# Hearing Beyond Categories: General Adaptation to Nonnative Speech

**Yuting Gu (yuting.gu@uci.edu)**
Department of Language Science, University of California, Irvine

**Michelle Chao (michellemhc.008@gmail.com)**
Department of Language Science, University of California, Irvine

**Xin Xie (xxie14@uci.edu)**
Department of Language Science, University of California, Irvine

**Abstract**

Listeners can rapidly adapt to non-native accented speech, yet the underlying mechanisms remain debated. This study examines whether accent adaptation reflects adjustments in phonetic representations or shifts in decision-making processes. Using a pretest-exposure-posttest paradigm, we examined native English listeners' perception of the Mandarin-accented /θ/-/s/ contrast across two exposure conditions: exposure to Mandarin-accented sentences (Experiment 1) or to pure tones (Experiment 2). In both experiments, listeners showed increased acceptance of ambiguous /θ/ and /s/ tokens when they formed real words, suggesting that adaptation stems from changes in lexical decision criterion reinforced through task repetition rather than accent exposure alone. Additionally, we observed evidence suggestive of rapid within-test distributional learning from limited trials. Our findings support the notion that listeners lower lexical decision criterion when processing accented speech, while also demonstrating remarkable adaptability to novel accent features even with minimal exposure.

**Keywords:** non-native speech; phonetic variation; speech perception; accent adaptation; lexical decision criterion

## Introduction

While speech from talkers with unfamiliar accents initially presents comprehension challenges, listeners demonstrate remarkable perceptual flexibility, rapidly improving their recognition of unfamiliar pronunciations with exposure. (Bradlow & Bent, 2008; Eisner & McQueen, 2005; Norris, McQueen, & Cutler, 2003; Reinisch & Holt, 2014; Samuel & Kraljic, 2009; Xie et al., 2018; Zheng & Samuel, 2020).

How do listeners adapt to unfamiliar and unexpected pronunciations? One approach to this question is by examining whether adaptation to specific sound categories leads to improved recognition in response to recent exposure. In a seminal study by Norris et al., (2003), listeners were exposed to an ambiguous fricative midway between /f/ and /s/ in lexical contexts where /f/ was expected but /s/ was not (e.g., [lea?], where *leaf* is a word, but *leas* is not; the original study was conducted in Dutch). Listeners exposed to the ambiguous sound in /f/-final words were subsequently more likely to classify other ambiguous sounds along the /f-s/ continuum as /f/, while those exposed to the ambiguous fricative in /s/-final words tend to classify the ambiguous sounds as /s/, appearing to have 'recalibrated' their phonetic categories according to the specific exposure. Numerous studies have built on this paradigm of 'phonetic recalibration,' demonstrating flexible adjustments to noncanonical pronunciations across different phoneme classes (e.g., stops, fricatives, vowels). Such flexibility is observed among speakers of different languages (e.g., Reinisch, Weber, & Mitterer, 2013), including monolinguals and bilinguals (e.g., Caudrelier et al., 2023), and across the life span (e.g., Gordon-Salant et al., 2010).

Despite the rich evidence supporting category-specific adaptation, the exact mechanism behind adaptation remains under debate. Critically, most studies have used crafted stimuli designed to maximize perceptual ambiguity, designed to enhance the power to detect changes between participants who receive the same acoustic manipulations (e.g., ambiguous /f-s/ morphed from clear /f/ and /s/) in different contexts (e.g., Mitterer, Scharenborg, & McQueen, 2013; Norris et al., 2003; Samuel & Kraljic, 2009). One drawback of this approach is that it distorts natural variation in accent features, reducing the generalizability of findings based on 'artificial accents.' In a natural nonnative accent, the same sound can be pronounced in drastically different ways, compared to native variants. These differences could involve a shift in phonetic cue distributions (e.g., Flege, Schirru, & MacKay, 2003), increased variability along one or more acoustic cues (e.g., Flege, Munro, & Skelton, 1992), or changes in the relative reliance of multiple acoustic cues (e.g., Flege, Bohn, & Jang, 1997). These distinct types of phonetic variation are predicted to directly influence whether and how listeners could adapt to them, according to theories of learning (Bent & Baese-Berk, 2021; Kleinschmidt & Jaeger, 2015; Xie, Jaeger, & Kurumada, 2023). Yet the morphing procedure used to create ambiguous tokens between the original productions of two sounds often eliminate accent-specific features. For instance, while several acoustic cues (e.g., frication duration, formant transitions) distinguish between fricatives across place of articulation and sibilance in natural speech (e.g., /θ/ vs /s/), these cues are typically rendered ambiguous in morphed stimuli (e.g., Melguy & Johnson, 2022). This leaves listeners with little information to learn about the accent-specific input statistics.

