Psychonomic Bulletin & Review (2026) 33:116
https://doi.org/10.3758/s13423-026-02862-2

BRIEF REPORT

Cumulative input sensitivity predicts both attenuation and stability of lexically guided perceptual learning

Shawn N. Cummings[1,2,3] · Emma C. Hodges[1] · Rachel M. Theodore[1,2,3]

Received: 23 April 2025 / Accepted: 17 January 2026

Abstract

Listeners can use lexical information to accommodate ambiguity in speech input. Some evidence suggests that lexically guided perceptual learning persists over time. However, other evidence suggests that lexically guided perceptual learning attenuates throughout the test session, consistent with learning that occurs given exposure to the test stimuli. The current study aimed to determine whether this apparent discrepancy could be resolved when viewed through the lens of the belief-updating theory of speech adaptation, which posits continuous sensitivity to statistical cues in speech input. During exposure, listeners (n = 160) heard speech sounds ambiguous between /ʃ/ and /s/ in a lexically-biasing context. At test, listeners categorized tokens from an *ashi-asi* continuum. Critically, the duration of the initial test phase was manipulated between subjects to be either brief or long. Approximately 24 hours later, all listeners completed a second test phase. Though evidence of lexically guided perceptual learning was present at all timepoints for both biasing groups, attenuation of learning was observed given continued testing. Strikingly, the position of the 24-hour time delay relative to duration of the initial test had no significant effect on learning. These results and a re-analysis of previous work align a seemingly mixed literature under a unifying account characterizing lexically guided perceptual learning as an iterative, continuous sensitivity to dynamic changes in speech input that is insensitive to time alone in the initial 24-hour learning window.

Keyword Language comprehension; speech perception; perceptual learning

Introduction

There is no prescribed reason that a certain acoustic pattern signals the meaningful unit /s/, or that a deviation in those acoustics rather signals /ʃ/. Yet, English speakers readily distinguish between the words *sign* and *shine*. Research demonstrates that prior experience guides the mapping between speech acoustics and meaning, allowing listeners to "tune in" to structured phonetic variation (e.g., Clarke & Garrett, 2004; Cummings & Theodore, 2023; Liu & Holt, 2015). This benefits listeners because acoustic speech patterns show considerable variation due to gender, age, vocal tract size, speaking rate, and idiolect (e.g., Byrd, 1992; Chodroff & Wilson, 2017; Johnson & Beckman, 1997; Newman et al., 2001; Theodore et al., 2009). Perceptual constancy therefore requires listeners to adjust expectations in response to the source and context of speech input (Kleinschmidt & Jaeger, 2015; McMurray & Jongman, 2011; Norris et al., 2003).

Novel input is mapped to meaning via a listener's current best guess given acoustic cues combined with informative context such as lexical knowledge (e.g., Cummings & Theodore, 2022, 2023; Drouin & Theodore, 2018; Norris et al., 2003; Samuel & Kraljic, 2009; Tzeng et al., 2021). When acoustic energy ambiguous between /ʃ/ and /s/replaces the natural /s/ in items such as *rehearsal*, listeners are more likely to report hearing *rehearsal*, which is a word, compared to *rehearshal*, which is not a word. Crucially, this mapping process establishes a link between the observed acoustics and recognized category such that when listeners subsequently encounter acoustic patterns ambiguous

✉ Rachel M. Theodore
rachel.theodore@uconn.edu

[1] Department of Speech, Language, and Hearing Sciences, University of Connecticut, 2 Alethia Drive, Unit 1085, Storrs, CT 06269-1085, USA

[2] Department of Psychological Sciences, University of Connecticut, 406 Babbidge Road, Unit 1020, Storrs, CT 06269-1020, USA

[3] Connecticut Institute for the Brain and Cognitive Sciences, University of Connecticut, 337 Mansfield Road, Unit 1272, Storrs, CT 06269-1272, USA

Published online: 17 March 2026

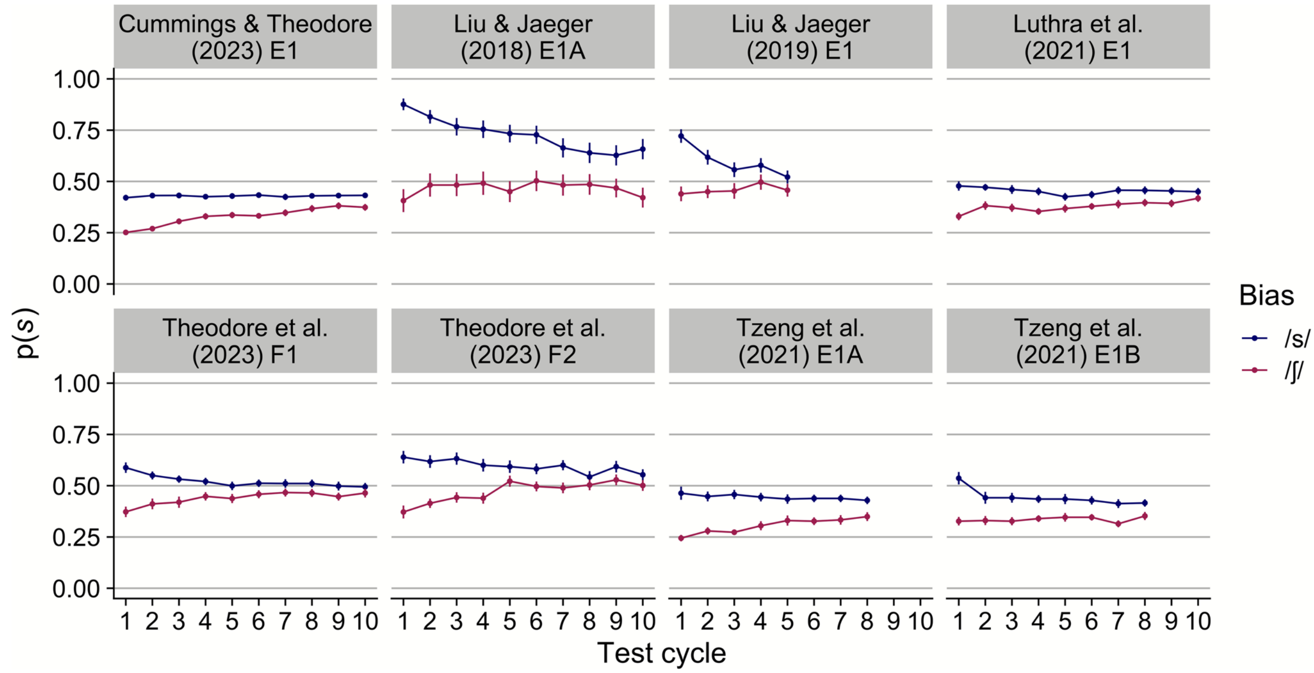


Fig. 1 Re-analysis of eight lexically guided perceptual learning experiments, which include: experiment 1 of Cummings & Theodore (2023), experiment 1 A of Liu and Jaeger (2018); experiment 1 of Liu and Jaeger (2019), experiment 1 of Luthra et al. (2021), talkers F1 and F2 in Theodore et al. (2023), and experiments 1 A and 1B in Tzeng et al. (2021). These studies were selected because trial-level data was publicly available. All experiments tested lexically guided perceptual learning for the /ʃ/-/s/contrast. In each experiment, mean proportion /s/ responses at test was calculated for each bias condition separately for each test cycle (i.e., each randomization of the test continuum). Means reflect grand means calculated over by-subject averages; error bars indicate standard error of the mean. In all experiments, the magnitude of the learning effect, as indexed by the displacement of the /s/-bias and /ʃ/-bias functions, attenuates across test cycles. As for the primary analyses, the data and analysis code to reproduce this analysis is available in the OSF repository, provided in the Open Practices Statement

between /ʃ/ and /s/ – even when lexical information is absent – they perceive the ambiguity in line with previous lexically-biasing exposure.[1]

