Masthead Logo

University of Connecticut
OpenCommons@UConn

Master's Theses

University of Connecticut Graduate School

4-23-2019

The Influence of Sentence Context on Phonetic Recalibration

Sahil Luthra
sahil.luthra@uconn.edu

This work is brought to you for free and open access by the University of Connecticut Graduate School at OpenCommons@UConn. It has been accepted for inclusion in Master's Theses by an authorized administrator of OpenCommons@UConn. For more information, please contact opencommons@uconn.edu.

The Influence of Sentence Context on Phonetic Recalibration

Sahil Luthra

B.S., Brown University, 2014

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Masters of Science

At the

University of Connecticut

2019

Copyright by

Sahil Luthra

2019

APPROVAL PAGE

Masters of Science Thesis

The Influence of Sentence Context on Phonetic Recalibration

Presented by

Sahil Luthra, B.S.

Major Advisor ____
James S. Magnuson

Major Advisor ____
Emily B. Myers

Associate Advisor ______________________________________
Gerry T. M. Altmann

Associate Advisor ______________________________________
Rachel M. Theodore

Connecticut 2019

Abstract

Individual talkers vary considerably in how they produce different speech sounds, and a challenge for the listener is to learn the appropriate mapping between acoustics and phonetic categories for an individual talker. Several studies have shown that listeners are able to leverage various sources of context (e.g., coincident visual information, lexical knowledge) to guide this process, sometimes termed perceptual learning of speech. Here, we examine how sentence-level semantic information – specifically, whether preceding sentence context is predictive of an upcoming word – might modulate the size of learning effects. Across a series of perceptual learning experiments, we manipulate how learning compares between groups who receive neutral or predictive sentence contexts, also varying whether contexts are presented in the auditory or written modality. Though we observed greater learning for subjects who read predictive contexts than for subjects who read neutral contexts, this finding did not replicate in an identical follow-up experiment, suggesting that potential influences of sentence context on phonetic recalibration may be small. These findings are discussed in the context of the broader literature on perceptual learning.

Keywords:

Sentence context, phonetic recalibration, perceptual learning, speech

Introduction

A core principle in psychology is that perception is guided by past experience. In his classic paper, Goldstone (1998) described *perceptual learning* as a process by which some stimulus dimensions in the environment are highlighted and others are deemphasized, with learning inducing "relatively long-lasting changes" in perception. Consider, for instance, the often-formidable challenge of trying to understand the speech of an unfamiliar talker, perhaps a person with a novel accent or a speech motor impairment. As listeners gain familiarity with that talker, they can make perceptual adjustments that facilitate comprehension in future encounters with that talker relative to an unfamiliar talker (e.g., Nygaard, Sommers, & Pisoni, 1994). Psycholinguists often refer to this form of perceptual learning for speech as *phonetic recalibration*, as it involves reconfiguring how the incoming acoustic signal maps onto known speech sounds (i.e., phonetic categories) for a given talker.

Listeners are able to avail themselves of a variety of contextual cues in order to resolve ambiguity during spoken language comprehension. For instance, a listener's interpretation of an acoustically ambiguous token is influenced by their lexical knowledge of which strings of sounds constitute real words (Ganong, 1980) as well as by the visible movements of a talker's articulators, such as the lips and jaw (McGurk & MacDonald, 1976). Importantly, these contextual cues not only affect perception in the moment but also guide future encounters with that talker when such contextual information is no longer available (e.g., Norris, McQueen, & Cutler, 2003; van Linden & Vroomen, 2007). In a seminal study, Norris et al. introduced a paradigm for *lexically guided perceptual learning* (LGPL; a specific case of phonetic recalibration in which lexical knowledge guides perceptual learning). In this paradigm, listeners typically hear a talker who produces a speech sound that is ambiguous between two fricatives, such as 's'-/s/ and 'sh'-/ʃ/; this ambiguous sound is often noted as /?/. Critically, /?/ is only encountered in disambiguating lexical contexts; that is, some participants will only hear /?/ in contexts where lexical information guides them to interpret it as /s/ (e.g., *dino?aur*), whereas other listeners only hear /?/ in /ʃ/-biased contexts (e.g., *flouri?ing*). All participants also hear clear productions of the contrastive phoneme (e.g., participants who hear *dino?aur* also hear *flourishing*). Following exposure, participants are given a phonetic categorization task using sounds on a continuum from /s/ to /ʃ/. Listeners who are exposed to /?/ in /s/-biased contexts categorize more of the continuum as /s/, whereas listeners who hear the /?/ in /ʃ/ contexts categorize more of the continuum as /ʃ/. That is, listeners adjust how they map that talker's acoustics onto phonetic categories on the basis of the exposure phase. Notably, Norris et al. found that this perceptual learning did not occur when both possible completions of the exposure items resulted in nonwords, suggesting an important role for disambiguating lexical context in guiding learning.

Consistent with Goldstone's (1998) characterization of perceptual learning more generally, LGPL seems to induce relatively long-lasting changes in perception; Eisner and McQueen (2006) found that learning was stable 12 hours after exposure, though recent work by Zhang and Samuel (2014) suggests learning might diminish within a week. Importantly, this form of perceptual learning of speech does not entail a general change in how all incoming acoustic information is mapped on to phonetic categories; rather, phonetic recalibration involves a change in understanding how a *particular* talker produces their speech (e.g., Eisner & McQueen, 2005). Phonetic recalibration does not tend to generalize to a different talker (at least for fricatives and vowels, which differ along spectral dimensions that tend to also provide speaker-specific information; Kraljic & Samuel, 2006, 2007). Further, the extent of learning depends on whether listeners intuit that the variation that they hear is characteristic of the talker (as opposed to being a result of the environment, such as if the talker has a pen in their mouth; Kraljic, Brennan, & Samuel, 2008; Kraljic & Samuel, 2011; Kraljic, Samuel, & Brennan, 2008; Liu & Jaeger, 2018). Finally, learning is limited by whether a particular token can believably be reinterpreted as belonging to a different phonetic category; phonetic recalibration effects tend to be smaller at continuum endpoints, where the stimuli are relatively unambiguous (e.g., Zhang & Samuel, 2014), and a study by Sjerps & McQueen (2010) found that listeners were more likely to resolve a non-speech noise token as /f/ than /s/, leading to asymmetries in perceptual learning. Taken together, these findings suggest that phonetic recalibration is quite a robust phenomenon and that listeners' learning is guided both by the details of the bottom-up signal as well as by higher-level inferences about the likely cause of any acoustically atypical tokens.

Studies using the standard LGPL paradigm have observed stronger learning the more that participants endorse the ambiguous stimuli as belonging to words during the lexical decision task (Scharenborg & Janse, 2013; but see McAuliffe & Babel, 2016), which might be taken as evidence that phonetic recalibration is specifically a result of having to make lexical decisions. However, LGPL has also been observed following several other exposure tasks, including counting the number of words heard during exposure (McQueen, Norris, & Cutler, 2006), old/new judgments (Leach & Samuel, 2007), same/different judgments (Clarke-Davidson, Luce, & Sawusch, 2008), loudness judgments (Drouin & Theodore, 2018), syntactic judgments (Drouin & Theodore, 2018), and even passive exposure (Eisner & McQueen, 2006; Maye, Aslin, & Tanenhaus, 2008; White & Aslin, 2011), suggesting that the phonetic recalibration can be induced following a broad range of exposure tasks, so long as the task encourages the listener to resolve the ambiguous sounds to the intended phonetic category (Kleinschmidt & Jaeger, 2015).

As such, successful phonetic recalibration depends both on successful processing of the acoustic (bottom-up) signal as well as on contextual (i.e., lexical) support for the intended phonetic category. Recent work has demonstrated that perceptual learning of speech is attenuated when attentional resources are directed away from the critical auditory stream (Jesse & Kaplan, 2019; Samuel, 2016) as well as when the critical item is embedded in speech-shaped noise (Zhang & Samuel, 2014). Learning is also determined by the degree of lexical support for the critical phoneme (Drouin & Theodore, 2018), as LGPL is not observed when the ambiguous fricative is presented in a nonword context (Norris et al., 2003; but see Cutler, McQueen, Butterfield, & Norris, 2008) or when it is presented in a word-initial position, in which case lexical support may be too weak to drive learning (Jesse & McQueen, 2011). Indeed, a critical determinant of whether learning will occur appears to be when the relevant information about phoneme identity is encountered, with theoretical accounts positing that listeners must have strong prior expectations about the signal they will receive in advance of receiving the bottom-up input; learning thus occurs when a listener is able to compare the signal they had expected to the signal they actually receive (Davis & Johnsrude, 2007; Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Hervais-Adelman, Davis, Johnsrude, & Carlyon, 2008).

The current study manipulated the amount of lexical support for the intended phonetic category at exposure by varying whether preceding sentence contexts were predictive of the target words, with the goal of seeing whether the strength of lexical support (as modulated by prior context) might influence the size of phonetic recalibration effects. While a number of perceptual learning studies have used sentences in their stimuli, they have not specifically manipulated the degree to which the sentence biases a listener to expect a particular word; rather, such studies have investigated perceptual learning of globally distorted (e.g., time-compressed or noise-vocoded) speech, manipulating whether the exposure sentences are composed of real words or phonotactically legal nonwords (Altmann & Young, 1993; Davis et al., 2005). However, the semantic content of a sentence frame has been shown to bias in-the-moment interpretations of phonetically ambiguous speech; for instance, participants who hear a token ambiguous between /g/ and /k/ are more likely to categorize it as /g/ in a context like *He milked the ?oat* but as /k/ in a context like *He hemmed the ?oat*. Theoretical frameworks have largely assumed that sentence-level semantic information should be able to guide phonetic recalibration (Davis & Johnsrude, 2007; Obleser & Eisner, 2009), but such an assumption has not been tested empirically.