Importantly, whatever type of learning underlies the observed boundary shifts in perceptual recalibration paradigms, it appears to be highly constrained. For instance, adaptation to an altered /b/-/d/ contrast in an *aba-ada* context does not readily transfer to the same contrast in an *ibi-idi*



In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

context (Reinisch et al., 2014). Similarly, studies have failed to find generalization across allophonic variants of the same phoneme category (Mitterer, Reinisch, & McQueen, 2018; Mitterer, Scharenborg, & McQueen, 2013). A closer examination of the acoustic details of adapted and generalized stimuli suggests that much of category-specific adaptation, as studied in perceptual recalibration paradigms, has an acoustic locus. For instance, Melguy and Johnson (2022) exposed listeners to artificially constructed ambiguous /θ/ tokens, midway between [θ] and [s]. They found that exposure to these tokens not only increased listeners' likelihood of categorizing sounds along the /θ-s/ continuum as /θ/ but also resulted in a similar shift in /θ-ʃ/ categorization. In contrast, there was no change in how tokens along a /θ-f/ continuum were categorized. They concluded that while perceptual recalibration does not have to be contrast-specific; rather, generalization to neighboring categories is conditioned on acoustic similarity, which is greater between /θ-ʃ/ and /θ-s/ than between /θ-f/ and /θ-s/. Relatedly, cross-talker generalization has also been found to be constrained by acoustic similarity (Lai & Tamminga, 2024; Reinisch & Holt, 2014). Taken together, these findings suggest learning is constrained by the degree of acoustic overlap between the phonetic variants encountered during exposure and test in perceptual recalibration studies.

Such highly specific pattern of learning stands in contrast with findings from work on natural accents, where listeners demonstrate greater flexibility in how they map sounds onto words, beyond the specific sounds experienced during exposure. For example, after exposure to an accented talker, listeners may accept *vaby* as *baby*, even if they have never heard the /v/ sound from that speaker (Schmale, Cristia, & Seidl, 2012). Similar effects have been found in vowel adaptation, where hearing lowered back vowels increases listeners' acceptance of previously unheard raised vowels as members of the same category (Weatherholtz, 2015).

These differences between natural accents and artificially constructed accented speech raise the question of whether the perceptual adjustments involved adapting to them are driven by the same mechanisms, and whether category-specific adjustments as discussed above can explain general adaptation to an accent (Bradlow, Bassard, & Paller, 2023; Bradlow & Bent, 2008; Sidaras, Alexander, & Nygaard, 2019). To investigate this, Zheng and Samuel (2020; henceforth ZS20) examined how the same group of native listeners adapt to both ambiguous speech artificially constructed from native-accented speech and natural nonnative-accented speech, focusing on the fricative contrast /θ/-/s/ in English. Because dental fricatives are absent in Mandarin Chinese, Mandarin speakers often struggle to produce /θ/ in English, with some speakers pronouncing /θ/ and /s/ with great overlap phonetically and some replacing /θ/ with the alveolar fricatives /s/ phonologically (Han, 2013; Zhang & Xiao, 2014). As a result, an intended word could be perceived as a nonword (e.g., *thief* pronounced as *sief*).

In the perceptual recalibration experiment of ZS20, different groups of listeners were exposed either to ambiguous /θ/ sounds (e.g., /θ/ in *thankful*) or to ambiguous /s/ sounds (e.g., /s/ in *episode*) in lexically disambiguating contexts; ambiguous tokens were created using the morphing procedure as aforementioned. The group who heard ambiguous /θ/s during exposure categorized more sounds along the /θ-s/ continuum as /θ/ and vice versa for the other group, demonstrating a shift in categorization boundary. In contrast, the accent accommodation experiment exposed listeners to naturally produced Mandarin-accented English sentences containing /θ/ and /s/ tokens. After talker-specific exposure, listeners increased their proportion of /θ/ responses not only for Mandarin-accented /θ/-words but also, crucially, for /s/-nonwords where /θ/ had been substituted with /s/ (e.g., accepting *sief* as *thief*).

ZS20 observed no reliable within-individual correlation between adaptation to an artificially constructed accent (derived from native English speech) and adaptation to a natural nonnative accent (Mandarin-accented English) for the same /θ/-/s/ contrast. Based on these findings, the authors concluded that phonemic boundary recalibration is not the primary mechanism underlying natural accent adaptation. Instead, they attributed the mechanism to a "general relaxation" strategy, whereby listeners become more tolerant of phonetic deviations from the accented speaker in word recognition. However, it is not clear whether this relaxation is due to changes in phonetic category representations or rather, it reflects a change in lexical decision criterion as both possibilities are compatible with an increase in /θ/ responses in the lexical decision task.

