experiment ([#]) consisted of a signal-correlated version of the [θ] sound used in the first experiment. In the signal-correlated noise procedure, a random decision is taken for every sample of a signal whether to multiply the sample by 1 or by –1 (Schroeder, 1968). This procedure results in a signal that has the same duration and amplitude as the original, but has a flat spectrum. The experiment was otherwise identical to Experiment 1.

Results

Training by lexical decision. Unlike the previous experiments, participants who did not reach the 50% criterion were included in the analyses. This decision was made because of the large number of participants who did not reach this criterion (11 participants, all in the /s/-trained condition). This high number of rejections reflects the fact that the nonspeech sound was judged to be a poor representative of the phoneme /s/. Table 4 shows a summary of the lexical decision phase.

Because of the high number of no responses, an analysis of RTs to the yes responses was not carried out. ANOVAs in errors revealed a strong interaction between training condition (did the nonspeech sound replace [f] or [s]?) and final fricative (experimental item in /s/- or /f/-final word): $F1(1, 30) = 78.67$, $p < .001$, $\eta_p^2 = .724$; $F2(1, 38) = 944.52$, $p < .001$, $\eta_p^2 = .961$. Main effects were revealed for both training condition, $F1(1, 30) = 80.37$, $p < .001$, $\eta_p^2 = .728$; $F2(1, 38) = 944.52$, $p < .001$, $\eta_p^2 = .961$, and final fricative, $F1(1, 30) = 98.55$, $p < .001$, $\eta_p^2 = .767$; $F2(1, 38) = 283.93$ $p < .001$, $\eta_p^2 = .882$. These strong effects all reflect the fact that the nonspeech sound did not seem to sound like an /s/. But the nonspeech sound was accepted as an /f/, to a similar extent to the [θ] and [fs] mix in Experiments 1 and 2.

Although no RT analyses were carried out, one aspect of the RT data should be mentioned: Participants in the /f/ condition responded much faster to the [#]-final words than the participants in the /s/ condition responded to the *natural* /f/-final words. The participants in the s-trained condition were confronted with s-final words that ended in the nonspeech sound. As a result, they frequently judged these items to be (possibly /f/-final) nonwords. These participants became more cautious, leading to longer RTs in the s-trained compared with the f-trained condition. The apparent difference between item types is thus likely to be a main effect of group.

Testing by cross-modal identity priming. Table 13 displays the mean RTs and errors in the testing phase. As in the previous experiments, the table shows the crossed factors target type (visual prime was f- or s-final), prime type (auditory prime was unrelated or ambiguous), and training condition (during the training phase [#] replaced [s] or [#] replaced [f]). Figure 1 displays these data as priming effects. ANOVAs revealed that the critical three-way interaction was not present in RTs, $F1(1, 24) = 1.11$, $p < .5$, $\eta_p^2 = .044$; $F2 < 1$, or errors ($F1$ and $F2 < 1$). The absence of this three-way interaction shows that participants in the two training groups did not respond differently to the primes.

Other effects that did reach significance in both the $F1$ and $F2$ analysis in RTs were the following. An interaction effect was found for target type and prime type, $F1(1, 24) = 73.05$, $p < .001$, $\eta_p^2 = .753$; $F2(1, 38) = 55.44$, $p < .001$, $\eta_p^2 = .593$, which reflects the fact that priming was much stronger in the f-target condition. Main effects were found for prime type, $F1(1, 24) = 24.72$, $p < .001$, $\eta_p^2 = .507$; $F2(1, 38) = 15.01$, $p < .001$, $\eta_p^2 = .283$ (the priming effect on the f targets was bigger than the inhibition on the s targets), target type, $F1(1, 24) = 42.28$, $p < .001$, $\eta_p^2 = .638$;