Notably, it is unclear from the extant literature exactly how sentential context would be incorporated with phonetic detail to guide perceptual learning. One possibility comes from *ideal observer* accounts (e.g., Kleinschmidt & Jaeger, 2015), which are based on principles of Bayesian inference and posit that listeners should be able to make use of every available source of information to recognize a token as a member of a particular phonetic category. Given that a predictive sentence context is thought to pre-activate upcoming lexical candidates (e.g., DeLong, Urbach, & Kutas, 2005), a predictive context should therefore boost the prior probability that a particular phoneme will be encountered, thus leading to a larger degree of learning than when phoneme activation is relatively lower. As such, ideal observer accounts would predict that LGPL should be enhanced when critical items follow predictive contexts relative to when they follow neutral contexts. An alternative hypothesis comes from accounts that describe perceptual learning as a result of changes in *attentional weighting* (Goldstone, 1998), whereby stimulus dimensions that are more informative become more heavily weighted than stimulus dimensions that are less informative. When applied to phonetic recalibration, an attentional weighting account would argue that a predictive sentence context permits listeners to resolve the identity of an ambiguous phoneme without attending strongly to the bottom-up signal. As such, listeners who hear predictive sentence contexts should increase the attentional gain on lexical information (as described, for instance, by Pitt & Szostak, 2012) at the expense of the attention directed toward the acoustic signal. Because reducing attention to the bottom-up signal has been shown to reduce phonetic recalibration (Samuel, 2016), an attentional weighting account thus predicts that the degree of perceptual learning should be reduced when critical items follow a predictive sentence context relative to when they follow a neutral one.

Here, we present the results of three experiments assessing how the predictive power of sentential context influences the degree of LGPL. In Experiment 1, we measured the extent of perceptual learning when critical words were presented in isolation (i.e., without any preceding sentence context). Experiments 2 and 3 employed auditory and written sentence contexts, respectively, to assess whether learning was larger for participants who received sentence contexts that were predictive of the critical word or for participants who received contexts that were neutral with respect to their predictive power. We also present the results of several replication experiments in order to validate any effects observed in our initial experiments.

Experiment 1

Experiment 1 was designed to provide a baseline estimate of the extent of perceptual learning when no sentence context was provided at exposure; in doing so, we also would validate that phonetic recalibration would be obtained with slight modifications to the standard LGPL paradigm. In particular, we anticipated that it might seem unnatural if nonwords were presented after sentence contexts, and so we opted to present only words during exposure. As such, we used a semantic categorization task (judging whether the word was a concrete noun) instead of a lexical decision task during exposure; learning was assessed using a standard phonetic categorization task. Additionally, a standard LGPL paradigm would expose a given participant to the ambiguous token in one biasing condition; for instance, a subject might either hear /?/ in /s/-biased contexts or in /ʃ/-biased contexts, but not both (i.e., Bias is treated as a between-subjects factor). Following Saltzman and Myers (2018), we instead manipulated Bias within subjects, first providing subjects with one set of biasing contexts (e.g., /s/-biased contexts), assessing their learning with a phonetic categorization task, and then repeating the procedure with the opposite set of biasing contexts (e.g., /ʃ/-biased contexts); this procedure is schematized in Figure 1, below.

If participants recalibrate on the basis of their previous exposure (as expected from previous LGPL studies), we should see an effect of the most recent Bias condition (/s/ or /ʃ/) on participants' responses during the phonetic categorization task. We also expect an effect of which Step along the continuum participants are hearing, with participants making more /s/ responses when presented with more /s/-like tokens.

Methods

Stimuli. Thirty-two words (16 with word-medial /s/, 16 with word medial /ʃ/) were selected; the full set of stimuli is provided in the Appendix. 16 items (7 with medial /s/, 9 with medial /ʃ/) were taken directly from Kraljic and Samuel (2005), and the remaining items were generated following the same constraints that Kraljic and Samuel used to generate their stimuli. Student *t*-tests indicated no significant difference between /s/-medial and /ʃ/-medial words in written frequency (Kučera & Francis, 1967), $t(28) = 1.2$, $p = 0.24$, or in number of syllables, $t(30) = 0.46$, $p = 0.65$.

To assess the amount of lexical support provided for each of the word-medial fricatives, we computed the frequency-weighted probability that the intended fricative would be the next phoneme given the preceding phonemes. For instance, for the word *episode*, we calculated the probability that the next phoneme would be /s/ given that the preceding phonemes were [ɛpɪ], accounting for the word frequency of each of the possible completions. Probabilities were calculated using the MRC Psycholinguistic Database (Coltheart, 1981). In particular, we used the database to generate phonetic transcriptions for each word, and these transcriptions were then used to find all the words that began with the same onset. In the cases where American English pronunciations differed from those in the MRC database, both British and American English transcriptions were used to find words that shared an onset. Only onset competitors with a written frequency (Kučera & Francis, 1967) of at least 1 were included. If the target word did not appear in the Kučera and Francis corpus (e.g., *Arkansas*), it was assigned a frequency of 1. In this way, we calculated that the intended fricative was predicted with a mean probability of 0.42 (SE: 0.07), which did not differ significantly between /s/-medial (0.37) and /ʃ/-medial (0.46) words, $t(30) = 0.66$, $p = 0.52$. Because the preceding phonemes in the word are not necessarily diagnostic of the upcoming fricative, we expected that the degree of lexical support would be able to be modulated by the preceding sentence context.

Though performance on the semantic categorization task was not of theoretical interest, we also ensured that there were roughly equal numbers of words that were concrete nouns and words that were not. Note that unlike a lexical decision task, answers for the semantic judgment task are rather subjective, as it is not immediately apparent whether some of our items (e.g., *Arkansas*) are concrete nouns or not. Based on experimenter judgment, however, approximately 14 items were concrete nouns and 18 were not; a chi-square test of independence indicated that status as a concrete noun was independent from whether words contained a medial /s/ or /ʃ/, $\chi^2(1) = 0.16$, $p = 0.69$.

Stimuli were produced by a female native speaker of American English, who was recorded in a sound-attenuated booth using a RØDE NT-1 condenser microphone with a Focusrite Scarlet 6i6 digital audio interface. The talker produced both a lexically consistent token (e.g., *episode*) as well as a lexically inconsistent token (e.g., *epishode*) for each item; as described below, these tokens ultimately served as endpoints to generate a word-nonword continuum from which an ambiguous token was selected. The talker produced each token (word and nonword) after each of its corresponding sentence contexts (see Experiment 2), with two productions recorded for each token. The speaker paused before each critical token to reduce the impact of coarticulation on the target item. Finally, the speaker also produced five productions each of the words *sign* and *shine* to generate stimuli for the phonetic categorization task.

Following recording, the default noise reduction filter was applied to the entire audio file in Audacity (Mazzoni & Dannenberg, 2015). Sentence-final tokens (e.g., *episode*, *epishode*) were cut at zero-crossings in Praat (Boersma & Weenik, 2017), and the first author then selected what he judged to be the best production of each lexically consistent item (*episode*) and each lexically inconsistent item (*epishode*). These tokens were scaled to a mean amplitude of 70 dB SPL. For each item, an 11-step word-nonword (e.g., *episode*-*epishode*) continuum was generated using STRAIGHT (Kawahara et al., 2008). An 11-step continuum was also generated from *sign* to *shine*, to be used in the phonetic categorization post-test. Based on experimenter judgment, we decided that step 7 would provide a suitably ambiguous fricative for each continuum; note that the continuum was asymmetric, as step 7 was not the middle step along the generated continuum but was perceptually judged to be the most ambiguous. Step 4 was selected to serve as the clear /s/

token for each item, and step 10 was selected to serve as the clear /ʃ/ token. Thus, all fricative-containing tokens had been morphed in STRAIGHT, and endpoint tokens were an equal number of steps away from the ambiguous token. Similarly, steps 4-10 from the *sign-shine* continuum were selected for use in the phonetic categorization task.

Procedure. The procedure is summarized in Figure 1.

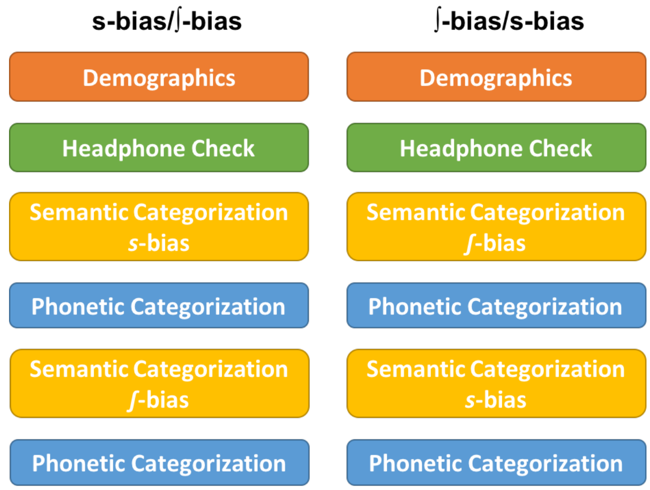

Figure 1. The general procedure for all experiments in this study.

Following a demographics screener, participants completed a short test to assess whether they were using headphones (Woods, Siegel, Traer, & McDermott, 2017). In the experiment proper, each participant completed four experimental blocks. In the first and third block, participants completed a semantic categorization task, one of which was a /s/-biased block and one of which was a /ʃ/-biased block; block order (/s/ or /ʃ/) was counterbalanced across participants. In the /s/-biased condition, participants heard the ambiguous fricative only in contexts where lexical information disambiguated the sound as a /s/ (e.g., *epi?ode*) and also heard clear /ʃ/ endpoints in lexically congruent contexts (e.g., *friendship)*. In the /ʃ/-biased condition, participants heard the ambiguous fricative only in /ʃ/-biasing conditions (*friend?ip*) and a clear /s/ in lexically congruent contexts (*episode*). For these blocks, participants were told that they would hear a word on every trial and would need to decide if it was a concrete noun. A concrete noun was defined during the instructions as a person, place or thing that can be experienced with any of the five senses (sight, sound, smell, touch, taste). Item order was randomized for each participant.

During the second and fourth blocks of the experiment, participants participated in a phonetic categorization task. They were told they would hear the word *sign* or *shine* on each trial and to indicate as quickly as possible which one they heard. Participants heard each step from the 7-step continuum ten times presented in a random order, yielding a total of 70 trials for each block.

For both tasks, participants were prompted to indicate their response with the keyboard after they heard the stimulus; button mappings were counterbalanced across participants. Participants had 4 seconds to make their response before the trial timed out, and there was a 1-second interval between trials.