One possibility is that relaxation reflects changes in lexical decision criterion—the threshold at which bottom-up acoustic phonetic input is accepted as a real word. In a lexical decision task, listeners must decide whether a given auditory signal corresponds to a real word. When the speech signal is ambiguous and thus activates multiple phonetic categories (e.g., /θ/ and /s/), it may cascade to partially activate multiple lexical candidates. Whether a listener responds 'yes' depends on both the strength of the lexical activation forwarded from the bottom-up acoustic phonetic input as well as listeners' decision threshold for endorsing a lexical candidate as a real word (Ratcliff & McKoon, 2008; Wagenmakers et al., 2008). Crucially, even if bottom-up activation remains unchanged, a lowered decision threshold (i.e., a more generous decision criterion) can increase lexical endorsement rates. This means that even if a /θ/ token remains as perceptually ambiguous as it was before exposure, listeners may modify their decision threshold and become more likely to identify the ambiguous sound as /θ/, as long as doing so results in a real word.

An alternative possibility is that accent exposure actually changes listeners' underlying category representations via distributional learning of the acoustic phonetic input (Kleinschmidt & Jaeger, 2015; Theodore & Monto, 2019). In this view, listeners track the statistical distribution of relevant cues in the nonnative input and update their phonetic representations of the categories accordingly. Evidence for such mechanisms comes from prior work on nonnative accent adaptation (e.g., Idemaru & Holt, 2011; Tan, Xie, & Jaeger, 2021; Xie, Theodore, & Myers, 2017). For instance, exposure to Mandarin-accented English tokens allow native listeners to adjust both the boundary and internal structure of the exposure category. Similarly, Melguy and Johnson (2022) found that listeners expanded category boundaries into neighboring phonetic space when the cue distributions of the exposure and test stimuli overlapped in the acoustic space.

ZS20 reported that the acoustic cue distributions along multiple dimensions (e.g., frication duration, frication amplitude) were almost entirely overlapped between /θ/ and /s/ in their accent accommodation experiment, unlike what is expected for the native-English contrast (McMurray & Jongman, 2011). If listeners engaged in distributional learning, then after hearing many ambiguous /θ/ tokens and some /s/ tokens during exposure, listeners' representations of the two categories will be broadened and shifted toward the contrastive category but with a greater change for /θ/. This will result in an overall shift in phonetic categorization, resulting in more /θ/ responses; when this change cascades to the lexical level, lexical items such as *thief* are more likely to be called words, while items such as *sief* are less likely to be called nonwords. Unlike a shift in lexical decision criterion, such representational changes would impact both lexical decision and phonetic categorization tasks, as they reflect a restructuring of the phonetic categories per se.

In sum, both levels of changes—lexical decision criterion shifts and changes in phonetic representation—could explain the pattern observed in ZS20.

## The Current Study

To distinguish between these possibilities, we conducted two experiments examining how native-English listeners adapt to Mandarin-accented /θ/. We followed the original design of ZS20 but modified it in two aspects. First, whereas they exposed participants to /θ/-words (e.g., *thief*) and /s/-nonwords (e.g., *ausentic*) during the test phase—creating a */θ/-biased condition* where both lexical contexts bias listeners towards a /θ/ interpretation, we introduce a new */s/-biased* condition where participants hear /s/-words (e.g., *sudden*) and /θ/-nonwords (e.g., *prothpect*). This new design allows us to tease apart predictions made by alternative hypotheses. These hypotheses differ in whether they attribute adaptation to changes in lexical decision criterion (H1) or to changes in phonetic category representations (H2), and whether such changes are category-specific or general across categories.

**H1: Changes in lexical decision criterion.** Under this hypothesis, exposure to accented tokens in real-word contexts lowers the decision threshold for lexical responses, leading to more tokens—especially ambiguous ones from the talker—being accepted as words. Therefore, it is predicted that in the */θ/-biased condition*, /θ/ responses should increase for both /θ/-words and /s/-nonwords, a condition where interpreting ambiguous sounds as /θ/ forms real words. In the */s/-biased condition*, the same criterion shift would manifest as an increase in /s/ responses—therefore a decrease in /θ/ responses—for both /s/-words and /θ/-nonwords. Of note, it is possible that the change in lexical decision criterion is category-specific and does not indiscriminately apply to all lexical items. Given that participants hear more /θ/-words and /s/-words, if the adjustment is sensitive to this statistic, then we expect greater changes in the */θ/-biased condition*; crucially though, any changes in the */s/-biased condition* will be in the opposite direction, regardless of its magnitude.

**H2: Changes in phonetic category representations.** As aforementioned, this hypothesis proposes that exposure to accented /θ/ tokens will lead to an overall shift in phonetic categorization resulting in more /θ/ responses overall. Crucially, given that the exposure phase is identical in the */θ/-biased* and */s/-biased conditions*, the same pattern is expected across conditions. This leads to increased /θ/ responses not only for words and nonwords in the */θ/-biased condition*, but also for those in the */s/-biased condition*. Therefore, comparing the patterns between the two conditions allows us to distinguish between H1 and H2.