Lexically guided perceptual learning occurs rapidly, detectable after only four lexically-biasing exposures (Cummings & Theodore, 2023), and appears to persist over time. For example, learning remains detectable 25 minutes, 12 hours, 24 hours, and even one week after exposure (Eisner & McQueen, 2006; Kraljic & Samuel, 2005; Zheng & Samuel, 2023). That lexically guided perceptual learning persists over time is remarkable given other evidence to suggest that the learning effect rapidly diminishes across a single test session (Cummings & Theodore, 2022, 2023; Giovannone & Theodore, 2021; Liu & Jaeger, 2018; Tzeng et al., 2021).

Figure 1 shows a re-analysis of eight lexically guided perceptual learning studies that examined the English /ʃ/-/s/ contrast. The learning effect is most often measured in a test phase, completed immediately after exposure, where listeners categorize members of a test continuum that extends beyond the range of ambiguity presented during exposure (e.g., 5–10 cycles of 6–9 steps of an *ashi-asi* continuum). Inspection of Figure 1 suggests that effect sizes for perceptual learning (i.e., the difference between the /s/- and /ʃ/-bias conditions, collapsed across the test continuum) are variable between experiments. This is not surprising given previous work indicating that myriad factors influence learning including stimulus properties (e.g., Cummings & Theodore, 2022) and the quantity of evidence (e.g., Cummings & Theodore, 2023). Despite variation in learning magnitude, all studies show rapid attenuation of learning over the course of test. This apparent "unlearning" can perhaps be explained as the consequence of listeners learning from the test stimuli. That is, the nature of testing for lexically guided perceptual learning provides additional exposure to the talker producing /ʃ/ and /s/. Indeed, the number of test trials often *exceeds*

[1] Though we focus on previous research for the English /ʃ/-/s/ contrast in the main text, it is important to note that lexically guided perceptual learning has been observed for a host of other speech sound categories representing fricative (Norris et al, 2003), stop (Kraljic & Samuel, 2006), vowel (Colby et al., 2018), liquid (Scharenborg et al., 2015), and tone (Mitterer et al., 2011) contrasts. Lexically guided perceptual learning has also been observed in many languages that include Dutch (Norris et al., 2003), English (Tzeng et al., 2021), German (Mitterer & Reinisch, 2017), and Mandarin (Burchfield et al., 2017).

the number of critical exposures provided during the exposure phase. Attenuation of lexically guided perceptual learning during test is consistent with other evidence suggesting that adaptation for speech perception reflects cumulative sensitivity to statistical cues in speech input (Cummings & Theodore, 2023; Tzeng et al., 2021).

A puzzle therefore emerges – how can perceptual learning appear so fragile, with effects attenuating over a brief test phase, and simultaneously be robust across longer timescales? We submit that this contradiction can be resolved by a theory of cumulative input sensitivity, such as the belief-updating theory of speech adaptation (Cummings & Theodore, 2022, 2023; Kleinschmidt & Jaeger, 2015; Tzeng et al., 2021). In this framework, speech sounds are represented as a distribution of acoustic-phonetic cues formed by long-term experience with cue-sound mappings. Talkers sample from these distributions and perception requires inferring talkers' generative distributions given listeners' cue-sound beliefs. Adaptation is the consequence of updating beliefs by integrating observed evidence with existing priors, reflecting continuous change in representational knowledge as listeners receive input. This framework accounts for adaptation driven by input with or without a supervisory signal, which may occur during the exposure and test phases of the lexically guided perceptual learning paradigm, respectively.

Predictions from this framework for observed behavior reflect intersecting parameters, including the central tendency and variance of cue-category mappings and confidence in these prior beliefs. Assume a system that represents /s/ with a modal frequency of 6000 Hz and narrow variance (e.g., 100 Hz) and has low confidence in those beliefs. Encountering a new /s/ token at 5000 Hz is predicted to change beliefs to reflect a lower modal value concomitant with increased variability. Now consider a system representing /s/ with the same modal value and low confidence, but with a high variance (e.g., 1000 Hz). Integration of the same novel /s/ token is predicted to yield a more minimal change in beliefs because the system expects greater variability *a priori*. Finally, consider that both systems instead have high confidence in the modal and variance beliefs. The belief-updating framework predicts that novel input will have a relatively weaker influence on belief updating because integration of new input for a system with high confidence is weighted lower than for a system with low confidence.

To connect the theory to behavior, this type of input – acoustic values intermediate between prototypical /s/ and /ʃ/ values in lexically-biased contexts – is exactly the kind of input presented in the lexically guided perceptual learning domain. The robust learning effect, with more of the test continuum perceived as /s/ for /s/-biased compared to /ʃ/-biased listeners, is well predicted by this theory. Additionally, it also explains atrophy of the lexically-guided learning effect (Fig. 1). Specifically, both bias groups receive the same test stimuli, providing additional input to listeners that spans the acoustic range between /ʃ/ and /s/. Because test stimuli are presented without explicit supervisory signals, listeners must provide their own label.[2] A theory of cumulative input sensitivity posits that listeners then use these self-provided labels, along with the acoustics they are tied to, to continue belief updating. That is, learning does not stop at the exposure phase but rather continues each time a listener encounters /s/ or /ʃ/. Over the test session, the composite of experienced input becomes more similar between biasing groups because both groups receive the same additional input provided by the test stimuli. Therefore, responses to subsequent input are predicted to become more similar. In addition, the relative weight of new evidence diminishes as evidence accumulates, suggesting that attenuation of the lexically guided perceptual learning effect should be most apparent early in test, when less total input has been accrued. Over the course of testing, learning effects are expected to stabilize at some floor higher than zero but lower than the effect size observed directly after exposure. Both predictions are broadly consistent with the patterns presented in Figure 1.

Though a theory of cumulative input sensitivity shows promise to reconcile the puzzle at hand, previous investigations of the stability of learning over time have not examined attenuation of learning over the course of test. Zheng and Samuel (2023) acknowledge this potential, but no formal assessment was presented. Eisner and McQueen (2006) examined learning before *and* after a 12-hour delay, finding no evidence of attenuated learning over time, but provided 78 critical lexically-biasing exposures – far more than is standard for studies in this domain – and only presented 50 trials at each test. Thus, consistent with the belief-updating theory of speech adaptation, the apparent lack of attenuation of the learning effect over time may reflect the relatively small proportion of test input compared to the total input, proportions that are generally reversed in this domain. No previous study can explicate the respective contributions of lexically-biased exposure, exposure to test stimuli, and the passage of time, which is required to optimally test the theory of cumulative input sensitivity.