The experiment was programmed using custom JavaScript code using functions from the jsPsych library (de Leeuw, 2015) and was hosted online using Google App Engine. All code and stimuli are publicly available at https://github.com/sahil-luthra/SenCoPL. The full session took approximately 30 minutes, and procedures for all the experiments reported in this paper were approved by the University of Connecticut's Institutional Review Board. In all experiments, subjects gave informed consent prior to participating.

Participants. Data were collected online using Amazon Mechanical Turk, a crowdsourcing platform that has previously been used to study phonetic recalibration (Kleinschmidt & Jaeger, 2015; Liu & Jaeger, 2018). To qualify for the study, participants had to have the US set as their location and also had to have indicated that they had normal or corrected-to-normal hearing in both ears, that their computer played sound, and that American English was the only language they spoke prior to age 13. After an individual participated in one experiment, they were deemed ineligible to participate in subsequent experiments reported in this paper. Participants were paid \$5.05 for completing the full experiment, and \$0.85 if they were deemed ineligible after completing the initial demographics screener. Payment amounts were based on estimated maximum time to complete the full experiment and the screening, respectively, multiplied by Connecticut's minimum wage of \$10.10 per hour.

Data were excluded from analyses if participants failed to respond to a substantial portion (≥ 10%) of the trials on either task or if they showed poor accuracy (≤ 70%) in phonetic categorization of the unambiguous endpoints. In this way, data from 9 eligible participants were excluded, leaving 40 participants whose data were included in analyses.

Results

Results from the phonetic categorization task are visualized in Figure 2. Visually, recalibration is noted as a difference in the pattern of phonetic categorization following /ʃ/-biased contexts (shown in red) as compared to the /s/-biased contexts (shown in blue).

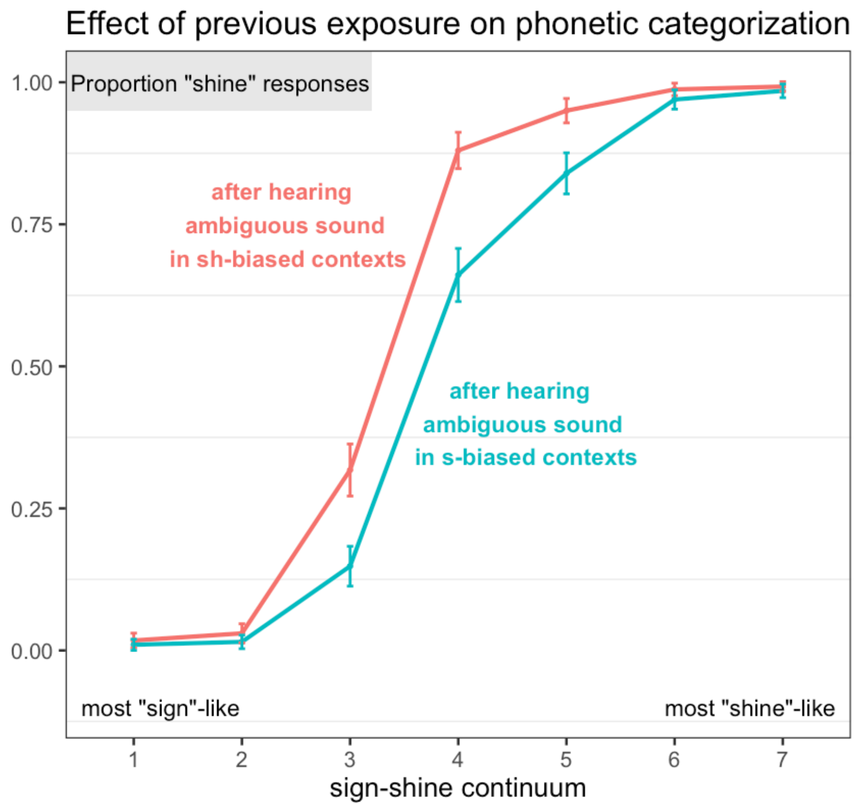

Figure 2. Data from the phonetic categorization task of Experiment 1, in which all words were encountered in isolation, without any sentence context. Error bars indicate 95% confidence intervals.

Data were analyzed using mixed effects logistic regression in R (R Core Team, 2018) implemented using the *mixed* function in the "afex" package (Singmann, Bolker, Westfall, & Aust, 2018). The model considered fixed factors of Bias (ʃ-bias, s-bias; coded with a [-0.5, 0.5] contrast) and Step (centered) as well as random by-subject slopes for Bias and Step, random by-subject interactions between Bias and Step, and random intercepts for each subject. There was an effect of Bias, $\chi^2(1) = 28.70$, $p < 0.001$, demonstrating that phonetic recalibration indeed occurred. We also observed an expected effect of Step, $\chi^2(1) = 224.48$, $p < 0.001$, indicating that participants made more /ʃ/ responses as the continuum tokens became more /ʃ/-like. The interaction between Bias and Step was not significant, $\chi^2(1) = 0.20$, $p = 0.66$.

Discussion

Data from Experiment 1 indicated that phonetic recalibration occurred when participants were given a semantic categorization task during exposure, consistent with previous studies that have demonstrated LGPL effects without using a lexical decision task (Drouin & Theodore, 2018; Eisner & McQueen, 2006; Leach & Samuel, 2007; Maye et al., 2008; McQueen et al., 2006); to our knowledge, this is the first LGPL study to use a concreteness judgment during exposure. In another departure from the standard paradigm, we manipulated Bias within-subject (following Saltzman & Myers, 2018), allowing us to minimize the influence of subject-to-subject variability on our conditions and facilitating the identification of any potential group differences in subsequent experiments. Having ascertained that our paradigm can successfully induce phonetic recalibration, we turn next to two experiments designed to examine whether the nature of preceding sentence context (i.e., whether it is predictive of the final word or neutral with respect to it) can modulate the size of perceptual learning effects.

Experiment 2

In Experiment 2, we examined if the extent of LGPL might be influenced by whether a preceding auditory sentence context predicted a critical lexical item. One group of participants heard a predictive auditory context (e.g., *I love "The Walking Dead" and eagerly await every new…*) before each target item (e.g., *episode*), while another group heard neutral sentence contexts (*My ballpoint pen ran out of ink when I was halfway through writing the word…*). We expected that Context (predictive / neutral) would modulate the size of the Bias (s-bias / ʃ-bias) effect (i.e., we expected a Bias x Context interaction). Of interest is whether the recalibration effect would be larger for predictive contexts, as would be predicted by an ideal observer account, or whether more learning would occur for neutral contexts, as would be predicted by an attentional weighting account.

Methods

Stimuli. As described in the Methods for Experiment 1, 32 words (16 with word-medial /s/, 16 with word-medial /ʃ/) had been selected for use in the current experiments. As described in the Methods for Experiment 1, the phonemes that precede the fricative do not unambiguously determine the identity of the critical fricative for our set of target words, suggesting that sentence context should be able to modulate the prior probability of the intended word. For each item, we created two predictive contexts and two neutral contexts. Two contexts were needed per item because every subject was exposed to each item twice (once in the /s/-biased exposure block and once in the /ʃ/-biased block), and we did not want subjects who were receiving neutral contexts to be able to predict the sentence-final target during their second exposure block (on the basis of their memory for sentence contexts from the first exposure block). As such, we created one set of sentence contexts for the first exposure block and a separate set of contexts for the second exposure block.

The predictive power of our sentence contexts was assessed with a norming pretest. In the pretest, participants were given sentence contexts and asked to complete each one with the first word that came to mind (Taylor, 1953). A given participant only saw one of the two sentences that was designed to predict a particular item. Occasionally, some participants withdrew before completing all sentences, so a total of 65 participants were recruited in order to collect 20 responses for each sentence context. Participants were recruited through Amazon Mechanical Turk and compensated at a rate of $10.10/hour. The cloze probability of the target word in each predictive sentence context is given in the Appendix. The intended target had a mean cloze probability of 0.74 in predictive contexts (SE: 0.23), and this did not differ between /s/-medial and /ʃ/-medial targets, $t(62) = 0.11$, $p = 0.91$. Neutral contexts never elicited their associated target items.

Sentence contexts did not include /s/ or /ʃ/ phonemes, as initial exposure to normal tokens would be likely to attenuate perceptual learning for those fricatives (Kraljic, Samuel, & Brennan, 2008).[1] However, sentence contexts did include other fricatives (e.g., /f/, /v/, /z/), as excluding those phonemes would have too dramatically limited the scope of possible words. This raises the possibility that hearing other unaltered fricatives could affect phonetic recalibration for /s/ and /ʃ/; we consider the possibility in Experiment 3. Sentence contexts had a mean length of 14.5 words. An analysis of variance indicated that there were no differences in sentence length as a function of the target's medial fricative (/s/ or /ʃ/), $F(1,30) = 0.18$, $p = 0.68$, as a function of the type of context (neutral or predictive), $F(1,30) = 1.18$, $p = 0.29$, or as a function of whether the sentence context would be used in the first or second exposure block, $F(1,30) = 1.77$, $p = 0.19$. There were also no significant interactions between any of these factors.[2]

Sentence contexts were recorded during the same recording session as critical target items (see Methods for Experiment 1). Sentence contexts were excised from the auditory file by cutting at zero-crossings in Praat; the first author then selected the context he deemed to be the best

---

[1] During recording, it was noted that three normed contexts each contained an instance of /s/, so we opted to record minimally altered sentences. While these new contexts were not identical to the ones normed, we do not expect these minimal changes to substantially affect cloze probabilities. In particular, "on the fir<u>s</u>t day of camp" was changed to "at the beginning of camp;" the word "intere<u>s</u>ting" was changed to "intriguing;" and an instance of "<u>s</u>o" was changed to "and."

[2] Note that the results of this ANOVA reflect the lengths of the modified sentences.

recording. These contexts were then scaled to a mean amplitude of 72 dB in Praat and concatenated with the sentence-final words (which had been scaled to a mean amplitude 70 dB). Note that we used different dB values here to equate for differences in perceived amplitude.

Procedure. The procedure followed for Experiment 2 was identical to that followed for Experiment 1, with the exception that the exposure trials involved the presentation of an auditory sentence context prior to the critical word. Participants were told that their task during the exposure blocks was to decide if the final word of each sentence was a concrete noun.

