Can listeners retune native categories across a phoneme boundary?

Michael D. Tyler & Mona M. Faris
Marcs Institute and School of Social Sciences and Psychology
University of Western Sydney, Sydney, Australia
m.tyler@uws.edu.au, m.faris@uws.edu.au

Abstract
Lexically guided perceptual retuning studies have demonstrated that listeners use their knowledge of phonemes in words to accommodate to artificially generated sounds that are halfway between two phonemes. However, it is unknown whether listeners accommodate in the same way when words are pronounced with an incorrect native phoneme (e.g., flower pronounced as “thlower”). Monolingual Australian-English listeners completed one of two exposure phases where the /θ/ or /s/ in words was pronounced as /θ/ (“th”), followed by a visual lexical decision task with cross-modal priming. If training is effective, identity-priming should be observed when a /θ/-bearing auditory prime (e.g., *thoil*) precedes a training-congruent matched visual target (e.g., *foil* or *soil*). Priming was observed for participants in the /θ/-training, but not the /s/-training condition. A second experiment with intact /θ/- or /s/-primes confirmed stimulus validity by showing an identity priming pattern. We conclude that lexically-guided perceptual retuning may be possible across a category boundary, but the native phonological system and/or acoustic similarity may impose limits on which native phonemes can be substituted effectively.

Index Terms: speech perception, lexically guided retuning, cross-modal priming, spoken word recognition.

1. Introduction
There is natural variability in spoken language, as a result of various inter-individual factors such as gender, age, and vocal tract shape, but despite these differences listeners are able to accurately detect the speaker’s intended utterance. One way in which listeners appear to attenuate to individual speaker idiosyncrasies is to use their knowledge of the phonemes that a word ought to contain to adjust phoneme category boundaries. This was demonstrated experimentally by Norris, McQueen, & Cutler [1], whose Dutch listeners completed an auditory lexical decision task, in which an ambiguous sound that was midway between /θ/ and /s/ was spliced into either /θ/- or /s/-final words. In a subsequent identification task using the /θ/-/s/ continuum from which the ambiguous sound was obtained, those in the /θ/-training condition were more likely to categorise ambiguous steps as /θ/, and those in the /s/-training condition were more likely to categorise them as /s/, suggesting that they had shifted their perceptual boundary according to the lexical context of the training. No learning effect was observed when participants were trained on non-words, which suggests that lexical knowledge is responsible for the phonemic adjustments. This type of learning has been termed lexically guided retuning [2], but the mechanism responsible for it is largely unknown. Norris et al. suggested that the existing categorical boundaries might expand to include the ambiguous phoneme.

To test the bounds of flexibility in lexically guided retuning, Sjerps and McQueen [3] tested whether accommodation is limited to sounds that lie on a continuum between two endpoints (e.g., /θ/ and /s/). They employed a voiceless dental fricative ([θ]) as the training sound, which was non-native for their Dutch participants. The training phase was the same as in [1], but /θ/ or /s/ was replaced with [θ] instead of an ambiguous [θs] sound. To test whether participants had been trained to perceive [θ] as identical to /θ/ or /s/, depending on the training condition, they employed a visual lexical decision task with cross-modal identity priming. Participants in the /θ/-training group showed priming for visual words containing /θ/ (e.g., reaction time to visual *doof* primed by *doil*) was faster than when preceded by the unrelated prime [krep]), but there was no priming for words containing /s/ (e.g., *doil*-[doof], suggesting that the participants had been successfully trained to perceive [θ] as /θ/. Those in /s/-training group showed the opposite pattern. Comparing the magnitude of priming to another experiment using unambiguous stimuli (e.g., [dof]-doof or [dos]-doos), Sjerps and McQueen concluded that [θ] was processed in the same way as the phoneme it replaced, although there was a slight bias toward interpreting [θ] as /θ/ rather than /s/.

It is clear from [3] that lexically guided retuning can result in expansion of a speech category into a phonetic region that is not simply in-between two native categories (e.g., between /θ/ or /s/). As [θ] is not a phoneme in Dutch, it is perhaps not surprising that the uncommitted phonetic space was available for remapping to /θ/ or /s/, depending on the training condition. One question that remains is whether the lexically guided retuning is possible when the training category is a different phoneme of the listener’s native language. That is, can categories expand across a phoneme boundary?

Here we investigated whether native Australian-English monolingual speakers could be trained to treat English /θ/ as /θ/ or /s/, in same way as the Dutch listeners of [3]. For the training phase we used a picture-verification task, adapted from [2], and for the test phase we created an English version of the cross-modal priming task of [3].