Towards this goal, listeners here heard ambiguous fricatives during an exposure phase, where lexical context was used to differentially bias perception of the ambiguity as either /s/ or /ʃ/. Immediately following exposure, listeners

[2] The ideal adapter framework posits a modification of Bayes theorem to operationalize phoneme decision, in alignment with prior work on Bayesian inference in the speech domain (e.g., Clayards et al., 2008; Norris & McQueen, 2008). However, because listeners explicitly provide a label for each stimulus in test, predictions can be made about adaptation without making explicit claims about the mechanism of perception.

completed a test phase assessing perception of an /ʃ/-/s/ continuum. The length of the initial test was manipulated to be extremely brief (36 trials, the T36 condition) or relatively longer (108 trials, the T108 condition). Twenty-four hours later, all listeners completed a second test phase consisting of 108 trials. This design provides a novel dissociation of input and time and thus provides a critical test of cumulative input sensitivity. Specifically, this theory predicts that the magnitude of the learning effect will be strongest immediately after exposure and become attenuated but not extinguished over time, reflecting increased test input and not the passage of time itself.

Methods

Participants

Participants (n = 160) were recruited from the Prolific participant pool (https://www.prolific.com; Palan & Schitter, 2018) according to the following criteria: 18–35 years of age, born in and currently residing the United States, monolingual English speaker, no history of language-related disorders, and a Prolific approval rating ≥ 98. In addition, no participants completed any previous lexically guided perceptual learning study for our laboratory. The participants included 98 men and 62 women, with a mean age of 29 years (SD = 4 years). Participants were randomly assigned to one of four cells formed by crossing lexical bias during exposure (/s/-bias vs. /ʃ/-bias) and the number of test trials at the initial test session (36 trials vs. 108 trials), yielding 40 participants in each cell. An additional 19 participants were tested but excluded from analyses due to failure to pass the headphone screen (n = 9) or failure to meet the accuracy criterion during the exposure phase (≥ 80% correct, n = 10), as described further below.

We followed emerging best practices for conducting a priori power analyses for mixed effects models (e.g., Green & MacLeod, 2016; Kumle et al., 2021), which entails (1) measuring effect sizes and the covariance structure from an existing model, (2) using those parameters to simulate new data sets with different numbers of participants, trials, and/or effect sizes, (3) analyzing each simulated data set to test for statistical significance of the fixed effect(s) of interest, and (4) calculating power based on the proportion of statistically significant effects relative to all simulations. Evaluating our hypotheses requires power to detect an effect of bias and, in some cases, interactions with bias. We executed our power analyses using the simr package (Green & MacLeod, 2016) based on data from previous work (Tzeng et al., 2021). This study is well-suited for this purpose because it provides data from a standard lexically guided learning task (experiment 1) and two input manipulations that also used the standard lexically guided learning task (experiments 2 and 3). The results showed a monotonic decrease in the magnitude of the bias effect across experiments. Thus, using these data allowed us to estimate effects sizes for the main effect of bias and for bias-by-input interactions. Moreover, each experiment was conducted twice, once with each of two stimulus sets, allowing us to assess convergence of our power analyses across effect size estimates. Based on these analyses, the current experiment includes 40 participants in each between-subjects condition. This sample size yields high power (≥ 87%) to detect effect sizes observed in Tzeng et al. (2021) for the main effect of bias, experiment by bias interactions, and a conservative estimate of half of the effect size for the main effect of bias. The results of our power analyses converge with others who have used a similar approach with a different data set as the starting point for simulations (Liu & Jaeger, 2019).

Stimuli

The stimuli were identical to those used in previous work (Cummings & Theodore, 2023; Tzeng et al., 2021), to which the reader is referred for comprehensive details on the stimulus creation procedure. In brief, the stimuli consisted of 240 auditory tokens used during the exposure phase and six auditory tokens used during the test phase. All stimuli were produced by a female speaker of American English (referred to as “f1” in Tzeng et al., 2021).

The exposure stimuli consisted of 20 words containing /s/ (e.g., *rehearsal*), 20 words containing /ʃ/ (e.g., *publisher*), 60 filler words that contained no instance of either /s/ or /ʃ/ (e.g., *camera*), and 100 nonwords (e.g., *rumatik*). The specific word and nonword items follow those used in Kraljic and Samuel (Kraljic & Samuel, 2005) and were balanced with respect to stress pattern and number of syllables, among other parameters. Two versions of the /s/ and /ʃ/ words were created, one that contained the natural /s/ or /ʃ/ production, and one in which the natural /s/ or /ʃ/ production was replaced by blend of natural /s/ and /ʃ/ energy that was judged to be perceptually ambiguous. As described in Tzeng et al. (2021), the ambiguous blend was customized for each word by digitally mixing energy from a production that contained the intended fricative (i.e., the /s/ energy in *rehearsal*) with energy from a production that contained the opposite fricative (i.e., the /ʃ/ energy in *rehearshal*). In most cases, the ambiguous blend consisted of a 50-50 mixture of the two fricatives; the specific mixture for each item is reported in Tzeng et al. (2021).

The test stimuli consisted of a six-step auditory continuum that perceptually ranged from *ashi* to *asi*. Continuum steps were created by digitally mixing energy from natural /ʃ/ and /s/ productions in different weights. Step 1 consisted of 70% /ʃ/ energy and 30% /s/ energy, step 6 consisted of 20% /ʃ/ energy and 80% /s /energy, and the intermediate steps reflected an equidistant change in proportion of /ʃ/ and /s/ energy across steps.

Procedure

The experiment was a web-based study hosted by Gorilla Experiment Builder (https://gorilla.sc; Anwyl-Irvine et al., 2020) and consisted of two experimental sessions, as illustrated in Figure 2. At session 1, listeners first provided informed consent and completed a headphone screen using tasks designed to measure headphone use on web-based testing platforms (Milne et al., 2021; Woods et al., 2017). Then, listeners completed an exposure phase and a test phase. During exposure, the 200 exposure stimuli appropriate for each bias condition were presented in a randomized order for each participant. On each trial, participants were asked to indicate whether the item was a real English word by pressing a button labeled either "Yes" or No." During test, listeners completed either one block (T36 test trials condition) or three blocks (T108 test trials condition) of phonetic identification. In each block, six cycles of the test continuum were presented in a separate randomized order for each participant, and participants were asked to identify each stimulus as either *asi* or *ashi* by pressing an appropriately labeled button. Thus, listeners in the T36 test trials condition completed 36 test trials at session 1 (6 continuum steps × 6 cycles × 1 block = 36 trials) whereas listeners in the T108 test trials condition completed 108 test trials at session 1 (6 continuum steps × 6 cycles × 3 blocks = 108 trials). For both the exposure and test phases, trials were separated by 1000 ms, timed from the participant's response to the onset of the next auditory stimulus, and no feedback was provided. At session 2, all listeners completed a test phase consisting of three blocks of phonetic identification (6 continuum steps × 6 cycles × 3 blocks = 108 trials) that was procedurally identical to the session 1 test phase. Session 1 lasted approximately 15 minutes and listeners were paid $2.50 for their participation; session 2 lasted approximately 10 minutes and listeners were paid $1.67 for their participation.

Statistical analyses

Performance for the exposure phase was analyzed in terms of proportion correct lexical decision responses. A response was considered correct if participants responded "Yes" to the /s/, /ʃ/, and filler word items and "No" to the nonword items.