Participants. Data from 94 participants were collected on Amazon Mechanical Turk, following the same procedures as in Experiment 1. As before, we excluded the data of participants who failed to respond to at least 10% of the trials on either the semantic categorization or phonetic categorization tasks as well as the data from participants whose categorization of the continuum endpoints was at or below 70%. Thus, we analyzed data from 80 participants, 40 of whom received predictive contexts and 40 of whom received neutral contexts.

Results

Data from Experiment 2 are visualized in Figure 3. As before, red lines indicate phonetic categorization responses after /ʃ/-biasing blocks, while blue lines indicate categorization after /s/-biasing blocks; as such, the difference between red and blue lines is taken as the size of the phonetic recalibration effect. Evidence for an influence of sentential context on the size of recalibration effects would be observed as a difference in the size of the phonetic recalibration effect after predictive sentence contexts (shown in solid lines) as compared to neutral contexts (shown in dashed lines).

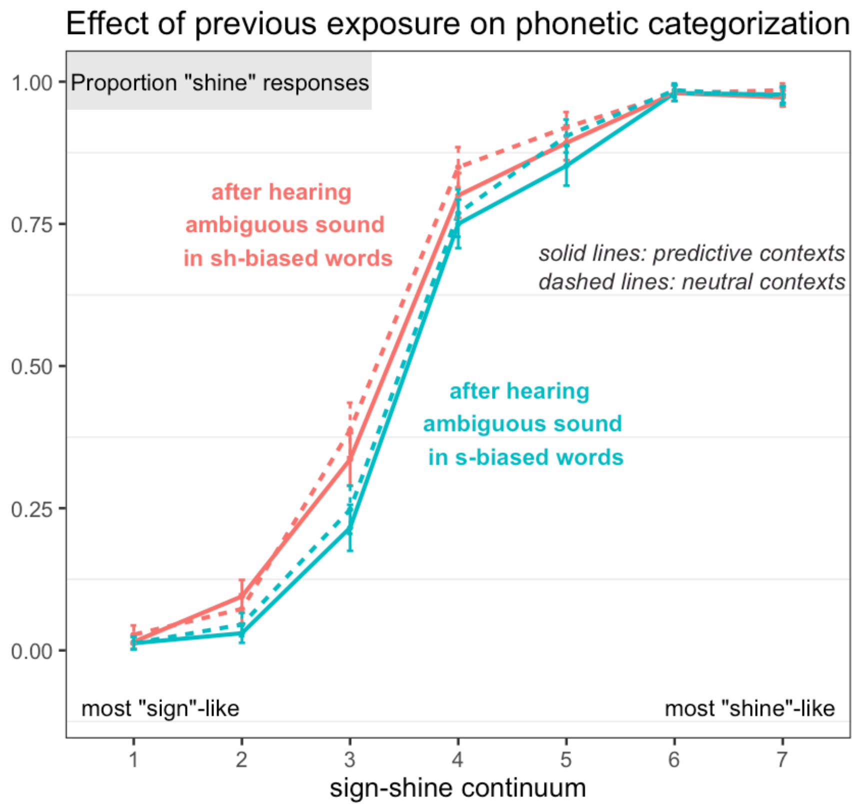

Figure 3. Data from the phonetic categorization task in Experiment 2, in which participants heard neutral or predictive sentence contexts prior to items with word-medial fricatives. Error bars indicate 95% confidence intervals.

Data from Experiment 2 were analyzed using mixed effects logistic regression that considered fixed factors of Bias (ʃ-bias, s-bias; coded with a [-0.5, 0.5] contrast), Step (centered), and Context (Neutral, Predictive; coded with a [-0.5, 0.5] contrast), random by-subject slopes for Bias and Step, random by-subject interactions between Bias and Step, and random intercepts for each subject. The model indicated the expected effect of Bias, $\chi^2(1) = 15.66$, $p < 0.001$, demonstrating phonetic recalibration. However, the expected interaction with Context was not observed, $\chi^2(1) = 0.01$, $p = 0.92$. Finally, we observed significant effects of Step, $\chi^2(1) = 224.48$, $p < 0.001$, and Bias x Step, $\chi^2(1) = 6.19$, $p = 0.01$.

A follow-up analysis considered the overall influence of providing a sentence context on phonetic recalibration. For this analysis, we compared the extent of phonetic recalibration in the Neutral sentence context group to the recalibration in the Word Only group from Experiment 1. The model included the factors of Bias (ʃ-bias, s-bias; coded with a [-0.5, 0.5] contrast), Step (centered), and Context (Word Only, Neutral; coded with a [-0.5, 0.5] contrast) and used the same random effect structure as above. We observed the expected significant effects of Bias, $\chi^2(1) = 32.50$, $p < 0.0001$, and Step, $\chi^2(1) = 433.21$, $p < 0.001$. Additionally, we observed a significant Bias x Context interaction, $\chi^2(1) = 3.85$, $p = 0.05$ and a significant Step x Context interaction, $\chi^2(1) = 4.51$, $p = 0.03$. To clarify the nature of the Bias x Context interaction, follow-up models considered the effects of Bias and Step within each level of Context (Word Only, Neutral). A larger Bias effect was observed for the group who heard words in isolation, $\chi^2(1) = 28.70$, $p < 0.0001$, than for the group who heard neutral auditory contexts, $\chi^2(1) = 7.29$, $p = 0.007$.

Discussion

In Experiment 2, we observed phonetic recalibration when the critical stimuli were presented in sentence context, consistent with other LGPL studies where critical items were embedded in sentence contexts (Eisner & McQueen, 2006; Maye et al., 2008). However, we did not observe an interaction between Context and Bias, suggesting that the size of the LGPL effect was not modulated by whether the sentence context was predictive of the critical word or neutral with respect to predicting the final word. This is contrary to the predictions of an ideal observer account, which predicts larger recalibration effects for predictive contexts compared to neutral ones, and also contrary to the predictions of an attentional weighting account, which predicts larger recalibration effects in neutral contexts than predictive ones.

Additionally, we found that phonetic recalibration effects were *smaller* when participants heard a neutral sentence context prior to each critical word than when they heard critical words in isolation. To our knowledge, no previous studies have compared the degree of LGPL that occurs when words are presented in isolation as compared to when they are presented in sentence contexts. It is not immediately apparent why learning should be attenuated when critical items are presented in the context of sentences. However, it is possible that the other acoustic information in the auditory sentence context reduced the degree of learning. Kraljic, Samuel, and Brennan (2008) demonstrated that if a listener encounters a talker producing a prototypical fricative (for instance, a clear production of /s/) prior to an atypical variant of that fricative in a lexically disambiguating context (e.g., producing /?/ in a /s/-biased context), learning will be attenuated. Though the sentence contexts we created did not include other instances of /s/ or /ʃ/ for this very reason, the contexts did include other fricatives, such as /f/ and /z/. It is possible that the perceptual units that are recalibrated during LGPL are sub-phonemic units (as suggested, for instance, by Kraljic & Samuel, 2006). If this is the case, the presence of these other fricatives in the auditory contexts may have reduced learning. Thus, even though a significant effect of Bias was observed in Experiment 2, an overall reduction in the extent of recalibration might have obscured differences in the size of recalibration effects as a function of sentence context and might also explain why there was relatively less learning when listeners received sentence contexts as compared to when they heard the critical words in isolation.

As such, there are at least two ways to interpret the results of Experiment 2. First, it is possible that the higher-level context provided by sentences diminishes the extent of phonetic recalibration, though it is not theoretically clear why this should be the case. If this is the case, this might suggest that LGPL effects in the lab may not reflect the size of recalibration effects in ecological instances of spoken word recognition, as words are typically encountered in the context of sentence-level information. Alternatively, it may be the case that the additional acoustic information about the talker provided by auditory sentence contexts diminished learning and thus potentially obscured group differences in learning effects. In Experiment 3, we attempt to distinguish between these alternatives by presenting critical items after *visually* presented sentence contexts, which do not provide additional auditory information about the talker.

Experiment 3

In Experiment 3, exposure blocks consisted of auditory stimuli with word-medial fricatives that were presented after visual sentence contexts. The use of visual contexts allows us to provide higher-level context for the target items but without providing subjects with auditory exposure to the fricatives (/f/, /z/) in the sentence contexts. Exposure blocks were followed by phonetic categorization blocks, as before. Here, we were specifically interested in whether we would observe larger recalibration effects with predictive contexts or with neutral ones, as well as whether we would still observe larger LGPL effects with neutral contexts as compared to when words were presented in isolation.

Methods.

Stimuli. We used the same stimuli as in Experiment 1. Sentence contexts were taken from Experiment 2 but presented visually.

Procedure. We followed the same procedure as in Experiment 2 with one important exception: Instead of presenting sentence contexts in the auditory modality during exposure blocks, contexts were provided visually. To encourage participants to read the full context, we used a self-paced reading design, where participants were only shown one word of the sentence at a time and had to press the spacebar to see the next word. Text was presented in size 2em center-aligned black Open Sans text on a white background. The final word of the sentence was presented in the auditory modality, and participants made the same semantic categorization judgment as in the previous experiments, with stimulus timings and button mappings set as in Experiment 1.

Participants. Participants were recruited from Amazon Mechanical Turk, as before. Following the same exclusion criteria as in previous experiments resulted in discarding the data of 15 participants, leaving 80 for analysis (40 who heard predictive contexts, 40 who heard neutral ones).

Results

Results from Experiment 3 are plotted in Figure 4.