2. Experiment 1: /θ/ Training and /θ/ Priming

2.1. Method

2.1.1. Participants
Forty-eight monolingual Australian-English speaking students were recruited from Introductory Psychology at the University of Western Sydney. None of the participants had any known hearing or language impairments, and had normal or corrected-to-normal vision. The participants were randomly assigned to one of the two experimental conditions: 24 to the /θ/-training condition (17 females, $M_{age} = 22.08$ years, Age range: 18-37 years), and 24 to the /s/-training condition (21 females, $M_{age} = 23.25$ years, Age range: 18-39 years). Participants received course credit for participation.
2.1.2. Stimuli and Apparatus

The words for the training phase were taken from [2]. The /θ/-bearing versions of the critical items had been recorded for that study to serve as base stimuli for splicing in the ambiguous [θ]s tokens. For example, for the word basket, recordings were made for /bǝskǝt/, /bǝtǝkǝt/, and /bǝtǝkǝt/. The female Australian-English speaker who had recorded the items for the training phase was invited back to record the primes for the test phase. The stimuli were recorded in a sound booth at a 44.1 kHz sampling rate and 16-bit resolution using a Shure SM10A headset microphone, and a laptop computer with an EDDIOL UA-25 USB audio capture device.

The 86 word items in the picture verification task consisted of 6 practice words and 40 filler words, none of which contained /θ/, /θ/, or /s/, 20 items containing /θ/ (e.g., buffalo), and 20 items containing /s/ (e.g., basket). All of the participants were presented with the same words, but in the /θ/-training condition, the /θ/ in /θ/-bearing words was pronounced as /θ/ (e.g., bathalo). In each set of training items, the critical phoneme appeared in 10 word-initial positions, in seven word-medial positions, and in three word-final positions. Refer to [2] for further detail regarding the selection criteria for this set of stimuli. We used the same 174 black line drawings that were used in the training phase of [2].

The test phase consisted of 160 trials of an auditory prime followed by a visual word (80 trials) or nonword (80 trials) target, constructed in the same way as the Dutch stimuli of [3]. Word stimuli were selected using the CELEX database [4]. The 40 critical trials consisted of 20 /θ/-/s/- minimal pairs, chafe/chase, feed/seed, filly/silly, fit/sit, flap/slap, slate/sleet, flick/slick, flight/slight, fling/sling, flop/slop, flow/slow, foil/soil, folk/soak, lift/list, turf/terse, confine/consign, defeat/deceit, foggy/soggy, offend/ascend, and relief/release. The overall written frequency of the /θ/- and /s/-bearing words were 23 and 26 per million, respectively. The critical phoneme (/θ/ or /s/) appeared 13 times in word-initial position, four in word-medial position, and three in word-final position. The minimal pair words served as visual targets and were paired with either a phonologically related /θ/-prime (e.g., /θǝbǝl/, derived from foil), or with a phonologically unrelated prime (e.g., /θǝdǝl/ [grade]). The remaining 120 trials served as fillers. For the filler-word targets, 10 were preceded by identity-primes (e.g., /bǝutǝ/boat), 10 by phonologically related primes (e.g., /mǝgǝ/hug), and 20 by phonologically unrelated primes in which /θ/ replaced /θ/ or /s/, depending on the training condition (e.g., /θǝbǝl/ based on fail for /θ/-training and /θǝbǝ/ based on toast for /s/-training). The 80 nonword filler trials consisted of a visual nonword target preceded by an auditory word or nonword prime (see [3] for details). Only the minimal pair word trials were included in the analysis.

Each participant was presented with both the /θ/ and /s/-bearing members of each minimal pair. Half of each were preceded by a /θ/-prime and the other half by an unrelated prime. The primes and targets were counterbalanced across participants so that each prime-target pair was presented an equal number of times across the experiment. Each member of the minimal pairs was separated across two halves of the experiment such that one member was paired with a /θ/-prime (e.g., /θǝbǝl/-foil) in one half, and the other member with an unrelated prime (e.g., /θǝdǝl/-soil) in the other half. A pseudorandom presentation order was created following [3].

Stimulus presentation and response collection was controlled using DMDX software on PC laptops [5]. Visual stimuli were presented in 24-point Courier New font in the centre of the screen and auditory stimuli were presented over headphones that were connected to an Edirol UA-25EX external USB sound card.