Table 13
Experiment 4: Visual Lexical Decision Performance (Mean Correct Reaction Times [RTs] in Milliseconds From Target Onset and Mean Error Rates) for f- and s-Final Targets in Each Priming Condition for Both Training Conditions

	f-final target		s-final target	
Measure/training condition	[#]-final related prime	Unrelated prime	[#]-final related prime	Unrelated prime
Mean RT				
[#] in /f/-final words	575	699	723	696
[#] in /s/-final words	578	659	679	641
Mean % error				
[#] in /f/-final words	4	8	8	3
[#] in /s/-final words	2	9	8	6

$F2(1, 38) = 7.55$, $p < .01$, $\eta_p^2 = .166$ (the priming for f-final targets was larger than the priming for s-final targets), and the first versus second half of the experiment, $F1(1, 24) = 4.98$, $p < .05$, $\eta_p^2 = .172$; $F2(1, 38) = 4.56$, $p < .05$, $\eta_p^2 = .107$ (participants became faster overall in the second half of the experiment). The error analysis also revealed an interaction between target type and prime type, $F1(1, 24) = 16.44$, $p < .001$, $\eta_p^2 = .407$; $F2(1, 38) = 7.01$, $p < .05$, $\eta_p^2 = .156$, confirming that priming was stronger in the f-target condition.

Planned pairwise comparisons were carried out on the data, collapsed over the first versus second half of the experiment. The results of these comparisons are displayed in Table 14. The pairwise comparisons confirm the priming pattern shown by the confidence intervals in Figure 1. Participants who had been trained to interpret [#] as /f/ responded significantly faster to the f-final visual targets (e.g., *doof*) when primed with a [#]-final prime (e.g., *doo#*) than after an unrelated prime. This effect was found in RTs but was absent in errors. Participants who had been trained to interpret [#] as being an instance of /s/ were also significantly primed when responding to an f-final target. Despite what they were encouraged to do during training, these participants apparently interpreted [#] as /f/. This effect was found in RTs and errors. Participants who were trained to interpret [#] as /f/ tended to respond more slowly to s-final targets. Participants who had been trained with [#] in an /s/-biasing context also tended to respond more slowly to s-final targets. Despite their training condition, these participants thus showed some evidence of inhibition in their responses to s-final targets.

A four-way analysis comparing Experiments 3 and 4 revealed an interaction in RTs, $F1(1, 48) = 21.25$, $p < .001$, $\eta_p^2 = .307$; $F2(1, 76) = 31.59$, $p < .001$, $\eta_p^2 = .294$, but not in errors, $F1(1, 48) = 1.83$, $p < .5$, $\eta_p^2 = .037$; $F2(1, 76) = 1.46$, $p < .5$, $\eta_p^2 = .019$. This effect shows that the processing of the nonspeech item is different from the processing of normal instances of [f] and [s], mainly due to the difference between the nonspeech stimulus and [s].

The final analysis was a comparison between the nonspeech /f/-trained and the natural /f/ conditions in Experiment 3. This comparison addressed whether the processing of [#] was, in this experimental design, different from the processing of a natural /f/. The three-way interaction between experiment, trial type, and prime type was not significant in RTs, $F1(1, 24) = 1.14$, $p < .5$, $\eta_p^2 = .045$; $F2(1, 38) = 0.85$, $p < .5$, $\eta_p^2 = .022$, or errors, $F1(1, 24) = 0.80$, $p < .5$, $\eta_p^2 = .032$; $F2(1, 38) = 1.06$, $p < .5$, $\eta_p^2 = .027$. This suggests that the signal-correlated noise was processed as a natural instance of /f/.

Discussion

No three-way interaction, indicative of a lexically induced learning effect, was found using a nonspeech sound as final fricative on the critical test trials. The nonspeech sound could thus not be learned, using lexical evidence, to represent both /f/ and /s/. This failure was due to the fact that the nonspeech sound proved to be a poor representative of the phoneme /s/, and was instead perceived as an instance of /f/ even when the training part of the experiment suggested otherwise. The most likely reason for this is the higher spectral similarity between [#] and [f], compared with that of [#] and [s] (see Table 7). But the unclaimed territory in acoustic–phonetic space that [#] occupies could apparently effortlessly be associated with /f/. Experiment 4 therefore leads to two conclusions. First, phonological categories are surprisingly flexible, as they can extend to nonspeech sounds. Moreover, adjustments to such sounds can be so thorough that the sounds appear to be treated like natural instances of phonemes. Second, however, there are limits on this flexibility: A nonspeech sound will be treated as a speech sound only if it shares sufficient spectral characteristics with the sound that it replaces.