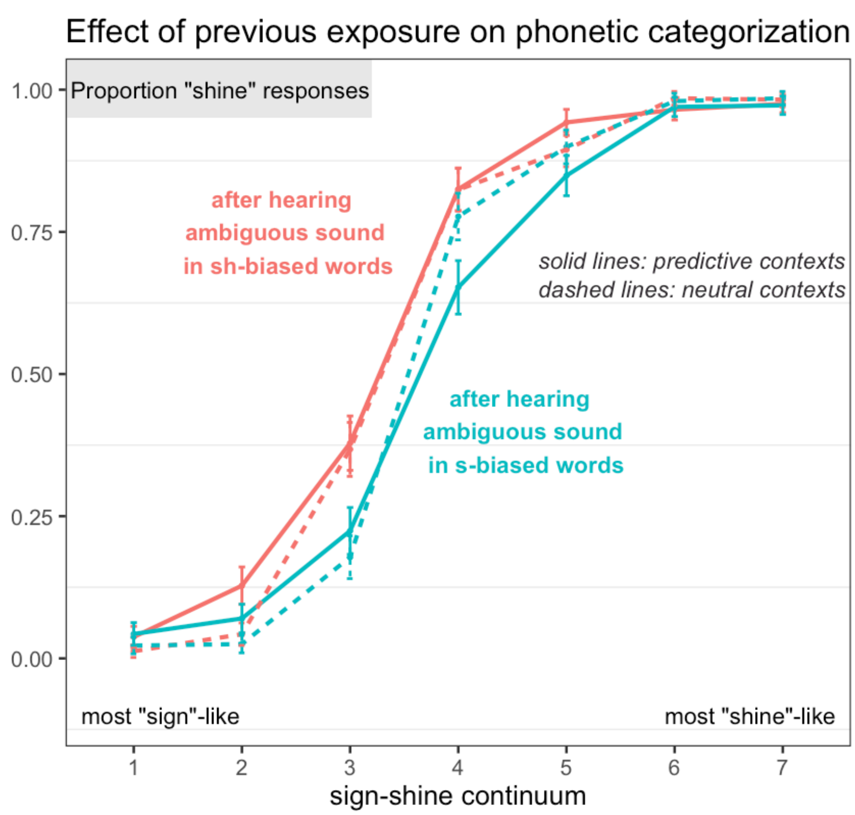

Figure 4. Data from the phonetic categorization task in Experiment 3, in which participants read neutral or predictive sentence contexts prior to items with word-medial fricatives. Error bars indicate 95% confidence intervals.

Data were submitted to mixed effects logistic regression that considered fixed effects of Bias (ʃ-bias, s-bias; coded with a [-0.5, 0.5] contrast), Step (centered), and Context (Neutral, Predictive; coded with a [-0.5, 0.5] contrast). The maximal model did not converge, so a simplified random effects structure was used that considered only by-subject random intercepts. We observed the expected effect of Bias, $\chi^2(1) = 66.31$, $p < 0.0001$, demonstrating that phonetic recalibration had occurred, as well as an effect of Step, $\chi^2(1) = 9112.60$, $p < 0.0001$. There was also a significant interaction between Bias and Context, $\chi^2(1) = 6.31$, $p = 0.01$, indicating that the size of the Bias effect differed as a function of the type of sentence context. The interaction between Step and Context was also significant, $\chi^2(1) = 40.92$, $p < 0.0001$. Finally, the interaction between Bias, Step, and Context approached but did not reach significance, $\chi^2(1) = 3.59$, $p = 0.06$.

To further examine the interactions with Context, we examine the simple effects of Bias and of Step at each level of context. To do so, we implemented models that examined each level of Context (Neutral, Predictive) separately and considered fixed effects of Bias and Step. For consistency with the omnibus model, only random by-subject intercepts were considered. Analysis of data from participants who received Neutral contexts revealed significant effects of Bias, $\chi^2(1) = 18.49$, $p < 0.001$, and Step, $\chi^2(1) = 5490.48$, $p < 0.0001$, as well as a non-significant interaction between Bias and Step, $\chi^2(1) = 3.09$, p = 0.08. Data from the Predictive group showed significant effects of Bias, $\chi^2(1) = 61.09$, $p < 0.0001$, and Step, $\chi^2(1) = 4671.18$, $p < 0.0001$, and a non-significant interaction between these factors, $\chi^2(1) = 1.76$, $p = 0.18$. These results indicated that the Step x Context interaction in the omnibus model was driven by a larger influence of Step for participants who heard Neutral contexts than for those who heard Predictive ones, though it is not readily apparent why that would be the case theoretically. Critically, the effect of Bias was larger for participants who received Predictive contexts than for those who received Neutral contexts.

As in Experiment 2, additional analyses were conducted to examine whether recalibration differed when words were presented in (neutral) sentence contexts as compared to when they were presented in isolation (the Word Only condition from Experiment 1). Specifically, an omnibus model considered fixed effects of Bias (ʃ-bias, s-bias; coded with a [-0.5, 0.5] contrast), Step (centered), and Context (Word Only [Experiment 1], Neutral; coded with a [-0.5, 0.5] contrast). The random effects structure consisted of by-subject random intercepts, random slopes for Bias and Step, and random Bias x Step interactions. Results indicated the expected significant effects of Bias, $\chi^2(1) = 32.50$, $p < 0.001$, and of Step, $\chi^2(1) = 433.21$, $p < 0.001$. We also observed a Bias x Context interaction, $\chi^2(1) = 3.85$, $p = 0.05$, as well as a Step x Context interaction, $\chi^2(1) = 4.51$, $p = 0.03$. Follow-up models considered each level of Context independently. As in Experiment 2, a larger effect of Bias was observed for the group who received words in isolation, $\chi^2(1) = 28.70$, $p < 0.0001$, than for the group who read neutral sentence contexts, $\chi^2(1) = 7.29$, $p = 0.007$.

Discussion

To test whether the lack of Bias x Context effects in Experiment 2 arose because the additional exposure to the talker provided by auditory sentence contexts attenuated LGPL overall, Experiment 3 used visual sentence contexts prior to each auditory stimulus. In this way, participants received the contextual support of the sentence context without the additional auditory information (in particular, clear variants of other fricative contrasts) that might have attenuated learning. Results indicated that there was a greater degree of recalibration for participants who read predictive contexts as compared to those who read neutral contexts. Such a finding is consistent with the prediction of ideal observer accounts, as listeners appear to be able to leverage sentence-level information to enhance perceptual learning.

However, results of Experiment 3 also demonstrated that there was less learning when words were preceded by written (neutral) sentence contexts than when they were encountered in isolation. Such a finding is not predicted by an ideal observer account, as a neutral context should not have affected the prior probability of encountering a /s/ or /ʃ/. As such, an ideal observer account would have predicted the same degree of learning to occur for neutral sentence contexts as for words presented in isolation.

Because the results thus far only provide partial support for an ideal observer account, we turn next to a set of experiments in which we attempt to replicate the findings described thus far. For these studies, data were collected from new samples that had not participated in any of the previous studies.

Replication: Experiment 1

In our replication of Experiment 1, participants encountered critical words in isolation (i.e., without sentence context) during exposure.

Methods

Data were collected from 43 subjects who were recruited from the University of Connecticut's psychology participant pool and completed the experiment in the lab. Data from 3 participants were excluded following the same exclusion criteria used in previous experiments, resulting in data from 40 participants being included in analyses. The same procedures were followed as in Experiment 1.

Results

Data from replication 1 are visualized in Figure 5.

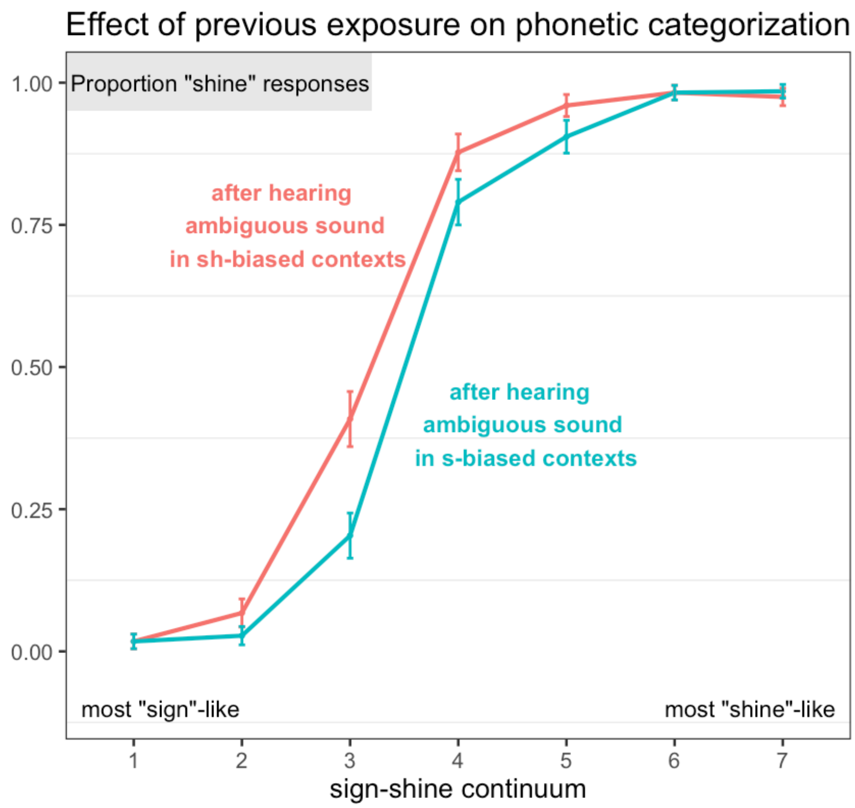

Figure 5. Data from the phonetic categorization task of Replication 1; as in Experiment 1, all words were encountered in isolation. Error bars indicate 95% confidence intervals.

Following the analysis procedure from Experiment 1, data were analyzed using a mixed effects model with fixed factors of Bias (ʃ-bias, s-bias; coded with a [-0.5, 0.5] contrast) and Step (centered). Because the model with the maximal random effects structure did not converge, we used a model that included random intercepts for each subject as well as random by-subject slopes for Bias. Results indicated that phonetic recalibration occurred, as shown by an effect of Bias, $\chi^2(1) = 27.25$, $p < 0.001$, and that participants were sensitive to the bottom-up stimulus properties, as shown by an effect of Step, $\chi^2(1) = 5465.50$, $p < 0.001$. The interaction between Bias and Step was also significant, $\chi^2(1) = 5.45$, $p = 0.02$, indicating that the size of the Bias effect was not constant at every step on the continuum.

Discussion

The findings of this replication study demonstrate that, as in Experiment 1, phonetic recalibration can be observed in a modified LGPL paradigm in which subjects make semantic decisions on exposure items (rather than performing the more standard lexical decision task). Notably, the original Experiment 1 was conducted using an online platform, whereas this replication experiment was conducted in the lab with university students; the consistency of the findings across these two samples provide evidence that online experiments can be used to successfully measure listener sensitivity to subtle acoustic differences, a finding that is particularly noteworthy since online studies require the experimenter to sacrifice a sizable degree of experimental control over the listening environment. The data collected in this replication study also provide a baseline against which to compare the extent of learning in participants exposed to neural sentence contexts in our replications of Experiments 2 and 3, below.