Performance at test was analyzed in a series of generalized linear mixed effects models (GLMMs) with the binomial response family as implemented by the glmer() function of the lme4 (Bates et al., 2015) package in R. To anticipate, the analytical strategy examines whether learning: (1) is apparent, (2) attenuates between sessions, (3) attenuates during session 1, and (4) attenuates during session 2. Accordingly, these four analyses assess learning at both coarse and fine grains. Finally, we examine whether (5) fine-grain learning attenuation between consecutive testing blocks is affected by the temporal separation of testing blocks.

The first analysis examined performance across the two test sessions for both the T36 and T108 conditions, with the goals of (1) confirming the expected perceptual learning effect and (2) determining whether, at the coarse granularity of an entire session, learning attenuates between sessions. In this model, the dependent variable was trial-level response (0 = *ashi*, 1 = *asi*). The fixed effects were continuum step, bias (/s/ = −0.5, /ʃ/ = 0.5), test trials condition (T36 = −0.5, T108 = 0.5), session (1 = −0.5, 2 = 0.5), and all interactions among these factors. Continuum step was entered into the model as a scaled/centered continuous variable that reflected percent /s/ energy in each step; all other factors were entered into the model as sum-coded contrasts (as specified above). The random effects structure consisted of random intercepts by subject and random slopes for continuum step and session by subject, which is the maximal random effects structure licensed by the experimental design.

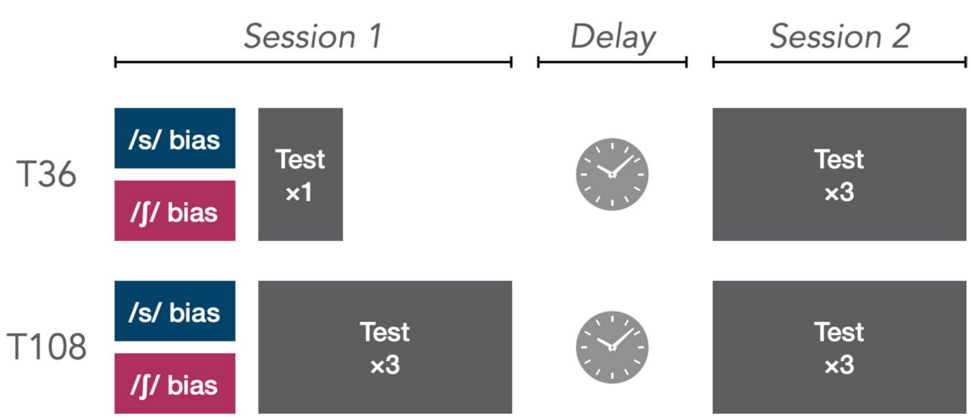


Fig. 2 Illustration of the experimental design. In session 1, listeners received either /s/- or /ʃ/-biasing lexical exposure prior to completing an initial test. Between-subjects, listeners completed either one test block totaling 36 test trials (the T36 condition) or three test blocks totaling 108 test trials (the T108 condition). After a 24-hour delay, all participants completed a second test in session 2 consisting of three test blocks totaling 108 test trials

The second analysis examined performance across the three test blocks in session 1 for the T108 test trials condition (recall that the T36 condition only received one test block in session 1), with the goal of assessing (3) more fine-grain learning attenuation in session 1. Trial-level data (0 = *ashi*, 1 = *asi*) in session 1 for the T108 condition were submitted to a GLMM as described previously. The fixed effects included continuum step, bias, block, and all interactions among these factors. Block was entered into the model as two sliding contrasts, one that compared block 1 to block 2, and one that compared block 2 to block 3. The random effects structure here and in the following two analyses consisted of random intercepts by subject and random slopes for continuum step and block by subject.

The third analysis examined performance across the three test blocks in session 2 for both the T36 and T108 conditions, assessing (4) fine-grain learning attenuation in session 2. Trial-level data (0 = *ashi*, 1 = *asi*) at session 2 for both test trials conditions were analyzed using a GLMM as described previously. The fixed effects included continuum step, bias, test trials condition, block, and all interactions among these factors. Factors were entered into the model as described previously.

The final analyses examined performance across the first two test blocks for both the T36 and T108 test trials conditions, regardless of which session they occurred in (i.e., the single test block in session 1 and the first test block of session 2 for the T36 test trials condition and the first two test blocks in session 1 for the T108 test trials condition). This analysis examines whether (5) fine-grain learning attenuation between consecutive testing blocks is affected by the temporal separation of testing blocks. The fixed effects were continuum step, bias, test trials condition, block, and all interactions among these factors. Continuum step, bias, and condition were entered into the model as described previously; block was entered into the model as a sum-coded contrast (block 1 = −0.5, block 2 = 0.5).

Results

Exposure

Table 1 shows mean lexical decision accuracy during the exposure phase. As expected, accuracy is near ceiling.

Table 1 Mean lexical decision accuracy (proportion correct) for each item type presented during the exposure phase and each test trials and bias condition. Means reflect grand means calculated over by-subject averages. Values in parentheses indicate standard deviation

		Item type			
Test trials	Bias	/s/ words	/ʃ/ words	Filler words	Nonwords
T36	SS	0.98 (0.03)	0.99 (0.03)	0.97 (0.03)	0.98 (0.03)
T36	SH	0.99 (0.02)	0.99 (0.02)	0.96 (0.04)	0.97 (0.04)
T108	SS	0.99 (0.02)	0.99 (0.01)	0.97 (0.02)	0.97 (0.04)
T108	SH	0.99 (0.04)	0.99 (0.04)	0.97 (0.03)	0.95 (0.08)

Test

Performance across sessions for both test trials conditions to examine whether (1) learning is apparent and (2) learning attenuates between sessions. Figure 3 shows test performance in each session (collapsing across test blocks and thus representing a coarse-grain analysis). Visual inspection suggests that both conditions show a robust learning effect in session 1, reflecting more *asi* responses for the /s/-bias compared to the /ʃ/-bias group; moreover, the magnitude of the learning effect appears larger for the T36 compared to the T108 condition. In session 2, a learning effect of similar magnitude is present for both conditions. The learning effect appears to wane across sessions for the T36 but not the T108 condition.

As shown in Table 2, the GLMM revealed a main effect of continuum step, reflecting an increase in *asi* responses across steps, and a main effect of session, reflecting an increase in *asi* responses from session 1 to session 2. A robust influence of bias was observed, confirming the learning effect. There was an interaction between step and bias, indicating that the magnitude of the bias effect differed across steps. Critically, the bias × session interaction was reliable, as was the bias × condition × session interaction. To explicate the 3-way interaction, GLMMs were constructed (following the procedure outlined for the omnibus model) to examine the bias × session interaction within each condition and the bias × condition interaction within each session. The learning effect diminished from session 1 to session 2 for the T36 condition ($\widehat{\beta} = -1.129$, $SE = 0.281$, $z = -4.021$, $p < 0.001$); no such attenuation of learning across sessions was observed for the T108 condition ($\widehat{\beta} = -0.219$, $SE = 0.258$, $z = -0.851$, $p = 0.395$). At session 1, there was weak evidence to suggest a larger learning effect for the T36 condition compared to the T108 condition ($\widehat{\beta} = -1.012$, $SE = 0.531$, $z = -1.906$, $p = 0.057$), and no evidence to suggest that the learning effect differed between the two conditions at session 2 ($\widehat{\beta} = -0.101$, $SE = 0.501$, $z = -0.203$, $p = 0.839$).