General Discussion

Dutch listeners can learn to include an exemplar of an acquired second-language phoneme ([θ]) in the native phonemic categories /f/ or /s/ (Experiment 1), even though [θ] is distinguished from these phonemes in its natural use. The same learning effect was obtained with a digitally mixed ambiguous [fs] sound (Experiment 2), providing a replication of the study of McQueen, Cutler, and Norris (2006). A comparison between the priming effects obtained in these two experiments revealed that the learning effects obtained with [θ] and [fs] were indistinguishable. Experiment 3 tested the priming effects obtained with natural instances of [f] and [s] and, hence, provided a benchmark for the effects obtained in Experiments 1 and 2. Cross-experiment comparisons showed that

Table 14
Experiment 4: Pairwise Comparisons of Priming Effects for f- and s-Final Targets in Each Training Condition

Target/training condition	Measure	Subject analysis				Items analysis			
		df	*F*1	*p*	η_p^2	*df*	*F*2	*p*	η_p^2
f-final									
[#] in /f/-final words	Reaction time	1, 15	119.22	.001	.888	1, 19	53.09	.001	.736
	Errors	1, 15	1.71	.21	.102	1, 19	1.88	.19	.090
[#] in /s/-final words	Reaction time	1, 15	17.54	.001	.539	1, 19	25.11	.001	.569
	Errors	1, 15	27.00	.001	.643	1, 19	4.44	.049	.189
s-final									
[#] in /f/-final words	Reaction time	1, 15	3.61	.077	.194	1, 19	2.16	.16	.102
	Errors	1, 15	7.74	.014	.340	1, 19	2.81	.11	.129
[#] in /s/-final words	Reaction time	1, 15	3.95	.066	.208	1, 19	7.56	.013	.285
	Errors	1, 15	0.32	.58	.021	1, 19	0.21	.65	.011

adjustments to newly learned instances of existing phonemes are very thorough: The priming effects with the new sounds patterned like those with natural [f] and [s] sounds. This thoroughness of learning is what makes this form of perceptual adaptation very useful in daily life; only a restricted number of occurrences of a particular odd phoneme is enough to cause subsequent effectively normal processing. In everyday listening, therefore, words containing such odd sounds can quickly be recognized correctly.

Experiment 1 also showed that it is not true that, in previous experiments (e.g., Norris et al., 2003; McQueen, Cutler, & Norris, 2006), listeners learned to interpret the ambiguous [fs] mixtures as either /f/ or /s/ by learning to filter out the incongruent fricative. This would have meant that these previous lexically guided perceptual learning findings would largely have been due to the synthetic nature of the stimuli that were used (the ambiguous sounds were composed of digital mixtures consisting of both /f/ and /s/). Experiment 1 demonstrated that a naturally produced sound that was neither /f/ nor /s/ could also be incorporated in both the /f/ and /s/ categories. A filtering mechanism is therefore not a very likely cause of lexically guided perceptual learning.