Replication: Experiment 2

In our replication for Experiment 2, participants heard neutral or predictive sentence contexts prior to the target exposure stimuli, with Context manipulated between subjects.

Methods

Data were collected from 83 participants who were recruited from the University of Connecticut's psychology participant pool and completed an in-lab experiment. Results from 3 participants were excluded based on our exclusion criteria, yielding in data from 80 participants (40 per group) included in each analysis.

Results

Data from Replication 2 are visualized in Figure 6.

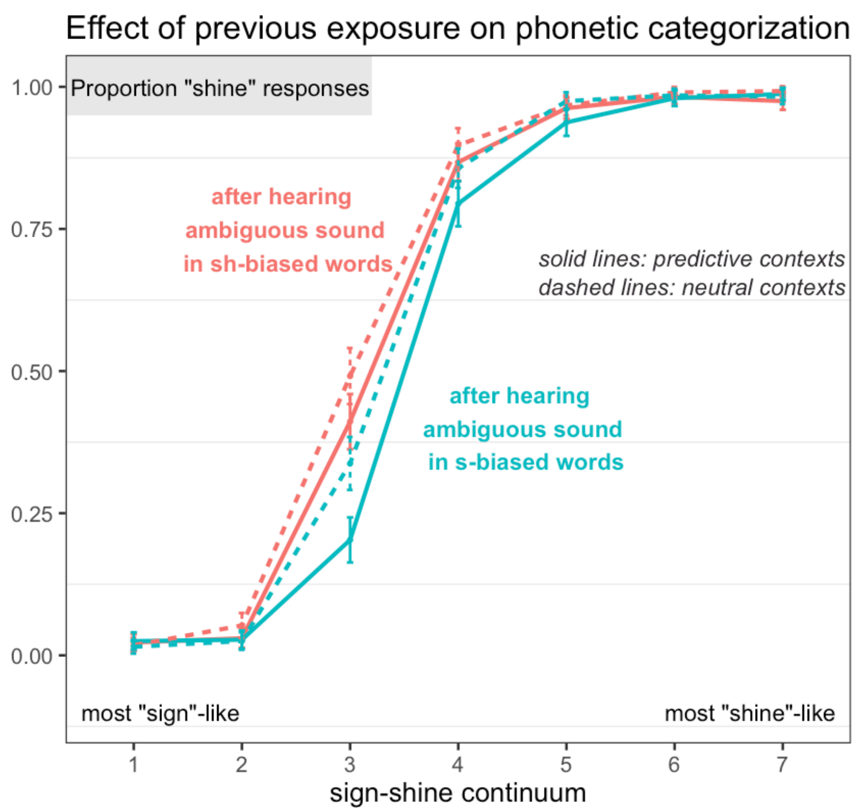

Figure 6. Data from the phonetic categorization task in Replication 2, in which participants heard neutral or predictive sentence contexts prior to items with word-medial fricatives. Error bars indicate 95% confidence intervals.

Data were submitted to a mixed effects logistic regression analysis using the same fixed and random effects structure as in Experiment 2. Results indicated the expected effects of Bias, $\chi^2(1) = 26.91$, $p < 0.0001$, and of Step, $\chi^2(1) = 401.55$, $p < 0.0001$. We also observed a significant Bias x Step interaction, $\chi^2(1) = 6.08$, $p = 0.01$. There was a marginal effect of Context, $\chi^2(1) = 3.76$, $p = 0.05$ and critically no Bias x Context interaction, $\chi^2(1) = 0.44$, $p = 0.51$.

We also examined how the size of LGPL effects compared when participants received sentence contexts compared to when participants heard words in isolation. As such, an additional analysis compared the size of the learning effects for those participants who heard neutral contexts in Replication Study 2 to those who heard the target words in isolation in Replication Study 1. The model included fixed factors of Bias (ʃ-bias, s-bias; coded with a [-0.5, 0.5] contrast), Step (centered), and Context (Word Only, Neutral; coded with a [-0.5, 0.5] contrast), and it also included random by-subject slopes for Bias as well as random intercepts for each subject. We observed the expected main effects of Bias, $\chi^2(1) = 36.85$, $p < 0.0001$, and Step, $\chi^2(1) = 11009.11$, $p < 0.0001$, and a Bias x Step interaction was also observed, $\chi^2(1) = 8.08$, $p = 0.004$. No other main effects or interactions were significant. Notably, the Bias (ʃ-bias/s-bias) x Context (Neutral/Word Only) interaction observed in Experiment 2 was not observed here, $\chi^2(1) = 2.53$, $p = 0.11$.

Discussion

As with Experiment 2, LGPL occurred in Replication 2, demonstrating that perceptual learning of speech does occur when the critical speech is presented in the context of sentences. Consistent with the results from Experiment 2, we did not find an influence of whether the context was neutral or predictive on the size of the Bias effect.

Additionally, we had observed in Experiment 2 that the size of the phonetic recalibration effect was larger when words were presented in isolation than when they were presented in neutral sentence contexts; however, this finding did not replicate here, suggesting that the mere presence of a sentence context may not actually lead to a smaller degree of perceptual learning, as had been suggested by the original Experiment 2.

Replication: Experiment 3

Methods

Data from 111 participants were collected online using Amazon Mechanical Turk. Following the same exclusion criteria as above, 31 participants were excluded, yielding 80 participants (40 in each group) whose data were considered in analyses.

Results

Data from Replication 3 are displayed in Figure 7.

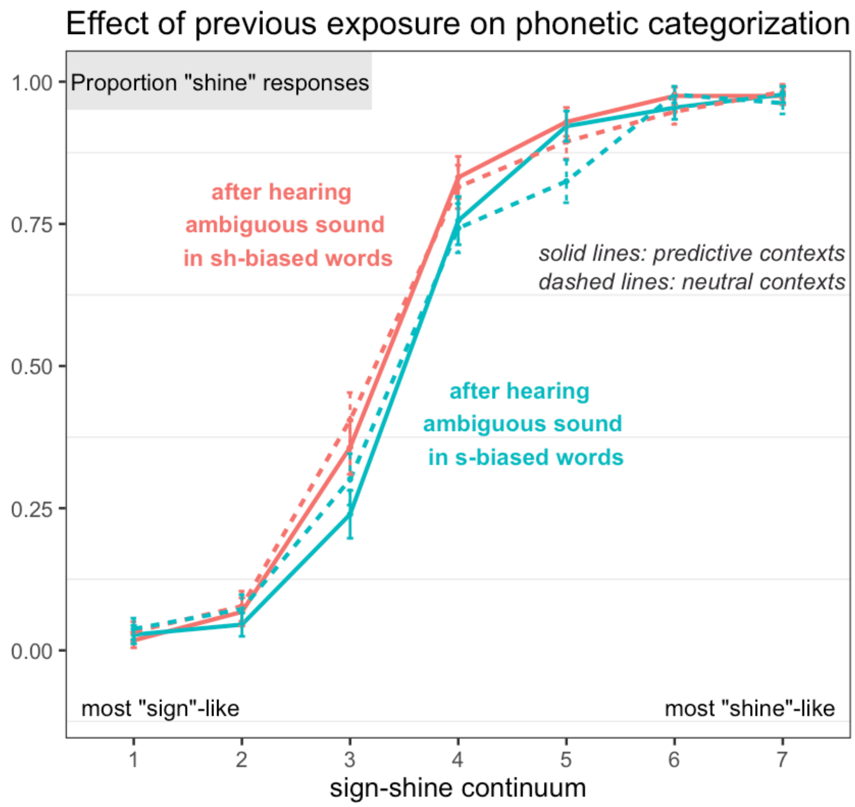

Figure 7. Data from the phonetic categorization task in Replication 3, in which participants read neutral or predictive sentence contexts prior to items with word-medial fricatives. Error bars indicate 95% confidence intervals.

Data were submitted to a logit mixed model that considered fixed effects of Bias (ʃ-bias, s-bias; coded with a [-0.5, 0.5] contrast), Step (centered), and Context (Neutral, Predictive; coded with a [-0.5, 0.5] contrast). The maximal model also considered random by-subject intercepts, random by-subject slopes for Bias and Step, and random by-subject interactions between those factors. In contrast to the original Experiment 3, this maximal model converged, and we report the results of this model. In particular, we observed the expected effects of Bias, $\chi^2(1) = 14.39$, $p =$

0.0001, and of Step, $\chi^2(1) = 324.14$, $p < 0.0001$. Critically, the Bias x Context interaction observed in Experiment 3 was not observed in this replication study, $\chi^2(1) = 0.01$, $p = 0.94$.

Finally, we compared the extent of learning for participants read neutral sentence contexts to the extent of learning for participants who did not receive any sentential context (participants from Replication 1). The model included fixed factors of Bias (ʃ-bias, s-bias; coded with a [-0.5, 0.5] contrast), Step (centered), and Context (Word Only, Neutral; coded with a [-0.5, 0.5] contrast) as well as the maximal random effect structure (random by-subject intercepts, random slopes for Bias and Step for each subject, and random by-subject interactions between Bias and Step). We observed a main effect of Bias, $\chi^2(1) = 26.82$, $p < 0.0001$ and a main effect of Step, $\chi^2(1) = 347.13$, $p < 0.0001$. The only significant interaction was between Step and Context, $\chi^2(1) = 4.59$, $p = 0.03$. Notably, the interaction between Bias and Context (Word Only / Neutral) that was observed in Experiment 3 was not replicated here.

Discussion

Though we did observe robust phonetic recalibration in Replication 3, we did not replicate the critical Bias x Context (Predictive / Neutral) interaction, indicating that the size of the phonetic recalibration effect was comparable for subjects who read Neutral and Predictive sentence contexts. This suggests that if real, the critical finding of Experiment 3 – that the predictive power of the sentence context modulates modulate the size of perceptual learning effects – may be relatively small. In the General Discussion, we consider possible reasons that Context did not modulate Bias effects in our experiments.

We also examined the overall influence of embedding words in sentences by comparing the extent of learning for subjects who received words in isolation to the extent of learning for subjects who heard neutral sentence contexts. We observed no difference in the size of the recalibration effect between these groups, suggesting that simply embedding words in sentences does not necessarily impact the extent of perceptual learning.