Second, we create Experiment 2 to serve as a strict control for Experiment 1. In Experiment 2, the pretest and posttest remain identical to those in Experiment 1, but the accent exposure phase is replaced with pure tone exposure which is similar in duration but contains no linguistic content. This allows us to determine whether perceptual changes from pretest to posttest are truly driven by the accent exposure. Specifically, if accent exposure in Experiment 1 were required to elicit changes in the perception of /θ/ and /s/ from pretest to posttest, we would not expect the same effect in Experiment 2. However, if both experiments yield similar results, it raises the question of whether it is accent exposure—or some other factor—that drives the changes in listeners' responses.

## Methods

### Participants

Two independent groups of participants aged between 18 and 35 were recruited from Prolific ($N$ = 77 and $N$ = 75 for experiment 1 and 2 respectively). All participants were monolingual American English speakers with normal hearing and vision and reported no regular exposure to Mandarin or Mandarin-accented English. One participant from Experiment 1 was excluded due to a below-chance performance during exposure (<50%). In both experiments, participants were randomly assigned to one of the two test conditions: */θ/-biased* vs. */s/-biased* (41 vs. 35 for Experiment 1; 40 vs. 35 for Experiment 2).

### Materials

All stimuli were produced by a female native speaker of Chinese whose /θ/ productions in English were ambiguous and exhibit an intermediate level of intelligibility (as judged by native-English speaking research assistants). Ambiguity was established by acoustic measurements (e.g., frication duration, four spectral moments), which confirmed a substantial overlap between the production distributions of /θ/ and /s/.

**Accent Exposure Stimuli (Experiment 1)** There were 96 semantically predictable sentences (e.g., *The dishcloth is very wet*), organized into 32 trials, with each trial presenting three auditory sentences followed by a visual probe. In half the trials, the probe was a reworded version of one of the three auditory sentences (e.g., '*Trees become different colors at Thanksgiving*' for '*Leaves change color at Thanksgiving*'); in the other half, the probe was unrelated. Trial order was randomized across participants. Across all 96 sentences, 48 contained the critical /θ/ phoneme, typically pronounced in an /s/-like manner. Due to the high frequency of /s/ in English words, eliminating its presence is impractical. We minimized the number of /s/ tokens (e.g., '*sweet*' in '*The orange was very sweet*') to 29.

**Tone Control Stimuli (Experiment 2)** Participants in Experiment 2 were exposed to pure tone sequences adapted from Xie and Myers (2015). Stimuli included 40 local tone pairs (20 'same,' 20 'different') and 40 global tone pairs (20 'same,' 20 'different'), resulting in 80 trials, with each pair played in a single trial. Each pure tone sequence consists of six pure tones (250ms duration, 20 ms interval between each tone within a sequence), presented at an interval of 1000 ms between sequences. In the local tone perception task, participants judged whether two sequences contained the exact same notes or differed by a single note (i.e., a pure tone with a specific fundamental frequency, used as a discrete unit within a pure tone sequence). In the global tone perception task, participants judged whether two sequences shared the same overall pitch contour, with one sequence elevated by a fixed interval relative to the other. All sequence pairs are randomized within each task. Pure tone exposure was chosen as a control because it engages general auditory discrimination without invoking phonetic or linguistic processing, thus isolating the effects of accent exposure. Additionally, the duration of pure tone exposure was matched to the duration of the accent exposure task, which ensured that differences in adaptation could not be attributed to the unequal length of exposure.

**Pretest And Posttest Stimuli (Both Experiments)** There were 40 critical test items, 20 filler words and 20 filler nonwords in each phase. Two test conditions were created: In the */θ/-biased condition*, there were 20 critical /θ/-words (e.g., *thief*) and 20 /s/-nonwords (e.g., *sief*). In the */s/-biased condition*, there were 20 /s/-words (e.g., *sudden*) and 20 /θ/-nonwords (e.g., *thudden*). Filler words and nonwords were identical for all participants. Filler nonwords were created by altering one phoneme in a filler word with a phoneme that is not challenging for Mandarin-accented English speakers to produce (e.g., '*program*' changed to '*probram*'). Stimuli were divided into two counterbalanced lists (A and B) such that half of the items (20 critical words, 20 critical nonwords, and 40 filler words) appear in the pretest, and the other half in the posttest. Syllable length (/θ/: M = 2.6, SD = .89; /s/: M = 2.55, SD = .71) and lexical frequency (/θ/: M = 8.76, SD = 1.30; /s/: M = 8.39, SD = 1.46) were comparable between the two lists, based on the English Lexicon Project (Balota et al., 2007).

## Procedure

Participants were screened using a headphone check to ensure proper audio setup (Milne et al., 2020). Both experiments follow a 3-block design (pretest-exposure-posttest). During the pretest and posttest, participants were randomly assigned to one of the two test conditions (*/θ/-biased* vs. */s/-biased*) and completed an auditory lexical decision task. At the offset of each word, participants pressed designated keys to indicate whether the auditory stimulus was a real word or a nonword.