2.1.3. Procedure

Participants were tested on individual computers in groups of up to three. For the picture-verification task, participants were told that they would see two pictures presented alongside each other on the screen, and they would simultaneously hear a spoken word through their headphones. They were asked to indicate as quickly and as accurately as possible which of those two pictures corresponded to the word they had heard by pressing one of two keys on the computer keyboard. The six practice trials, with feedback, were followed by two blocks of 80 trials, without feedback. Within each block, the target appeared an equal number of times on the left- and right-hand side of the screen. In the second block the same target-distractor pairs were repeated, in a different order, with the position of the visual target and distractor reversed. Stimuli were presented pseudorandomly, such that participants were never presented with more than four critical targets in a row, and the last five trials of the first block would be different to the first four trials of the second block. The presentation order was reversed for half of the participants.

In the subsequent lexical decision task with cross-modal priming, an auditory prime was presented through headphones, followed by a visual target on the screen at its acoustic offset. The target remained on the screen for 2 s, and was followed by a blank screen for 300 ms, after which responses were considered too late. Participants were asked to indicate as quickly and as accurately as possible whether the string of letters on the screen was an existing word in English by pressing one of two keys on a computer keyboard. Reaction time was calculated from the onset of the visual stimulus. After the test phase, participants completed a background information sheet.

2.2. Results and Discussion

Performance in the picture verification task was very close to ceiling in both training conditions, with mean accuracy for /θ/-bearing words, /s/-bearing words, and fillers ranging between 98% and 100%. For the test phase, inspection of the error rate data showed that many participants did not endorse certain minimal-pair items as words. We therefore removed from all further analyses target words with error rates higher than 40% in any of the Training (/θ=/θ/ vs. /θ=/s/) × Prime Type (/θ/-prime vs. unrelated conditions). Using this criterion, five target words were removed, and they are listed here with error rates from the four conditions in the following order: /θ=/θ/ training /θ/-prime, /θ=/s/ training unrelated prime, /θ=/s/ training /θ/-prime, /θ=/s/ training unrelated prime: chafe (50%, 83%, 83%, 75%), filly (58%, 67%, 83%, 75%), consign (17%, 42%, 33%, 50%), sleet (58%, 8%, 33%, 17%), terse (83%, 67%, 92%, 83%). With these items removed, the overall error rate was 4.23% (SE = 0.92%). There were no reliable effects of priming on the error data, so we report here only the results for reaction time (RT) to correct responses.

The priming effects for RT (based on the subject analysis) are presented in the left part of Figure 1. Note that the bars represent the difference in RT between unrelated primes and /θ/-primes. A positive priming effect means that RTs were shorter when the target was preceded by a /θ/-prime than an unrelated prime. It appears that when participants had been trained that /θ=/θ/, they were primed to respond to a related visual f-target (e.g., /θǝbǝl/ followed by foil), but not to a related visual s-target (e.g., soil). However, the reverse pattern was
not obtained for /θ/= /s/ training. An analysis of variance (ANOVA) revealed a significant three-way interaction between training condition (/θ/= /f/ vs. /θ/= /s/), target type (f-target vs. s-target), and prime type (related vs. unrelated) in both the subject analysis (denoted by \( F_1 \)) and the items analysis (denoted by \( F_2 \)):

- For the subject analysis, \( F_1 (1, 46) = 6.99, p = .011, \eta^2_p = .132; F_2 (1, 33) = 4.90, p = .034, \eta^2_p = .129.\) No other effects involving prime type were significant by both subjects and items. Thus, it appears that the difference in priming for f-targets versus s-targets is significantly greater in the /θ/= /f/ training condition than the /θ/= /s/ training condition.

### Results and Discussion

In Experiment 2, we used intact /f/ or /s/ primes instead of /θ/-bearing primes. That is, rather than /θ/- and /s/-training conditions, participants were assigned to /f/- and /s/-priming conditions (in terms of Figure 1). Following [3], we also conducted planned comparisons of related versus unrelated prime conditions (in terms of Figure 1), this would be testing whether the priming effect differs significantly from zero. The results of the analysis are presented in Table 1, which shows that the only reliable priming effect (i.e., significant by both subjects and items) was in the /θ/= /f/ training condition when a /θ/-prime was followed by an f-target. There were no reliable priming effects in the /θ/= /s/ training condition (the priming for f-targets was only significant by subjects, not by items).