General Discussion

The current study was designed to examine how the predictive power of a sentence context might affect learning, as existing theoretical accounts make opposing predictions of how sentence-level semantic information will influence the extent of recalibration. In a series of experiments using a modified LGPL paradigm, listeners were exposed to a talker who produced a fricative that was ambiguous between /s/ and /ʃ/ in lexical contexts that biased the listener to interpret the ambiguity in a particular way. While some participants heard the critical words without any preceding sentence context (Experiment 1), others received auditory (Experiment 2) or visually presented (Experiment 3) sentence contexts that were either predictive of or neutral with respect to the upcoming critical word. Phonetic recalibration was observed across all experiments, as participants' subsequent categorization of items along a *sign-shine* continuum was guided by their previous exposure (i.e., whether they heard the ambiguous fricative in /s/-biased or /ʃ/-biased contexts). In Experiment 2, we found that the predictive power of the sentence context did not modulate the size of the recalibration effect when participants heard auditory contexts; this finding was also observed in a separate replication study. Notably, an interaction between Context and Bias was observed in Experiment 3, with larger phonetic recalibration observed when participants read predictive sentence contexts than when they read neutral ones. While such a finding suggests that predictive sentence contexts can boost learning (so long as participants don't hear other fricatives from that talker, as in Experiment 2), this finding was not observed in a separate replication study. In general, then, findings from the experiments presented here suggest that if there is an influence of the predictive power of sentential context on the size of phonetic recalibration effects, it is not a very strong one.

These findings are striking when considered alongside the fact that perceptual learning of speech has been shown to occur on the basis of several different types of cues, including lexical knowledge (e.g., Norris et al., 2003), lipread information (van Linden & Vroomen, 2007) and written text (Bonte, Correia, Keetels, Vroomen, & Formisano, 2017; Keetels, Schakel, Bonte, & Vroomen, 2016; Mitterer & McQueen, 2009). Based on these findings, one might in principle expect phonetic recalibration to occur on the basis of *any* cue that guides interpretation of how acoustics should map onto a phonetic category. Why, then, did we not observe strong influences of sentence context on phonetic recalibration in the current study?

One possibility is that the current experiments employed conditions that favored robust learning, and with a relatively large degree of learning, it might be hard to observe robust effects of sentence context. This sort of a ceiling effect might have emerged in part because in the current study, listeners never encountered nonwords, potentially biasing listeners to always interpret the ambiguous phonemes in a lexically consistent manner and therefore generating strong learning effects (Kleinschmidt & Jaeger, 2015; Scharenborg & Janse, 2013). Additionally, listeners always heard the talker's speech in the clear (i.e., absent any background noise), potentially leading to ceiling levels of learning; an open question is thus whether effects of sentence context might emerge in the context of background noise. Indeed, previous work has shown that LGPL is diminished when listeners encounter simultaneous background noise (Zhang & Samuel, 2014), and the effects of sentence context on online processing of degraded speech are most pronounced at intermediate signal-to-noise ratios (Davis, Ford, Kherif, & Johnsrude, 2011).

The present findings also raise important questions about how the timing of disambiguating phonetic information guides learning. Some theoretical accounts suggest that perceptual learning is enhanced when listeners have strong expectations about the speech sounds they will encounter *prior to* encountering any ambiguities (Davis & Johnsrude, 2007; Davis et al., 2005; Hervais-Adelman et al., 2008); that is, an operative factor for the perceptual learning of speech appears to be the *timing* of one's expectations. Indeed, despite robust phonetic recalibration when phonetic ambiguities are in word-medial or word-final positions, at least one study has found that the extent of LGPL is reduced when these ambiguities are encountered in word-initial positions (Jesse & McQueen, 2011; Norris et al., 2003). In the current study, the ambiguous phonemes were only encountered in word-medial positions, but the phonemes that preceded the ambiguous phonemes did not unequivocally predict the intended phoneme. If it were truly the case that only the phonemes that precede the critical fricative determine the degree of phonetic recalibration, then we should have observed a modulatory influence sentential context on phonetic recalibration in the current study, as the probability of the critical phoneme (/s/ or /ʃ/) based on the preceding phonemes was often less than 1. Notably, the identity of the critical phoneme could be unambiguously determined given the entire lexical context (i.e., both the phonemes that preceded the critical fricative and those that followed), and sentence context might not be expected to play a large role if lexical context unambiguously resolves the identity of the ambiguous phoneme. Thus, the fact that we did not consistently observe effects of sentence context on phonetic recalibration in the current experiments might suggest that there is some influence of the phonemes that follow the critical fricative on the degree of perceptual learning. We suggest that additional work examine the role of sentence context when phonetic ambiguities are in word-initial, word-medial and word-final positions; it is possible that effects of sentence context would be more pronounced for some positions than for others, and some initial findings suggest that sentence context may play a modulatory role on the degree of phonetic recalibration when ambiguous speech sounds are encountered in word-initial positions (Jesse & Laakso, 2015).

Finally, we suggest that additional work is needed to clarify how computational accounts of speech perception relate to perceptual learning. A sentence context may modulate how likely an upcoming word is, but it is unclear whether a listener's calculation of the prior probability for a particular phoneme necessarily incorporates sentence-level cues. What cues does a listener take advantage of in calculating these probabilities? Furthermore, while sentence context may make some upcoming phoneme more or less likely to be interpreted in a particular way, it is unclear whether this larger prior probability necessarily translates to increased perceptual learning. That is, the degree of phonetic recalibration may not necessarily be proportional to the activation of a particular phoneme. More highly specified computational accounts of the mechanisms underlying perceptual learning will allow for a more precise understanding of how the degree of learning interacts with other factors, such as sentence context.

Conclusions

In the current study, we observed robust effects of phonetic recalibration across several experiments that used a non-standard LGPL paradigm. In one such study, the extent of phonetic recalibration was found to be larger when participants read sentence contexts that predicted critical words during exposure than when they received neutral contexts; however, this effect was not observed in a follow-up replication study. Overall, our results do not provide evidence for a strong influence of sentence context on phonetic recalibration. We suggest that additional work is needed to clarify how the influence of sentence context on phonetic recalibration might be dependent on other factors, including the timing of disambiguating information, the presence of other acoustic information (e.g., noise) in the bottom-up signal, and the factors a listener considers when determining how probable a particular phoneme is *a priori*.

References

Altmann, G. T. M., & Young, D. (1993). Factors affecting adaptation to time-compressed speech. In *Third European Conference on Speech Communication and Technology* (pp. 333–336).

Boersma, P., & Weenik, D. (2017). Praat: Doing phonetics by computer.

Bonte, M., Correia, J. M., Keetels, M., Vroomen, J., & Formisano, E. (2017). Reading-induced shifts of perceptual speech representations in auditory cortex. *Scientific Reports*, *7*(1), 1–11.

Clarke-Davidson, C. M., Luce, P. A., & Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Perception and Psychophysics*, *70*(4), 604–618.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, *33*, 497–505.

Cutler, A., McQueen, J. M., Butterfield, S., & Norris, D. (2008). Prelexically-driven perceptual retuning of phoneme boundaries. *Proceedings of Interspeech*, 2056.

Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through "top–down" processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, *23*(12), 3914–3932.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, *229*(1–2), 132–147.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*(2), 222–241.

de Leeuw, J. R. (2015). jsPsych : A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121.

Drouin, J. R., & Theodore, R. M. (2018). Lexically guided perceptual learning is robust to task-based changes in listening strategy. *Journal of the Acoustical Society of America*, *144*(2), 1089–1099.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, *67*(2), 224–238.

Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, *119*(4), 1950–1953.

Ganong, W. F. (1980). Phonetic Categorization in Auditory Word Perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *14*, 29–56.

Hervais-Adelman, A., Davis, M. H., Johnsrude, I. S., & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: Effects of feedback and lexicality. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(2), 460–474.

Jesse, A., & Kaplan, E. (2019). Attentional resources contribute to the perceptual learning of talker idiosyncrasies in audiovisual speech. *Attention, Perception, & Psychophysics*, 1–14.

Jesse, A., & Laakso, S. (2015). Sentence context can guide the retuning of phonetic categories to speakers. *Poster Presented at the 56th Annual Meeting of the Psychonomic Society.*

Jesse, A., & McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic Bulletin and Review*, *18*(5), 943–950.
Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 3933–3936).
Keetels, M., Schakel, L., Bonte, M., & Vroomen, J. (2016). Phonetic recalibration of speech by text. *Attention, Perception, and Psychophysics*, *78*(3), 938–945.
Kleinschmidt, D. F., & Florian Jaeger, T. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203.
Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust Speech Perception: Recognizing the familiar, Generalizing to the similar, and adapting to the novel. *Psychological Review*, *122*(2), 148.
Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, *107*(1), 54–81.
Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, *51*(2), 141–178.
Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review*, *13*(2), 262–268.
Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*(1), 1–15.
Kraljic, T., & Samuel, A. G. (2011). Perceptual learning evidence for contextually-specific representations. *Cognition*, *121*(3), 459–465.
Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, *19*(4), 332–338.
Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Dartmouth Publishing Group.
Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement : When adults learn new words. *Cognitive Psychology*, *55*, 306–353.
Linden, S. Van, & Vroomen, J. (2007). Recalibration of phonetic categories by lipread speech versus lexical information. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1483–1494.
Liu, L., & Jaeger, T. F. (2018). Inferring causes during speech perception. *Cognition*, *174*, 55–70.
Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*(3), 543–562.
Mazzoni, D., & Dannenberg, R. B. (2015). Audacity.
McAuliffe, M., & Babel, M. (2016). Stimulus-directed attention attenuates lexically-guided perceptual learning. *The Journal of the Acoustical Society of America*, *140*(3), 1727–1738.
McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.
McQueen, J. M., Norris, D., & Cutler, A. (2006). The dynamic nature of speech perception. *Language and Speech*, *49*(1), 101–112.
Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS ONE*, *4*(11), 4–8.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42–46.

Obleser, J., & Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences*, *13*(1), 14–19.

Pitt, M. A., & Szostak, C. M. (2012). A lexically biased attentional set compensates for variable speech quality caused by pronunciation variation. *Language and Cognitive Processes*, *27*(7–8), 1225–1239.