The exposure phase differed between the two experiments. In Experiment 1, participants heard three Mandarin-accented auditory sentences per trial and judged whether a visually presented probe sentence was a rewording of one of the previously heard sentences. In Experiment 2, participants completed two tone perception tasks (local and global tone perception) and judged whether tone sequences were the same or different. Task order was counterbalanced across participants.

# Results

## Experiment 1

Overall recognition accuracy during the accent exposure phase is reasonably high (*/θ/-biased condition*: M =.85, SD = .09; */s/-biased condition*: M = .81, SD = .10), suggesting that listeners were comprehending the spoken sentences as instructed. Recognition accuracy during the test phase is presented in Table 1.

Table 1: Lexical decision accuracy across test phases by condition and word type for Experiments 1 and 2

| Experiment | Condition | Word Type | Pretest (M, SD) | Posttest (M, SD) |
|---|---|---|---|---|
| Exp 1 | /θ/-biased | /θ/-words | .79 (.11) | .87 (.11) |
| | /θ/-biased | /s/-nonwords | .24 (.10) | .16 (.11) |
| | /s/-biased | /s/-words | .85 (.11) | .87 (.10) |
| | /s/-biased | /θ/-nonwords | .28 (.10) | .17 (.12) |
| Exp 2 | /θ/-biased | /θ/-words | .78 (.13) | .85 (.14) |
| | /θ/-biased | /s/-nonwords | .23 (.15) | .18 (.13) |
| | /s/-biased | /s/-words | .81 (.15) | .87 (.14) |
| | /s/-biased | /θ/-nonwords | .25 (.12) | .19 (.15) |

To facilitate the comparison of predictions from H1 and H2, we examined the proportion of /θ/ responses during the test phase (Figure 1). To investigate how this dependent measure has changed as a function of accent exposure, we fit a generalized mixed-effects linear regression model (GLMM) using the statistical software R (v4.4.1; R Core Team, 2024) with the packages *lme4* (v1.1.35.5; Bates et al., 2015) and *lmerTest* (v3.1.3; Kuznetsova et al., 2017). The model's fixed effects (all deviation coded) include test phase (pretest = -.5, posttest = .5), sound presented (e.g., '*authentic*' vs. '*ausentic*'; /θ/ = -.5, /s/ = .5), test condition (/θ/-biased = -.5, /s/-biased = .5), and their interactions. We start with the most comprehensive random effect structure supported by our design and remove higher-order interactions stepwise if convergence issues arose. Random effects include by-participant intercepts, and by-item intercepts and slopes for condition.

We find significant main effects for all three fixed effects: test phase, sound, and condition. The main effect of condition ($\widehat{\beta}$ = -3.43, *SE* = .41, *z* = -8.30, $p < .001$) suggests that participants in the */θ/-biased condition* show a higher proportion of /θ/ responses compared to those in the */s/-biased condition*, as expected from a lexical bias. There is a main effect of sound ($\widehat{\beta}$ = -.33, *SE* = .15, *z* = -2.27, $p < .001$), suggesting that listeners could discriminate between /θ/ and /s/ to some extent. There is also a main effect of test phase ($\hat{\beta}$ = .79, *SE* = .16, *z* = 4.83, $p < .001$). Importantly, we find significant interactions between condition and test phase ($\hat{\beta}$ = -1.64, *SE* = .23, *z* = -7.29, $p < .001$), and between condition and sound ($\widehat{\beta}$ = -.66, *SE* = .21, *z* = -3.10, $p < .05$). Follow-up simple effects analysis shows that, in the */θ/-biased condition*, there is a significant increase in /θ/ responses both for /θ/-words ($\widehat{\beta}$ = .79, *SE* = .16, *z* = 4.84, $p < .001$) and for /s/-nonwords ($\hat{\beta}$ = .84, *SE* = .15, *z* = 5.44, $p < .001$), replicating the findings of ZS20. In contrast, in the */s/-biased condition*, there is a significant decrease in /θ/ responses for both /θ/-nonwords ($\widehat{\beta}$ = -.85, *SE* = .16, *z* = -5.49, $p < .001$) and /s/-words ($\hat{\beta}$ = -.35, *SE* = .18, *z* = -1.97, $p < .05$). These patterns align with our hypothesis H1, suggesting that a change in lexical decision criterion may be at play.