The speech-perception system thus appears to be so flexible that even a familiar nonnative sound can rapidly be accepted as an instance of a native category. Acquiring new instances of native phonemes is not a process of unbounded flexibility, however. Experiment 4 showed that [#] could be learned to represent /f/, but not /s/, reflecting, as Jongman et al. (2000) showed, that fricative sounds with a flat spectrum are closer to the average [f] than to the average [s]. The findings by Jongman et al. also explain why [θ] appears to be a better representative of /f/ than of /s/, as [θ] is spectrally closer to [f] than to [s] on a number of important indicators such as spectral peak and spectral variance. These differences reflect to a great extent the difference between sibilants (including /s/) and nonsibilants (including /θ/ and /f/). Spectral analyses of the sounds that we used here resulted in comparable patterns of similarity (see Table 7). Another example of the importance of spectral characteristics comes from the perceptual learning article by Kraljic and Samuel (2005). Their study revealed that training on a female [s]–[ʃ] ambiguous item transferred to both a female and a male test continuum, whereas training with a male item only transferred to a male testing continuum. Spectral analysis revealed that the spectral means of the female training items were in the middle of the female and male testing items. The spectral means of the male training items, however, were close only to the mean of the male testing continuum. When listeners need to learn to accept a new instance of an existing phoneme, similarity to more prototypical instances on acoustic aspects of the signal thus seems to be an important prerequisite. Moreover, once an ambiguous sound has been learned, the same factors constrain generalization to other ambiguous sounds.

This finding suggests that bottom-up information played an important role in determining the interpretation of the nonspeech sound in Experiment 4. But the learning in Experiments 1 and 2 must also have relied on feedback from the lexicon; lexical feedback during exposure caused a retuning of our participants' phoneme categories, leading to enhanced interpretation of the ambiguous sounds during the testing phase (as in Norris et al., 2003). It has been suggested that lexical feedback also influences online perceptual processing, such that lexical activation feeds back immediately to prelexical representations (McClelland, Mirman, & Holt, 2006). Feedback for learning in speech perception does not require online feedback, however (McQueen, Norris, & Cutler, 2006; Norris et al., 2003; Norris & McQueen, 2008). Feedback for learning alone could cause changes to prelexical representations that lead to different phonological interpretations of sounds only over time, that is, on subsequent presentation of the critical sound.

The results of Experiment 4 have implications for this debate: Data from selective-adaptation studies have been taken as evidence of online feedback, both when the adaptors were phonemes that had been replaced with signal-correlated noise (Samuel, 1997) and when they were ambiguous phonemes (Samuel, 2001). As Norris et al. (2003) argued, however, these effects could reflect lexically guided perceptual learning about the adaptors (i.e., adjustments to accept the adaptors as tokens of the phonemes they replaced) followed by selective adaptation in response to those retuned adaptors. Vroomen, van Linden, de Gelder, and Bertelson (2007) have since shown that adaptation after perceptual learning does occur in response to ambiguous audiovisual stimuli, and, in a reanalysis of the Samuel (2001) data, in response to ambiguous sounds in lexically biased contexts. Specifically, performance changed over the course of the Samuel (2001) experiment from an early lexical retuning effect (more ambiguous sounds identified in a lexically consistent fashion) to a subsequent adaptation effect in the other direction. No data were collected during the initial adaptation phase in the Samuel (1997) study, so a time-course analysis of that phase is impossible. Nevertheless, Experiment 4 shows that lexically guided retuning in response to signal-correlated noise stimuli (i.e., like those in Samuel, 1997) can occur rapidly. Thus, while the broader debate on feedback concerns more than just the selective adaptation data, Experiment 4 supports the learning-based account of the Samuel (1997) study that was proposed by Norris et al. (2003).

Our primary argument, however, concerns flexibility in speech perception. Research suggests that flexibility in mapping novel perceptual input onto existing categorical knowledge is not limited to the situation we used here, that is, where the lexicon provides the training signal for adjustments to novel speech sounds. Using an exposure–test paradigm like that of Norris et al. (2003); Cutler, McQueen, Butterfield, and Norris (2008) exposed listeners to an ambiguous [fs] sound in nonword contexts in which one or other interpretation of the ambiguous sound was phonotactically illegal (e.g., /s/ is illegal in [fs]rar, /f/ is illegal in [fs]narm). Although lexical knowledge could not be used to adjust perceptual categories during exposure, a retuning effect was found in the subsequent test phase. Perceivers can therefore use different types of prior knowledge (e.g., lexical and phonotactic knowledge) to retune perception. Indeed, knowledge-based retuning of perception appears to be a domain-general capability, extending beyond speech perception to lexically guided retuning of printed letter perception (Norris, Butterfield, McQueen, & Cutler, 2006) and to retuning based on prototypical color knowledge in color perception (Mitterer & de Ruiter, 2008).