Performance across test blocks at session 1 for the T108 test trials condition to examine whether (3) learning attenuates during session 1. The top right panel of Figure 4 shows performance for each of the three test blocks in session 1 for the T108 condition. Visual inspection suggests that the learning effect, though robust, wanes across the three blocks. The GLMM, shown in Table 3, revealed an interaction between bias and block for the block 1 vs. block 2 contrast, indicating that the bias effect diminished from block 1 to block 2. The bias × block interaction was not reliable for the block 2 vs.

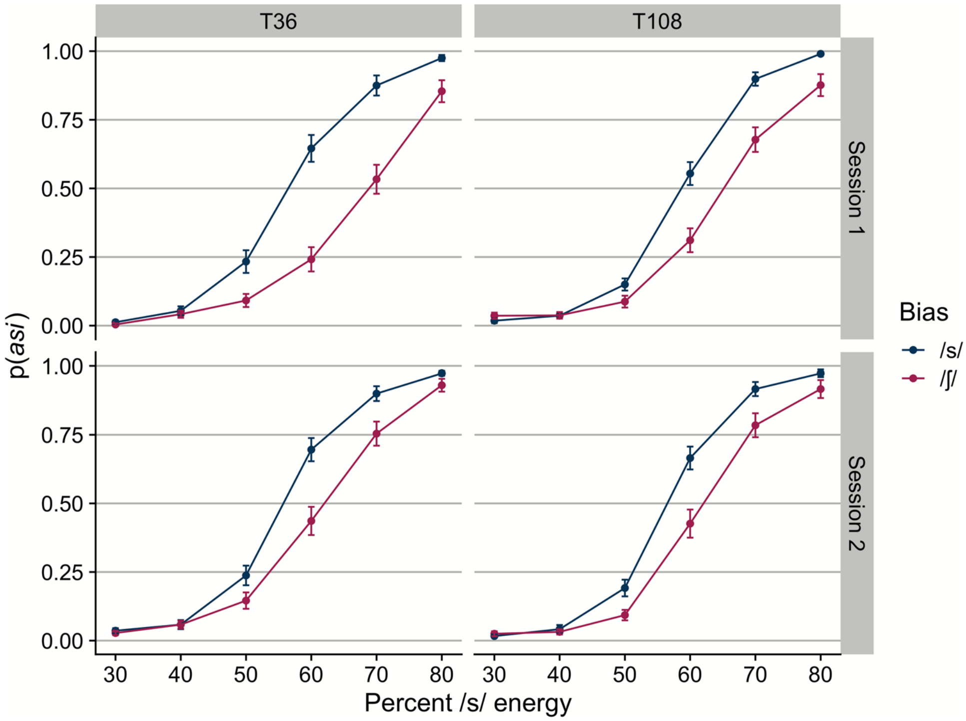


Fig. 3 Mean proportion *asi* responses at test for each biasing group as a function of continuum step, lexical bias, test trials condition, and session (collapsing over test blocks within each session). Step is represented by the proportion of /s/ energy in each step. Means reflect grand means calculated over by-subject averages; error bars indicate standard error of the mean

Table 2 Results of the GLMM to examine performance across sessions for both test trials conditions. The model contained 28, 800 observations in total across 160 listeners. As described in the main text, continuum step (Step) was entered into the model as a continuous variable; all other factors were entered into the model as sum-coded contrasts

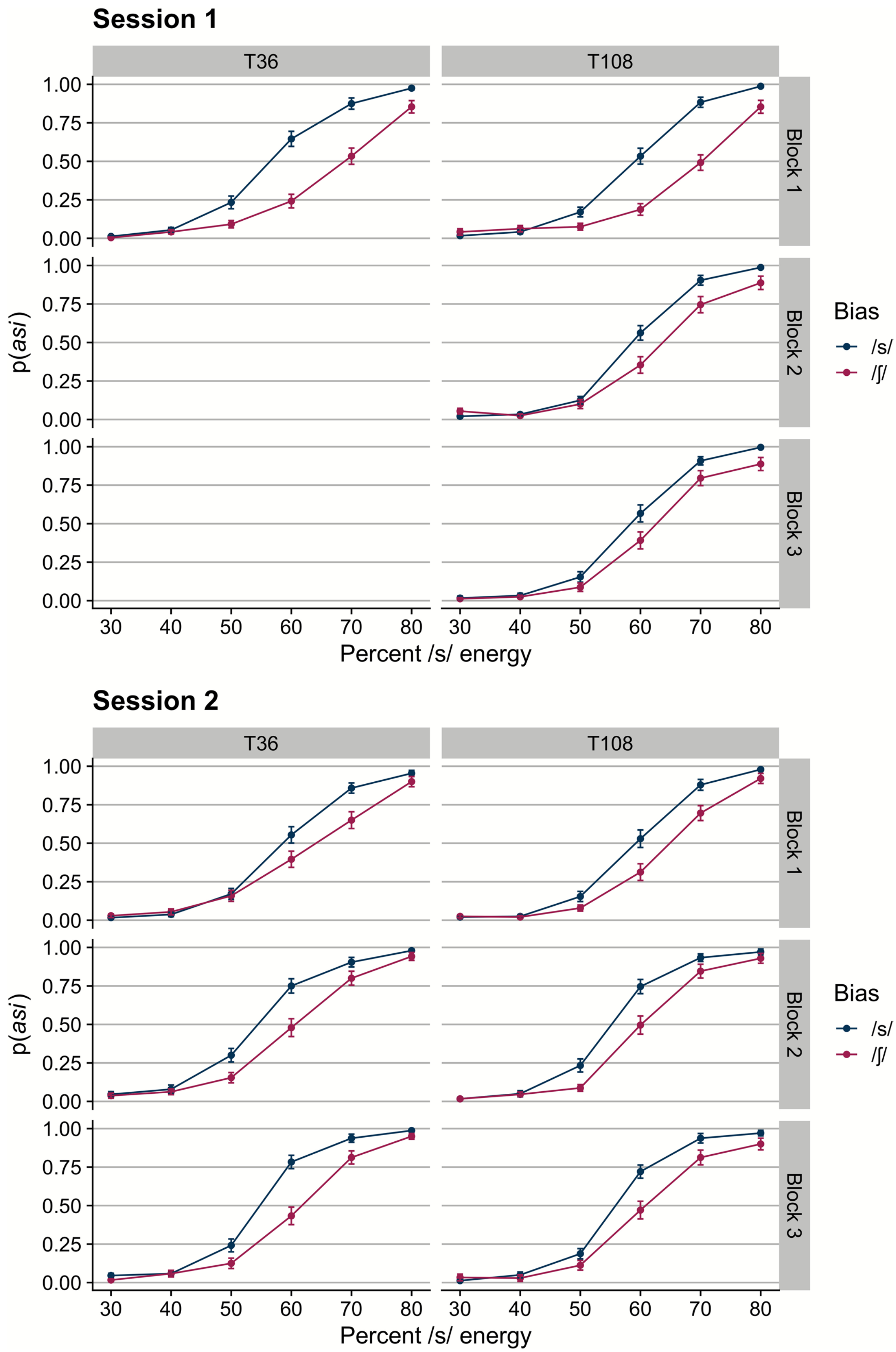
Fixed effect	$\widehat{\beta}$	*SE*	*z*	*p*
(Intercept)	−1.229	0.121	−10.150	< 0.001
Step	4.090	0.146	27.919	< 0.001
Bias	1.440	0.240	5.993	< 0.001
Condition	−0.114	0.240	−0.474	0.636
Session	0.560	0.099	5.640	< 0.001
Step × Bias	0.649	0.285	2.274	0.023
Step × Condition	0.196	0.285	0.687	0.492
Bias × Condition	−0.496	0.477	−1.040	0.299
Step × Session	0.130	0.080	1.628	0.104
Bias × Session	−0.408	0.196	−2.077	0.038
Condition × Session	−0.320	0.196	−1.637	0.102
Step × Bias × Condition	−0.072	0.564	−0.128	0.898
Step × Bias × Session	−0.500	0.152	−3.278	0.001
Step × Condition × Session	0.142	0.153	0.927	0.354
Bias × Condition × Session	0.967	0.391	2.477	0.013
Step × Bias × Condition × Session	−0.210	0.304	−0.689	0.491