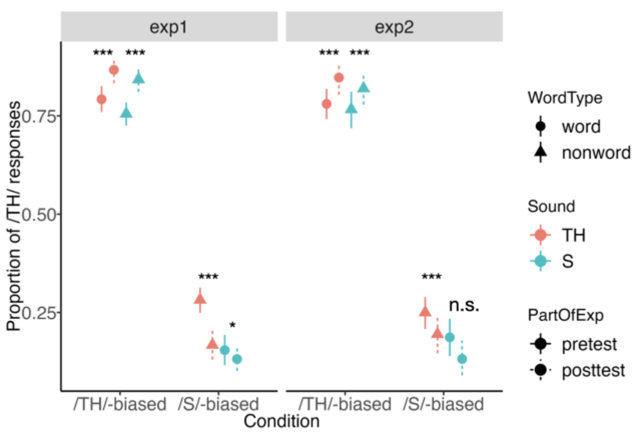


Figure 1: Proportion of /θ/ responses during the test phase across conditions for Experiments 1 and 2.

## Experiment 2

During the pure tone exposure phase, recognition accuracy was generally higher for local pure tone exposure compared to global pure tone exposure. In the */θ/-biased condition*, participants showed a mean accuracy of .73 (SD = .15) for local pure tones, which decreased to .54 (SD = .12) for global pure tones. Similarly, in the */s/-biased condition*, local tone accuracy was .78 (SD = .13), dropping to .52 (SD = .11) for global tone exposure. Recognition accuracy during the test phase matched that in Experiment 1 (see Table 1).

For the test phase, the regression model specifications are similar to those in Experiment 1, except that we simplify the random effects structure to include only by-participant and by-item intercepts.

We find a main effect of condition ($\widehat{\beta}$ = -3.32, *SE* = .42, *z* = -7.87, $p < .001$). Surprisingly, there is no main effect of sound ($\hat{\beta}$ = -.12, *SE* = .14, *z* = -.86, *p* = .39), suggesting that participants do not reliably distinguish between /θ/ and /s/, unlike participants in Experiment 1. In addition to the two-way interactions observed in Experiment 1 (condition X test phase; condition X sound), there is a significant three-way interaction between condition, sound, and test phase ($\widehat{\beta}$ = .77, *SE* = .33, *z* = 2.31, $p < .05$). Simple-effects analysis show a significant increase in /θ/ responses in the */θ/-biased condition* for both /θ/-words ($\widehat{\beta}$ = .64, *SE* = .16 , *z* = 4.11, $p < .001$) and /s/-nonwords ($\hat{\beta}$ = .51, *SE* = .15, *z* = 3.37, $p < .001$), replicating the findings of ZS20 and aligning with the results from Experiment 1. However, in the */s/-biased condition*, the pattern was asymmetrical: Specifically, we find a significant decrease in /θ/ responses for /θ/-nonwords ($\widehat{\beta}$ = -.85, *SE* = .17, *z* = -5.08, $p < .001$), but no such significant change for /s/-words ($\widehat{\beta}$ = -.21, *SE* = .18 , *z* = -1.14, *p* = n.s.).

### Changes During Pretest And Posttest

The results from Experiment 2 were surprising. If perceptual changes were driven solely by exposure to accented speech, we would have expected no shift in lexical endorsement in Experiment 2. Why, then, did listeners show signs of adaptation even in the absence of explicit accent exposure? It is unlikely that a mere elapse of time or exposure to any auditory input could alter the way listeners make lexical decisions. A more plausible explanation is that repeatedly making lexical decisions about nonnative-accented stimuli during pretest was sufficient to induce change. This interpretation is supported by a closer examination of pretest and posttest responses (see Figure 2).

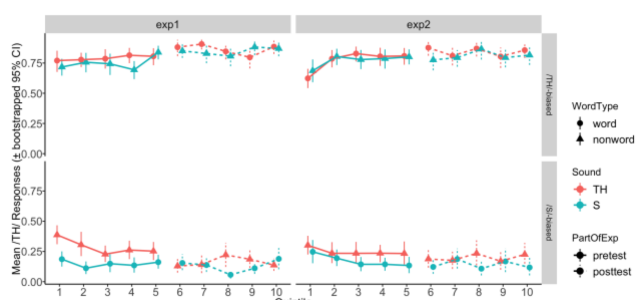


Figure 2: Proportion of /θ/ responses across pretest and posttest trials by word type, sound category, and condition (Experiments 1 and 2). The data is the same as that in Figure 1 but now plotted by quintile (16 critical trials per bin) within each test phase.

Specifically, there were noticeable changes in the proportion of /θ/ responses within the first quintile of the pretest; this early change was also present in Experiment 2, though the pattern was different between experiments. This means with as few as 8 critical trials, listeners had already begun to change their lexical decision responses, in a way consistent with what we observed for the change from pretest to posttest.

While this observation helps explain the lack of difference between Experiment 1 and Experiment 2, it raises an important question: what, then, led to the changes in the pretest? One possibility is that upon hearing a nonnative-accented voice, listeners rapidly adopted a more liberal lexical decision criterion, accepting more acoustic tokens as real words despite that the perceptual evidence in support of such a decision was weaker than expected for native-accented speech. The greater changes within quintiles 1-3, relative to the remainder of the pretest and the posttest phase, are compatible with this idea.