Flexibility in speech perception, however, is not limited to knowledge-based retuning. The speech input itself can also provide a training signal to guide adaptations in the speech-perception system. For example, the audiovisual retuning effects found by Vroomen et al. (2007), and earlier by Bertelson, Vroomen, and de Gelder (2003), show that adaptations can arise on the basis of purely bottom-up, signal-driven information. Allen and Miller (2004) found that listeners can learn to identify talker-specific voice onset times (VOTs) in the stop consonants of multiple talkers, who, during exposure, demonstrated that they typically used different amounts of VOT. Similarly, Clayards, Tanenhaus, Aslin, and Jacobs (2008) showed that listeners have steeper categorization slopes when they listen to sounds from a bimodal VOT distribution with little variance within the categories than when they listen to sounds with a wide variance within the categories. Experiment 4 also demonstrates that perceptual adjustments can arise from the signal alone. Irrespective of the lexical bias in the exposure phase, listeners learned to interpret the noise stimulus as a token of /f/, based, as we have already argued, on the greater spectral similarity between [#] and [f] than between [#] and [s]. The input signal can thus invoke adaptations without reference to (or even in spite of) higher level information.

Lexically guided retuning therefore seems to be an instance of a much more general property of the speech-perception system. Flexibility, driven both by prior knowledge and by information in the signal, is a property the system requires to be able to extract stable phonological categories out of a perceptual input that varies considerably. In line with the experiment in McQueen, Cutler, and Norris (2006), we have shown that adaptations to phonemic categories generalized to words that did not occur during the training phase. Moreover, a restricted number of presentations of these sounds led to subsequent effectively normal activation of lexical candidates. When processing first-language input, this is a convenient state of affairs; only a limited number of instances of an ambiguous phoneme can cause adjustments that are applicable to the entire lexicon. These adjustments help processing by making it easier to recognize all words that contain the new sound.

For the second-language learner, however, there could be a cost associated with this flexibility. A substantial body of research has shown that first-language phonology causes problems for second-language listeners (e.g., Best, 1995; Best & Tyler, 2007; Flege, 1995; Kuhl & Iverson, 1995; Weber & Cutler, 2004). Our results suggest that these problems are not only because of the passive state of the prelexical system (i.e., due to the mere existence of native-language phonological categories) but also because of an active adjustment process that attempts to modify the boundaries of native-language categories so as to include second-language sounds in those categories. We have not tested whether [θ] is included in the /f/ and /s/ categories if the experiment is run in an English setting, nor do we know whether native speakers of English would include a good exemplar of [θ] in their /f/ or /s/ categories when given the proper training stimuli. However, the ease with which our Dutch participants learned to reinterpret English [θ], combined with the wealth of documented evidence that native phonological categories are strong attractors of sounds in second-language listening situations, suggests that our results have at least two implications for second-language listening. First, linguistic borders might not be aligned with the bounds of perceptual learning. That is, tokens of foreign sounds can easily come to be treated as members of native-language phonemic categories. Second, the flexibility that benefits first-language listening may come with a cost for second-language listening.

Conclusion

These experiments investigated the bounds on the flexibility of the perceptual system when dealing with variability in spoken language. The results argue for a process characterized by fast and thorough adaptations, based here on lexically guided retuning, restricted only by the spectral properties of an ambiguous sound. These findings provide a partial answer to the problem of invariance: Connections are rapidly made between new sounds in acoustic–phonetic space and the phoneme repertoire acquired in infancy. This type of learning assists speech decoding in daily life, where sounds and listening situations are hardly ever the same. There might be a disadvantage arising from this flexibility, however. The strength of the native phoneme repertoire, aided by fast and thorough perceptual flexibility, can turn into a nuisance when trying to acquire a second language. Further research could look into the effects that a learning task like Experiment 1 might have on the already vulnerable /θ/ category in our participants. For their sake, we hope that perceptual learning in fricatives is as speaker specific as previous experiments have suggested.