block 3 contrast, providing no evidence to suggest that learning continued to attenuate beyond block 2.

Performance across test blocks at session 2 for both test trials conditions to examine whether (4) learning attenuates during session 2. The bottom panel of Figure 4 shows performance across the three test blocks in session 2 separately for each condition. As shown in Table 4, there was a significant interaction between bias and block for the block 1 vs. block 2 contrast, with the direction of the beta estimate indicating that the learning effect slightly increased in block 2 compared to block 1. The bias × block interaction for the block 2 vs. block 3 contrast was not reliable, nor was the bias × condition × block interaction for either block contrast.[3]

Performance across the first two test blocks for both test trials conditions to examine whether (5) fine-grain learning attenuation between consecutive testing blocks is affected by

[3] The results of the session 1 analysis for the T108 condition showed an attenuation of learning from block 1 to block 2, whereas the results of the session 2 analysis (which included both conditions) provided no evidence of attenuation of learning across test blocks. A GLMM was constructed for the T108 condition that included continuum step, bias, block, session, and all interactions among these factors as fixed effects to confirm the bias × block × session interaction. The model results showed that the three-way interaction was reliable for the block 1 vs. block 2 contrast ($\widehat{\beta}$ = 0.876, *SE* = 0.297, *z* = 2.950, *p* = 0.003). This analysis can be viewed by executing the script provided in the OSF repository (see Open Practices Statement).


Session 1
T36
T108
Block 1
Block 2
Block 3
p(asi)
Bias
/s/
/ʃ/
1.00
0.75
0.50
0.25
0.00
30 40 50 60 70 80
Percent /s/ energy
Session 2
T36
T108
Block 1
Block 2
Block 3
p(asi)
Bias
/s/
/ʃ/
Percent /s/ energy

◂**Fig. 4** Mean proportion *asi* responses at test for each biasing group as a function of continuum step, lexical bias, test trials condition, and test block for each session. Continuum step is represented by the proportion of /s/ energy in each step. Means reflect grand means calculated over by-subject averages; error bars indicate standard error of the mean

the temporal separation of testing blocks. The final analysis examined performance across the first two test blocks for both conditions, visualized in Figure 4, regardless of in which session they occurred. Specifically, this analysis examined the first block of sessions 1 and 2 for the T36 condition and blocks 1 and 2 in session 1 for the T108 condition. As shown in Table 5, there was a reliable interaction between bias and block, with the direction of the interaction indicating that the learning effect attenuated across the first two blocks. The interaction between bias, condition, and block was not reliable, providing no evidence that the attenuation of learning across the first two blocks differed between conditions.

Discussion

On the one hand, lexically guided perceptual learning rapidly diminishes over the course of a brief test session (Cummings & Theodore, 2023; Liu & Jaeger, 2018). On the other hand, perceptual learning persists over time (Eisner & McQueen, 2006; Zheng & Samuel, 2023). Here we assessed whether this apparent discrepancy could be explained by a theory of cumulative input sensitivity (Cummings & Theodore, 2023; Kleinschmidt & Jaeger, 2015; Tzeng et al., 2021). Listeners heard lexically-biased input and then completed two tests separated by 24 hours. Between subjects, the length of the initial test was manipulated to include either 36 or 108 trials.

In the aggregate, learning was reduced after 24 hours for the T36 condition, but not for the T108 condition. We interpret this result as evidence that both conditions initially achieved the same degree of learning, which then attenuated to different degrees. Of course, this pattern is less informative if learning initially differed between the two conditions. A post-hoc analysis of the first test block of session 1 found no evidence to suggest that learning in the initial block differed across conditions ($\widehat{\beta} = -0.252$, $SE = 0.609$, $z = -0.413$, $p = 0.679$). This analysis can be viewed by executing the script provided in the OSF repository (see Open Practices Statement).

The T108 condition replicates the closest parallel to Eisner and McQueen (2006) in that both found stable learning over a 12-hour delay. In contrast, the T36 condition – on the surface – demonstrated more transient learning. This discrepancy is resolved when results are considered at a finer grain. In session 1, learning significantly diminished in the T108 condition between the initial 36 test trials and subsequent trials. This initial attenuation contributes to the lack of difference across sessions for the T108 condition. Importantly, learning did not diminish across testing blocks within session 2 for either the T108 or T36 condition. In addition, analysis of the first two test blocks, that spanned a 24-hour delay for the T36 but not the T108 condition, showed no reliable differences between the two conditions.

These results, and the reanalysis of extant data presented in Figure 1, align with predictions made by a theory of cumulative input sensitivity (Kleinschmidt & Jaeger, 2015; Cummings & Theodore, 2023; Tzeng et al., 2021). Specifically, listeners entered the experiment with beliefs reflecting typical acoustic patterns for /s/ and /ʃ/. Lexically informed ambiguous input promoted shifts in those beliefs that differed between biasing groups. The differential beliefs were most apparent at initial test. However, over the course of test, listeners continued to update their beliefs. Because the test input was identical for both biasing groups, cumulative experience (and thus test behavior) homogenized over test. Finally, because increased input results in a concomitant increase in belief confidence, changes in beliefs are predicted to minimize given increased input. Specifically, the relative weight of new evidence diminishes as evidence accumulates, which predicts that given continued testing, learning effects will stabilize at some floor higher than zero but lower than the effect size observed directly after exposure, as observed in the current data.

Limitations of course exist. For example, the current study was not designed to preclude the influence of alternative determinants of adaptation such as sleep consolidation, memory decay, and interference from other speech input. In addition, factors unrelated to perceptual learning including decision-level biases may have contributed to performance. Here we consider their relevance to the current findings and as areas for future research. Though we discuss each separately, we note that multiple factors could be operating simultaneously, making it difficult to fully disentangle their relative contributions.