Another possibility is that the actual acoustic distributions of accented /θ/ and /s/ sounds presented in pretest elicited changes in category representations. Consider the */θ/-biased condition*: in an extreme scenario where listeners rely entirely on lexical knowledge to make lexical decisions, any instance containing /θ/ (e.g., *thief*) or /s/ (e.g., *sief*) would be recognized as '*thief*.' Under this scenario, the pretest in our experiment would function similarly to the exposure phase typically employed in perceptual recalibration studies (Norris et al., 2003), where the lexical biasing context serves as a teaching signal for recalibrating underlying categories. If so, the /s/-biased pretest would elicit an opposite shift towards /s/, which aligned with our observations. Of note, in traditional perceptual recalibration studies, listeners hear clear tokens from the contrastive category—hearing ambiguous *thief* and clear *sudden*, for example. This setup arguably makes it more likely for listeners to interpret all ambiguous tokens as /θ/ *in* real words rather than as /s/ in nonwords, compared to our experiment. That said, the rapid changes we observed are reminiscent of previous findings using artificially crafted speech, where lexically guided learning effects was detected after just four critical exposures and increased monotonically with more exposure (Cummings & Theodore, 2023).

## Discussion

The current study investigates whether hearing nonconical pronunciations in a natural nonnative accent elicits adaptation—and if so, through which mechanisms. Specifically, we asked whether exposure to Mandarin-accented English words containing /θ/ alters how native-English listeners perceive both categories in the /θ/-/s/ contrast in service of lexical decision.

Replicating Zheng & Samuel (2020), participants in the */θ/-biased condition* showed increased /θ/ responses when doing so rendered the auditory stimulus a real word (e.g., interpreting both *thief* and *sief* as *thief*). Critically, participants in the */s/-biased condition* showed the opposite pattern. In other words, listeners in both conditions demonstrated increased word recognition accuracy and decreased nonword recognition accuracy for the sound they were biased toward from pretest to posttest. These results support Hypothesis H1, which posits a change in lexical decision criterion—consistent with the "general relaxation" account of accent adaptation. If the observed changes were due to category-specific lexical criterion shift, we would expect increased /θ/ responses only for /θ/-words and /s/-nonwords in the */θ/-biased condition*, but no change in the */s/-biased condition*. Notably, this explanation of a general shift in lexical decision criterion readily accounts for the surprising changes between pretest and posttest in Experiment 2, despite the absence of accent exposure.

However, we cannot rule out the possibility that listeners' category representations were altered as well. Specifically, challenging an implicit assumption made by ZS20 and our own design, listeners clearly changed their responses to accented words and nonwords even during pretest. If changes in category representations indeed occurred as in lexically guided perceptual learning, it could account for the opposite pattern observed in the /θ/-biased and /s/-biased conditions. However, it is not clear why hearing a large amount of /θ/ tokens embedded in semantically predictable sentence during the exposure phase did not trigger enough relearning that ultimately resulted in an overall increase in /θ/ responses in the /s/-biased condition.

Our findings have several implications. First, they call for a reconsideration of the assumption that perceptual adaptation arises exclusively from dedicated exposure phases. Future studies should distinguish between adaptation driven by explicit accent exposure and adaptation that emerges from task repetition or stimulus-driven perceptual learning. Modifying experimental designs to isolate these effects could help disentangle the respective contributions.

Second, future research could consider incorporating phonetic categorization tasks using minimal pairs and acoustically manipulated continua (e.g., varying frication duration between /θ/ and /s/) to directly probe changes in cue-to-category mapping. This would allow researchers to determine the specific acoustic dimensions driving perceptual shifts and how those shifts vary across listeners.

In sum, our findings suggest that adaptation to natural nonnative accents can influence perception across phonetic categories. The results support a general relaxation account mediated by shifts in lexical decision criterion, while also pointing to possible roles of distributional learning—potentially occurring even with minimal exposure. Understanding when and how these mechanisms operate is critical for advancing theories of speech perception.