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Saltzman, D., & Myers, E. B. (2018). Listeners are maximally flexible in updating phonetic beliefs over time. *Psychonomic Bulletin and Review*, *25*(2), 718–724.

Samuel, A. G. (2016). Lexical representations are malleable for about one second: Evidence for the non-automaticity of perceptual recalibration. *Cognitive Psychology*, *88*, 88–114.

Scharenborg, O., & Janse, E. (2013). Comparing lexically guided perceptual learning in younger and older listeners. *Attention, Perception, & Psychophysics*, *75*(3), 525–536.

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2018). afex: Analysis of Factorial Experiments. R package version 0.21-2. https://CRAN.R-project.org/package=afex.

Sjerps, M. J., & McQueen, J. M. (2010). The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 195–211.

Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Bulletin*, *30*(4), 415–433.

White, K. S., & Aslin, R. N. (2011). Adaptation to novel accents by toddlers. *Developmental Science*, *14*(2), 372–384.

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, and Psychophysics*, *79*(7), 2064–2072.

Zhang, X., & Samuel, A. G. (2014). Perceptual learning of speech under optimal and adverse conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 200–217.

Acknowledgements

This research was supported by NIH R01 DC013064, by NSF IGERT grant DGE-1144399, and by NSF Research Traineeship grant (NRT) IGE1747486. I also gratefully acknowledge the support of a Graduate Research Fellowship from the National Science Foundation. Thanks are extended to members of the UConn community for valuable discussions at various stages of this project, especially the members of the Cognitive Neuroscience of Language Lab, the Language and Brain Lab, and the Spoken Language Processing lab. Particular gratitude is given to Rachael Steiner, who was instrumental in helping to program the experiment and deploy it online. Finally, I'd like thank my advisors, James Magnuson and Emily Myers, as well as the other members of my thesis committee, Gerry Altmann and Rachel Theodore, for their ongoing support and feedback.

Appendix

Stimuli. For predictive contexts, cloze ratings for each target word are provided in parentheses.

Target	Predictive Contexts	Neutral Contexts
absent	I had gone to the bathroom when the teacher took roll, which is why he marked me… (0.9) Though he was physically present, it was readily apparent that he was mentally… (0.5)	During the board meeting, the employee advocated that the new campaign tagline be the word… I tuned out because he had babbled on for a while, but I remember he kept using the word…
accent	Even though he has lived here for many years, I can detect a bit of a foreign… (0.75) Before he could play the role of a German, the actor needed to learn how to talk with a fake German… (1)	After I knocked a jug of water on my paper, the only word I could read was… In order to win the game, I would have to get my team to correctly come up with the word…
answer	Even though he did not raise his hand, the teacher called on him for the… (0.75) I told her I didn't want more food, but my mother wouldn't take no for an… (1)	You will now hear the target item… The little girl demanded to know the meaning of the word…
Arkansas	My grandmother lives in Little Rock, which is the capital of… (0.95) Before running for president, Bill Clinton was the governor of… (0.75)	I got annoyed when the typewriter jammed as I was typing the word… I don't know why, but I am never able to remember the word…
colosseum	He dreamed of being a gladiator and fighting in the… (0.25) The Romans would congregate in a giant amphitheater called the… (0.8)	It was on the tip of his tongue, but he could not remember the word… I've never been able to make out that lyric definitively, but I've always thought the word there was…
currency	In the UK, the pound is used as the local… (0.8) If you are traveling abroad, the bank can convert your money into the local… (0.9)	I had a bizarre dream in which my friends were jumping around a fire and chanting the word… Whoever owned this book before me repeatedly underlined the word…
dinosaur	Though I love the troodon and the pteranadon, the raptor is my favorite kind of… (0.85) Long before humankind roamed the planet, the world was home to many kinds of… (0.35)	I do not believe he knows the meaning of the word… Now that I know Braille, I know that these characters make up the word…
diversity	Troubled that there were no people of color on the faculty, the college talked about ways to promote… (0.8) By referring to America as a "melting pot" of different backgrounds, he hoped to convey the idea that the country values… (0.45)	The teacher called on me and told me to define the word… I only caught the occasional word over the crackle of the PA, but I definitely heard the word…
episode	I love The Walking Dead and eagerly await every new… (0.95) I know I need to go to bed, but after that cliffhanger, I have to watch another… (0.75)	My ballpoint pen ran out of ink when I was halfway through writing the word… I do like that word, but I wonder if it might be better to use the word…
eraser	After copying the math problem incorrectly, he needed to borrow an… (0.85) We could not clean the chalkboard after you took the… (0.85)	The director berated the actor for continually forgetting the word… My one critique of that debater is that he tends to overuse the word…

insane	When the defendant was proven to be mentally ill, he was carted off to a home for the criminally… (0.9) If a defendant is mentally unable to tell right from wrong, a court might declare them to be legally… (0.75)	The five-year-old looked at me blankly when I used the word… A word that came immediately to mind was…
parasite	The literary critic argued that because Dracula feeds off other organisms without conferring any benefit to them, he can be viewed as a… (0.65) In biology, an organism that leeches off of a different organism is known as a… (0.75)	All that was written on the billboard was the word… The only reason I got a bad grade on that German exam was because I couldn't remember the word for…
peninsula	After a week in Morocco, we headed up to the Iberian… (0.7) Florida is not an island but rather a… (0.7)	Painted on the wall of the modern art museum was the word… I was losing for a while, but I took the lead in the board game when I played the word…
pregnancy	The doctor told the future mother not to drink alcohol during her… (1) You can find out the gender of the baby halfway through the mother's… (0.8)	Reading over his printed final paper, he was mortified by the highly apparent typo in the word… When I was a kid, I did not know the meaning of the word…
receipt	After he rang up my coffee order, the employee printed out my… (0.8) The boutique will allow you to return anything you bought if you bring it back with the… (0.95)	Many of the words had faded over time, but if you look carefully, you can kind of make out the word… I really cannot fathom why the only thing on the blackboard is the word…
rehearsal	The band director yelled at the drummer who came late to… (0.2) With opening night on Friday, the director told the actors they would have to work extra hard during… (0.25)	There are many great words out there, but I think my all-time favorite word is… You're mumbling, and the only word I could hear was…
adoption	If you cannot have your own child biologically, there are many children who are available for… (0.9) Before they could legally become the child's guardians, they had to file for… (0.35)	He cut random words out of the magazine, finding words like "dandelion" and… Prominent in the headline on the front page was the word…
brochure	To attract more biology majors, the college included a whole page on the biology program in the annual recruitment… (0.25) To attract new employees, the recruiter handed out new copies of a tri-fold company… (0.5)	He did not like to admit it, but he did not know the meaning of the word… Partway through reading the royal decree, the duke tripped over the word…
definition	In relatively little time, we have gone from watching TV in black and white to being able to watch TV in high… (0.9) When you use a word that many people won't know, it can be helpful to provide a… (0.45)	The microphone cut out partway through, but I think the word he was in the middle of was… The only word I could think of in the moment was…
efficient	To minimize our carbon footprint, we bought bulbs that were highly energy… (0.85) Because I want to protect the environment, I am looking for a car that is very fuel… (1)	My writing was too big, and I ran out of room to write the word… I could only catch the occasional word, but one word I definitely overheard was the word…

friendship	They did not know it when they met at the beginning of camp, but that day marked the beginning of a lifelong… (0.85) There is nothing romantic going on between the two of them; what they have is nothing more than a deep… (0.75)	My vision is not great, but I can faintly make out the word… When I was reading the article, I highlighted the word…
graduation	The valedictorian did not know what to talk about at the junior high… (0.75) He got a good enough grade on the twelfth grade exit exam that he would be allowed to walk at… (0.75)	I had trouble remembering the French word for… I don't want to harp on the point, but I found it really intriguing that the poet used the word…
handshake	The corporate executive greeted me with a firm… (0.9) I went in for a hug, but in that kind of formal meeting, it might have been more appropriate go for a… (1)	I could not believe how many times the writer reused the word… Hurriedly jotted down on the napkin was the word…
impatient	I am usually accommodating, but after waiting for five hours, even I was feeling… (0.3) The car behind me honked the moment the light turned green -- clearly, the driver was feeling rather… (0.5)	The mother was quite taken aback to learn that her two-year-old daughter already knew the word… The old man wandered the halls, looking at his feet and mumbling the word…
invitation	The bride and groom told their friends to mark their calendars before they mailed a formal wedding… (0.95) I thought we were good friends, and I was taken aback when I found out he was having a party but I hadn't gotten an… (0.85)	Written prominently in large type at the top of the paper was the word… It was unclear if there was any particular reason for him to repeatedly reiterate the word…
ocean	The mighty Amazon river flows into the Atlantic… (1) The majority of the Earth is covered by miles and miles of blue… (0.35)	The improv comedians wanted a word to riff off of, and one guy in the crowd yelled out the word… As if to belabor the point, he kept on repeating the word…
parachute	Before you can jump out of an airplane, you need to have a working… (1) The airplane deployed food and equipment to the village, delivering the load by… (0.35)	At long last, the codebreaker figured out that the letters were an anagram for the word… Written in large print on the album cover was the word…
pediatrician	A doctor for kids is called a… (1) An adult needs to go to an adult primary care doctor, but an infant needs to visit a… (0.85)	The only vocabulary word I got wrong was the word… I don't remember every word of the memo, but it definitely included the word…
permission	Before they were allowed to go on the field trip, the children needed to get a parent to grant them… (1) Before proposing to his girlfriend of many years, the man went to her father to get… (0.8)	The author entertained many options for the title of her book, eventually opting for it to be called… In her paper, the writer contemplated the meaning conveyed by the word…
pressure	When I went in for my appointment, the doctor measured my blood… (0.8) The mother reminded her daughter not to give in to peer… (1)	I could not believe how many times the writer reused the word… I've been working on getting better at calligraphy and am particularly proud of how I wrote the word…

professional	The college athlete trained very hard, hoping one day to be recruited to play as a… (0.55) If you want the job done right, don't go to an amateur; hire a… (0.8)	The word to evaluate now is the word… The director told the actor to be more emphatic, particularly on the word…
vacation	The whole family went to Hawaii for a weeklong… (0.9) After four years without a day off, the couple was ready for a lengthy… (0.8)	I need help thinking of an antonym for the word… Preoccupied by the crying baby, he broke off mid-thought and midway through the word…