A stabilizing effect of sleep-based memory consolidation, orthogonal to cumulative input sensitivity, well characterizes results for the T108 condition. All subjects had a night of rest between biasing exposure and the second test session, and memory encoding processes during sleep have been implicated as important for stabilizing speech adaptation (Drouin et al., 2023; Earle & Myers, 2014). Consistent with this view, the learning effect in the T108 condition attenuated through session 1 but remained stable after sleep. However, a different pattern was observed for the T36 condition. Specifically, learning attenuation was observed from the single test block in session 1 to the first test block of session 2, the latter of which occurred after sleep. Collectively, these

Table 3 Results of the GLMM to examine performance across test blocks at session 1 for the T108 test trials condition. The model contained 8,640 observations in total across 80 listeners. As described in the main text, continuum step was entered into the model as a continuous variable. Bias was entered as a sum-coded contrast, and block was entered as two sliding contrasts (i.e., block 1 vs. block 2, block 2 vs. block 3)

Fixed effect	$\hat{\beta}$	*SE*	*z*	*p*
(Intercept)	−1.719	0.199	−8.650	< 0.001
Step	4.511	0.253	17.829	< 0.001
Bias	1.545	0.391	3.948	< 0.001
Block: 1 vs. 2	0.543	0.164	3.318	0.001
Block: 2 vs. 3	−0.025	0.137	−0.179	0.858
Step × Bias	0.326	0.485	0.673	0.501
Step × Block: 1 vs. 2	0.305	0.158	1.927	0.054
Step × Block: 2 vs. 3	0.462	0.181	2.550	0.011
Bias × Block: 1 vs. 2	−0.956	0.313	−3.051	0.002
Bias × Block: 2 vs. 3	0.182	0.259	0.703	0.482
Step × Bias × Block: 1 vs. 2	−0.220	0.296	−0.744	0.457
Step × Bias × Block: 2 vs. 3	−0.436	0.324	−1.345	0.179

findings are more consistent with a theory of cumulative input sensitivity than sleep-based memory consolidation.

While memory is clearly an important facet of learning, and its interaction with time makes clear predictions towards the decay of adaptation effects as listeners forget input, we do not find support for this occurrence within a 24-hour window. This contrasts with previous findings suggesting that perceptual learning is diminished when tested 24 hours later compared to immediately after exposure (Zheng & Samuel, 2023). However, in their study, no difference in the magnitude of the learning effect was observed when test was delayed by 24 hours versus a week, which would be expected by memory decay.

Though interfering input was not explicitly presented in the current experiment, intervening exposure may interfere with perceptual learning (Earle & Myers, 2015) and it is likely that most subjects heard speech between sessions that included natural variants of /ʃ/ and /s/. However, for lexically guided perceptual learning specifically, interference has only been observed when intervening input is produced by the same talker as exposure and test (e.g., Kraljic

Table 4 Results of the GLMM to examine performance across test blocks at session 2 for both test trials conditions. The model contained 17,280 observations in total across 160 listeners. As described in the main text, continuum step was entered into the model as a continuous variable. Bias and condition were entered as sum-coded contrasts, and block was entered as two sliding contrasts (i.e., block 1 vs. block 2, block 2 vs. block 3)

Fixed effect	$\hat{\beta}$	*SE*	*z*	*p*
(Intercept)	−1.000	0.135	−7.435	< 0.001
Step	4.584	0.178	25.731	< 0.001
Bias	1.246	0.266	4.677	< 0.001
Condition	−0.344	0.266	−1.295	0.195
Block: 1 vs. 2	0.875	0.092	9.526	< 0.001
Block: 2 vs. 3	−0.038	0.092	−0.409	0.683
Step × Bias	0.665	0.340	1.952	0.051
Step × Condition	0.496	0.340	1.460	0.144
Bias × Condition	−0.140	0.531	−0.263	0.792
Step × Block: 1 vs. 2	0.115	0.118	0.979	0.327
Step × Block: 2 vs. 3	0.029	0.119	0.243	0.808
Bias × Block: 1 vs. 2	0.343	0.172	2.001	0.045
Bias × Block: 2 vs. 3	0.026	0.177	0.150	0.881
Condition × Block: 1 vs. 2	0.123	0.170	0.723	0.470
Condition × Block: 2 vs. 3	−0.009	0.175	−0.052	0.959
Step × Bias × Condition	−0.057	0.678	−0.084	0.933
Step × Bias × Block: 1 vs. 2	−0.424	0.209	−2.030	0.042
Step × Bias × Block: 2 vs. 3	0.350	0.214	1.636	0.102
Step × Condition × Block: 1 vs. 2	0.062	0.209	0.295	0.768
Step × Condition × Block: 2 vs. 3	−0.514	0.215	−2.395	0.017
Bias × Condition × Block: 1 vs. 2	−0.433	0.339	−1.279	0.201
Bias × Condition × Block: 2 vs. 3	−0.321	0.349	−0.921	0.357
Step × Bias x Condition × Block: 1 vs. 2	0.271	0.418	0.648	0.517
Step × Bias x Condition × Block: 2 vs. 3	0.290	0.429	0.676	0.499

Table 5 Results of the GLMM to examine performance across the first two test blocks for both test trials conditions. The model contained 11,520 observations in total across 160 listeners. As described in the main text, continuum step was entered into the model as a continuous variable; bias, condition, and block were entered as sum-coded contrasts

Fixed effect	$\hat{\beta}$	*SE*	*z*	*p*
(Intercept)	−1.626	0.145	−11.222	< 0.001
Step	4.068	0.172	23.717	< 0.001
Bias	1.551	0.282	5.507	< 0.001
Condition	−0.262	0.279	−0.938	0.348
Block	0.372	0.113	3.277	0.001
Step × Bias	0.633	0.316	2.003	0.045
Step × Condition	0.234	0.314	0.745	0.456
Bias × Condition	0.104	0.557	0.186	0.852
Step × Block	0.132	0.118	1.121	0.262
Bias × Block	−1.133	0.218	−5.198	< 0.001
Condition × Block	0.202	0.216	0.939	0.348
Step × Bias × Condition	−0.169	0.628	−0.269	0.788
Step × Bias × Block	−0.080	0.217	−0.370	0.711
Step × Condition × Block	0.397	0.216	1.834	0.067
Bias × Condition × Block	0.473	0.430	1.100	0.272
Step × Bias × Condition × Block	−0.292	0.433	−0.674	0.500

& Samuel, 2005), consistent with other evidence suggesting that these learning effects may be talker-specific (Cummings & Theodore, 2022; Kraljic & Samuel, 2007). Collectively, these findings suggest that interference is unlikely to be the primary locus of the recalibration effects observed in the current study.

Finally, it is important to note that factors unrelated to perceptual learning may have contributed to performance, including attentional lapses. In addition, participants' responses may have been conditioned by their awareness of the bounds of the acoustic space (Yamada & Tohkura, 1992). Moreover, decision-level factors may have biased listeners to use the two response options equally often. If this bias accumulated in strength over time, it could in broad strokes approximate the pattern observed in Figure 1 and in the current study. However, this factor alone does not predict the fact that learning effects seem to stabilize at a level lower than initial test but greater than zero, which is observed in the reanalysis of prior work (Figure 1) and the current study, and predicted by the ideal adapter framework. Thus, we conclude that the results are consistent with a belief-updating account, without claiming it is the only factor that may influence performance.