To ensure that our minimal pairs would result in an identity priming effect when they are not mispronounced, in Experiment 2 we used intact /f/- or /s/- primes instead of /θ/-primes. That is, rather than /f/- and /s/-training conditions, participants were assigned to /f/- and /s/-priming conditions. Participants in the /f/-priming condition should be primed when the /f/-bearing prime precedes an identical f-target (e.g., /fæl/ followed by /foil/) but not when preceding a matched s-target (e.g., /fæl/ followed by /soil/). Conversely, for those in the /s/-priming condition, /s/-primes should prime s-targets (e.g., /soil/ followed by /soil/) but not f-targets (e.g., /soil/ followed by /foil/). Although there was no training component in Experiment 2, to keep the conditions as similar as possible across experiments they also completed a picture verification task using intact words.

### 3. Experiment 2: No Training and Intact Primes

#### 3.1. Method

Participants were 48 Australian-English monolinguals from the same population as Experiment 1. None of the participants reported any known hearing or language impairments, and had normal or corrected-to-normal vision. The participants were randomly assigned to one of two conditions: 24 to the /θ/-priming condition (19 females, \( M_{age} = 24.08 \) years, Age range: 17 years – 48 years), and 24 to the /s/-priming condition (21 females, \( M_{age} = 23.41 \) years, Age range: 18-48 years). Participants received course credit for participation.

The stimuli for the picture verification task of Experiment 2 were the same as those in Experiment 1 except that the auditory stimuli were unambiguous. For example, participants heard /baskat/ instead of /baθkat/. As all of the /θ/-bearing items of the Experiment 1 were converted to intact items, there was only one stimulus list for Experiment 2, presented in forward or reverse order, counterbalanced across participants.

The stimuli and the procedure for the test phase in Experiment 2 were similar to that in the previous experiment, except for the following deviations. Whereas participants in Experiment 1 were first trained that /θ/= /f/ or /θ/= /s/ and then tested with /θ/-bearing primes to gauge whether training had been successful, here the /θ/-prime was replaced with either an intact /f/-prime or an intact /s/-prime. Thus, the /θ/-priming condition in Experiment 2 is analogous to the /θ/= /f/ training condition of Experiment 1, and the /s/-priming condition is analogous to the /θ/= /s/ training condition.

#### 3.2. Results and Discussion

Performance in the picture verification task was at ceiling, just as it was in Experiment 1, with scores ranging from 98%-100%. To ensure comparability between experiments, the same five items were removed from the analysis of the test phase of Experiment 2 as were removed from Experiment 1. They are listed here with error rates from the four conditions in the following order: /θ/-prime, unrelated prime (/θ/-priming condition), /s/-prime, unrelated prime (/s/-priming condition): chafe (50%, 50%, 58%, 67%), filly (33%, 67%, 92%, 50%), consign (50%, 42%, 0%, 50%), sheet (50%, 42%, 25%, 17%), terse (67%, 58%, 67%, 58%). No other items had greater than 40% errors in any condition. With these items removed, the

---

**Table 1. Planned paired comparisons for priming effects (unrelated vs. related primes) for each training/priming condition by target in Experiments 1 and 2. Significant effects are highlighted in bold.**

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Condition</th>
<th>Target</th>
<th>Subject Analysis</th>
<th>Items Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/θ/= /f/ training → /θ/-prime</td>
<td>F-target</td>
<td>23</td>
<td>5.219</td>
</tr>
<tr>
<td></td>
<td>/θ/= /f/ training → /θ/-prime</td>
<td>S-target</td>
<td>23</td>
<td>1.298</td>
</tr>
<tr>
<td></td>
<td>/θ/= /s/ training → /θ/-prime</td>
<td>F-target</td>
<td>23</td>
<td>2.771</td>
</tr>
<tr>
<td></td>
<td>/θ/= /s/ training → /θ/-prime</td>
<td>S-target</td>
<td>23</td>
<td>1.102</td>
</tr>
<tr>
<td>Experiment 2: Intact</td>
<td>No training → /f/-prime</td>
<td>F-target</td>
<td>23</td>
<td>3.684</td>
</tr>
<tr>
<td></td>
<td>No training → /f/-prime</td>
<td>S-target</td>
<td>23</td>
<td>3.811</td>
</tr>
<tr>
<td></td>
<td>No training → /s/-prime</td>
<td>F-target</td>
<td>23</td>
<td>2.592</td>
</tr>
<tr>
<td></td>
<td>No training → /s/-prime</td>
<td>S-target</td>
<td>23</td>
<td>3.987</td>
</tr>
</tbody>
</table>
that participants in Experiment 1 showed priming when a /θ/ interaction was observed in both experiments, the patterns of identity were significant by both subjects and items. Planned comparisons to test priming effects are presented in Table 1, which showed a significant three-way interaction between prime identity (/θ/-prime vs. /s/-prime), target type (f-target vs. s-target), and prime type (related vs. unrelated) in both the subjects and items analyses: $F_1(1, 46) = 32.52, p < .001, \eta^2_p = .414; F_2(1, 33) = 31.20, p < .001, \eta^2_p = .516$. No other effects involving prime type were significant by both subjects and items. Planned comparisons to test priming effects are presented in Table 1, and they confirm that correct responses to identity-prime trials (i.e., /θ/-prime→f-target, /s/-prime→s-target) were significantly faster than correct responses to unrelated trials (i.e., unrelated-prime→f-target, unrelated prime→s-target).