## References

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English lexicon project. *Behavior research methods*, *39*, 445-459.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bent, T., & Baese-Berk, M. (2021). Perceptual learning of accented speech. *The handbook of speech perception*, 428-464.
Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*(2), 707-729.
Bradlow, A. R., Bassard, A. M., & Paller, K. A. (2023). Generalized perceptual adaptation to second-language speech: Variability, similarity, and intelligibility. *The Journal of the Acoustical Society of America*, *154*(3), 1601-1613.
Caudrelier, T., Martin, C. D., Samuel, A. G., Beausoleil, M.-M., Ménard, L., & Tiede, M. (2023). *Lexically-guided phonetic recalibration transfers across languages in French-English bilinguals.* Paper presented at the International Congress on Phonetic Science (ICPhS).
Cummings, S. N., & Theodore, R. M. (2023). Hearing is believing: Lexically guided perceptual learning is graded to reflect the quantity of evidence in speech input. *Cognition*, *235*, 105404.
Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67*(2), 224-238.
Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of phonetics, 25*(4), 437-470.
Flege, J. E., Munro, M. J., & Skelton, L. (1992). Production of the word-final English/t/–/d/contrast by native speakers of English, Mandarin, and Spanish. *The Journal of the Acoustical Society of America, 92*(1), 128-143.
Flege, J. E., Schirru, C., & MacKay, I. R. (2003). Interaction between the native and second language phonetic subsystems. *Speech communication, 40*(4), 467-491.
Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., & Schurman, J. (2010). Short-term adaptation to accented English by younger and older adults. *The Journal of the Acoustical Society of America, 128*(4), EL200-EL204.
Han, F. (2013). Pronunciation problems of Chinese learners of English. *Ortesol Journal, 30*, 26-30.
Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1939.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models (Version 3.1.3) [R package]. *Journal of Statistical Software*, *82*(13), 1–26.
Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological review, 122*(2), 148.
Lai, W., & Tamminga, M. (2024). Phonetics–phonology mapping in the generalization of perceptual learning. *Journal of Phonetics, 103*, 101295.
McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological review*, *118*(2), 219.
Melguy, Y. V., & Johnson, K. (2022). Perceptual adaptation to a novel accent: Phonetic category expansion or category shift? *The Journal of the Acoustical Society of America, 152*(4), 2090-2104.
Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., et al. (2020). An online headphone screening test based on dichotic pitch. *Behavior Research Methods, 53*(4), 1551-1562.
Mitterer, H., Reinisch, E., & McQueen, J. M. (2018). Allophones, not phonemes in spoken-word recognition. *Journal of Memory and Language, 98*, 77-92.
Mitterer, H., Scharenborg, O., & McQueen, J. M. (2013). Phonological abstraction without phonemes in speech perception. *Cognition, 129*(2), 356-361.
Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*(2), 204-238.
Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, *20*(4), 873-922.
R Core Team. (2024). *R: A language and environment for statistical computing.* (Version 4.4.1) [Computer Software]. R Foundation for Statistical Computing. www.R-project.org/
Reinisch, E., & Holt, L. L. (2014). Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance, 40*(2), 539-555.
Reinisch, E., Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance, 39*(1), 75.
Reinisch, E., Wozny, D. R., Mitterer, H., & Holt, L. L. (2014). Phonetic category recalibration: What are the categories? *Journal of phonetics, 45*, 91-105.
Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics, 71*(6), 1207-1218.
Schmale, R., Cristia, A., & Seidl, A. (2012). Toddlers recognize words in an unfamiliar accent after brief exposure. *Developmental Science, 15*(6), 732-738.
Sidaras, S. K., Alexander, J. E., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, *125*(5), 3306-3316.
Tan, M., Xie, X., & Jaeger, T. F. (2021). Using rational models to interpret the results of experiments on accent adaptation. *Frontiers in Psychology*, *12*, 676271.
Vaughn, C., Baese-Berk, M., & Idemaru, K. (2018). Re-examining phonetic variability in native and non-native speech. *Phonetica, 76*(5), 327-358.
Wagenmakers, E. J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of memory and language*, *58*(1), 140-159.

Weatherholtz, K. (2015). *Perceptual learning of systemic cross-category vowel variation*: The Ohio State University.

Xie, X., Jaeger, T. F., & Kurumada, C. (2023). What we do (not) know about the mechanisms underlying adaptive speech perception: A computational framework and review. *Cortex, 166*, 377-424.

Xie, X., & Myers, E. (2015). The impact of musical training and tone language experience on talker identification. *The Journal of the Acoustical Society of America, 137*(1), 419-432.

Xie, X., Theodore, R. M., & Myers, E. B. (2017). More than a boundary shift: Perceptual adaptation to foreign-accented speech reshapes the internal structure of phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(1), 206.

Xie, X., Weatherholtz, K., Bainton, L., Rowe, E., Burchill, Z., Liu, L., et al. (2018). Rapid adaptation to foreign-accented speech and its transfer to an unfamiliar talker. *The Journal of the Acoustical Society of America, 143*(4), 2013-2031.

Theodore, R. M., & Monto, N. R. (2019). Distributional learning for speech reflects cumulative exposure to a talker's phonetic distributions. *Psychonomic Bulletin & Review*, *26*, 985-992.

Zhang, Y., & Xiao, J. (2014). An Analysis of Chinese Students' Perception and Production of Paired English Fricatives: From an ELF Perspective. *Journal of Pan-Pacific Association of Applied Linguistics, 18*(1), 171-192.

Zheng, Y., & Samuel, A. G. (2020). The relationship between phonemic category boundary changes and perceptual adjustments to natural accents. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(7), 1270-1292.