This investigation adds to a growing evidence base indicating that perceptual learning is not a binary outcome, but rather a continuous, gradient, and iterative process (Tzeng et al., 2021; Cummings & Theodore, 2023). Listeners are sensitive to initial lexically-biasing input yet continue to refine their expectations of a talker's speech, even when additional input comes from the stimuli intended to test for lexically-informed learning. The current findings suggest that not all evidence appears equally weighted, and that how new input affects subsequent categorization is sensitive to both the difference between new input and prior expectations and the extent of previously encountered input from the same talker. Though future research is needed to fully explicate the role of sleep-based memory consolidation, memory decay, and interfering speech input for perceptual learning, the current results suggest lexically guided perceptual learning is consistent with the theory that listeners exhibit cumulative sensitivity to speech input over time (Cummings & Theodore, 2022; Kleinschmidt & Jaeger, 2015; Tzeng et al., 2021; Xie et al., 2023).

Acknowledgements This research was supported by U.S. National Science Foundation (BCS) grant 2146885 to RMT and by U.S. National Science Foundation (DGE) grant 1735225 to the University of Connecticut. The views expressed here reflect those of the authors and not the U.S. National Science Foundation.

Funding This research was supported by U.S. National Science Foundation (BCS) grant 2146885 to RMT and by U.S. National Science Foundation (DGE) grant 1735225 to the University of Connecticut. The views expressed here reflect those of the authors and not the U.S. National Science Foundation.

Data availability The data, materials, and analysis code for all experiments are available at https://osf.io/nwxmb/.

Code availability The data, materials, and analysis code for all experiments are available at https://osf.io/nwxmb/.

Declarations

Ethics approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study was approved by the University of Connecticut Institutional Review Board under protocol X22-0038.

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent to publish Not applicable; no identifying participant information is included in the manuscript.

Open practices statement The data, materials, and analysis code for all experiments are available at https://osf.io/nwxmb/. The experiments were not preregistered.

Conflicts of interest/competing interests The authors have no competing interests to declare that are relevant to the content of this article.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388–407.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Burchfield, L. A., Luk, S.-H., Antoniou, M., & Cutler, A. (2017). Lexically guided perceptual learning in Mandarin Chinese. *Interspeech, 2017*, 576–580.

Byrd, D. (1992). Preliminary results on speaker-dependent variation in the TIMIT database. *The Journal of the Acoustical Society of America, 92*(1), 593–596.

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics, 61*, 30–47.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America, 116*(6), 3647–3658.

Colby, S., Clayards, M., & Baum, S. (2018). The role of lexical status and individual differences for perceptual learning in younger and older adults. *Journal of Speech, Language, and Hearing Research, 61*(8), 1855–1874.

Cummings, S. N., & Theodore, R. M. (2022). Perceptual learning of multiple talkers: Determinants, characteristics, and limitations. *Attention, Perception & Psychophysics, 84*, 2335–2359.

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*, Article 105404.

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

Drouin, J. R., Zysk, V. A., Myers, E. B., & Theodore, R. M. (2023). Sleep-based memory consolidation stabilizes perceptual learning of noise-vocoded speech. *Journal of Speech, Language, and Hearing Research, 66*(2), 720–734. https://doi.org/10.1044/2022

Earle, F. S., & Myers, E. B. (2014). Overnight consolidation promotes generalization across talkers in the identification of non-native speech sounds. *The Journal of the Acoustical Society of America, 137*(1), Article EL91–EL97. https://doi.org/10.1121/1.4903918

Earle, F. S., & Myers, E. B. (2015). Sleep and native language interference affect non-native speech sound learning. *Journal of Experimental Psychology. Human Perception and Performance, 41*(6), 1680–1695. https://doi.org/10.1037/xhp0000113

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.

Giovannone, N., & Theodore, R. M. (2021). Individual differences in lexical contributions to speech perception. *Journal of Speech, Language, and Hearing Research, 64*(3), 707–724.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498.

Johnson, K., & Beckman, M. E. (1997). Production and perception of individual speaking styles. In *Working Papers in Linguistics* (Vol. 50, pp. 115–125). Ohio State University, Department of Linguistics.

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), 148–203.

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 & 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.

Kumle, L., Vo, M.L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods, 53*(6), 2528–2543.

Liu, L., & Jaeger, T. F. (2018). Inferring causes during speech perception. *Cognition, 174*, 55–70.

Liu, L., & Jaeger, T. F. (2019). Talker-specifi c pronunciation or speech error? Discounting (or not) atypical pronunciations during speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 45*(12), 1562–1588. https://doi.org/10.1037/xhp0000693

Liu, R., & Holt, L. L. (2015). Dimension-based statistical learning of vowels. *Journal of Experimental Psychology. Human Perception and Performance, 41*(6), 1783–1798.

Luthra, S., Mechtenberg, H., & Myers, E. B. (2021). Perceptual learning of multiple talkers requires additional exposure. *Attention, Perception, & Psychophysics, 83*(5), 2217–2228.

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), 219–246.

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch. *Behavior Research Methods, 53*(4), 1551–1562.

Mitterer, H., Chen, Y., & Zhou, X. (2011). Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm. *Cognitive Science, 35*(1), 184–197.

Mitterer, H., & Reinisch, E. (2017). Surface forms trump underlying representations in functional generalisations in speech perception: The case of German devoiced stops. *Language, Cognition and Neuroscience, 32*(9), 1133–1147.

Newman, R. S., Clouse, S. A., & Burnham, J. L. (2001). The perceptual consequences of within-talker variability in fricative production. *The Journal of the Acoustical Society of America, 109*(3), 1181–1196.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*(2), 204–238. https://doi.org/10.1016/S0010-0285(03)00006-9

Palan, S., & Schitter, C. (2018). Prolific. Ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22–27.

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception & Psychophysics, 71*(6), 1207–1218. https://doi.org/10.3758/app.71.6.1207

Scharenborg, O., Weber, A., & Janse, E. (2015). The role of attentional abilities in lexically guided perceptual learning by older listeners. *Attention, Perception & Psychophysics, 77*(2), 493–507.

Theodore, R. M., Cummings, S. N., Shattuck-Hufnagel, S., & Choi, J. Y. (2023). Linking lexically guided perceptual learning to statistical patterns in speech input. Poster presented at the 184th meeting of the Acoustical Society of America, Chicago, Illinois.

Theodore, R. M., Miller, J. L., & DeSteno, D. (2009). Individual talker differences in voice-onset-time: Contextual influences. *The Journal of the Acoustical Society of America, 125*(6), 3974–3982. https://doi.org/10.1121/1.3106131

Tzeng, C. Y., Nygaard, L. C., & Theodore, R. M. (2021). A second chance for a first impression: Sensitivity to cumulative input statistics for lexically guided perceptual learning. *Psychonomic Bulletin & Review, 28*, 1003–1014.

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

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. https://doi.org/10.1016/j.cortex.2023.05.003

Yamada, R. A., & Tohkura, Y. I. (1992). The eff ects of experimental variables on the perception of American English/r/and/l/by Japanese listeners. *Perception & Psychophysics, 52*(4), 376–392.

Zheng, Y., & Samuel, A. G. (2023). Flexibility and stability of speech sounds: The time course of lexically-driven recalibration. *Journal of Phonetics, 97*, Article 101222.