To test whether the priming effects in Experiments 1 and 2 were significantly different, we ran an additional ANOVA combining both sets of data with Experiment as an additional factor. The four-way interaction between experiment, training condition/prime identity, target type (f-target vs. s-target), and prime type (related vs. unrelated) was significant by subjects and by items: $F_1(1, 92) = 8.41, p = .005, \eta^2_p = .084; F_2(1, 33) = 8.93, p = .005, \eta^2_p = .213$. Thus, while a three-way interaction was observed in both experiments, the patterns of priming are different, as can be seen in Figure 1. It appears that participants in Experiment 1 showed priming when a /θ/-prime preceded an f-target, but not an s-target, regardless of training.

4. Discussion

Previous research into lexically-guided retuning has shown that listeners can shift their category boundaries when presented with words containing an ambiguous mixture of phonemes (e.g., [fis], [1]) or a non-native phone (i.e., [θ] for Dutch listeners, [3]). The aim of the present study was to test whether lexically guided retuning can occur when the training sound is a different phoneme in the listener’s language. In Experiment 1, participants treated /θ/-primes as equivalent to words containing /θ/, but the training was apparently ineffective for words containing /s/. We ruled out the possibility that our stimulus items were responsible for this finding by running a second experiment with intact words containing /θ/ or /s/. Results of Experiment 2 showed a clear identity priming pattern for both /θ/ and /s/ words.

The design and stimuli of this study were based on [3], who demonstrated that Dutch listeners could be trained to perceive /θ/ as either /θ/ or /s/. As /θ/ is a phoneme in English but not Dutch, it is possible that our English monolinguals were prevented from retuning across the /θ/-/s/ boundary. Indeed, /θ/ and /θ/ are more acoustically similar than /θ/ and /s/ [6], and that may have facilitated the /θ/-/θ/ retuning for English monolinguals. /θ/ and /θ/ are similar for Dutch monolinguals too, but in spite of that Dutch listeners tend to assimilate /θ/ as /s/, and rate /θ/ as more similar to /θ/ than English listeners do [7]. Another difference between our approach and [3] is that they spliced the same distinct [θ] token into /θ/- or /s/-bearing words whereas our stimuli retained natural phonetic variability. Retuning might be observable for /s/ if a single acoustic token is used for training.

An additional possibility, that cannot be ruled out from our data, is that English monolinguals have a natural bias to perceive /θ/ as /θ/ in our /θ/-primes. In that case, the /θ/-training would have served to reinforce the bias, and the /s/-training may not have been sufficient to turn that bias towards /s/. The finding that /θ/-primed f-targets in the /s/-training condition, by subjects but not items, suggests that this possibility is worth investigating. Future research should, therefore, establish a baseline level of priming /θ/ by running the test phase of Experiment 1 without any prior training. Nevertheless, it should be noted that if a /θ/→/θ/ bias exists, the /s/-training appears to have been effective enough neutralize the bias, as supported by a significant three-way interaction between training, target type, and prime type.

In conclusion, it is clear from this study that participants in Experiment 1 showed strong priming when f-targets were preceded by a prime in which the /θ/ had been pronounced as /θ/, but no reliable priming for s-targets. It is therefore possible that lexically-guided learning can occur across a phoneme boundary, but the native phonological system and/or acoustic similarity may impose limits on which pairs of phonemes can be returned in response to training.

5. Acknowledgements

This study was supported by an Australian Research Council Discovery Grant to the first author (DP0880913). We thank Sarah Fenwick and Kylie Tyler for their research assistance.

6. References