References

Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America, 115,* 3171–3183.

Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science, 14,* 592–597.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Baltimore: York Press.

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning. In honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins.

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

Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System, 36,* 22–34.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition, 108,* 804–809.

Cutler, A., McQueen, J. M., Butterfield, S., & Norris, D. (2008). Prelexically driven perceptual retuning of phoneme boundaries. In: J. Fletcher, D. Loakes, M. Wagner, & R. Goecke. (Eds.), *Proceedings of Interspeech 2008,* ISCA, Brisbane, Australia. September 2008.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America, 116,* 3668–3678.

Eisner, F. (2006). *Lexically guided perceptual learning in speech processing.* Doctoral dissertation, Radboud University Nijmegen (MPI Series in Psycholinguistics, 36).

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

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

Flege, J. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Baltimore: York Press.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America, 108,* 1252–1263.

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

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review, 13,* 262–268.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language, 56,* 1–15.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect." In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). Baltimore: York Press.

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

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences, 10,* 363–369.

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

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science, 30,* 1113–1126.

McQueen, J. M., & Mitterer, H. (2007, July). *Should you read the subtitles when watching a foreign film?* Presented at the joint meeting of the Experimental Psychology Society and the Psychonomic Society, Edinburgh.

McQueen, J. M., Norris, D., & Cutler, A. (2006). Are there really interactive speech processes in speech perception? *Trends in Cognitive Sciences, 10,* 533.

Mitterer, H., & de Ruiter, J. P. (2008). Recalibrating color categories using world-knowledge. *Psychological Science, 19,* 629–634.

Norris, D., Butterfield, S., McQueen, J. M., & Cutler, A. (2006). Lexically guided retuning of letter perception. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 59*(A), 1505–1515.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115,* 357–395.

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

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

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

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

Schroeder, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America, 44,* 1735–1736.

Stevens, M. (2007). *Perceptual adaptation to phonological differences between language varieties.* Unpublished doctoral dissertation, University of Ghent.

van Alphen, P. M., & McQueen, J. M. (2006). The effect of voice onset time differences on lexical access in Dutch. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 178–196.

van Linden, S., & Vroomen, J. (2007). Recalibration of phonetic categories by lipread speech versus lexical information. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 1483–1494.

Vroomen, J., Van Linden, S., de Gelder, B., & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory–visual speech perception: Contrasting build-up courses. *Neuropsychologia, 45,* 572–577.

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

Appendix

Critical Words Used in All Four Experiments

/f/- or /s/-Final Words Used in the Training Phase

/f/-final words. rif*, druif*, braaf, proef, lijf, witlof, aanhef, olijf, karaf, octaaf, achterneef, middenrif, onderlijf, ongeloof, biograaf, landbouwbedrijf, indicatief, locomotief, choreograaf, kwalitatief

/s/-final words. nis*, muis*, baas, roes, krijs, naaldbos, hakmes, radijs, karkas, relaas, pimpelmees, hagedis, paradijs, grandioos, geitekaas, ingenieus, anekdotisch, champagneglas, democratisch, problematisch

Minimal Word Pairs Used in the Testing Phase

doos/doof, bries/brief, hoes/hoef, les/lef, kuis/kuif, kas/kaf, los/lof, mus/muf, gaas/gaaf, gros/grof, roos/roof, lies/lief, kluis/kluif, poes/poef, bes/bef, gras/graf, bos/bof, wijs/wijf, hals/half, loos/loof

* Items not used in Pretests 2 and 3.

Received August 19, 2008
Revision received January 7, 2009
Accepted February 11, 2009 ■