

Short-term exposure alters adult listeners' perception of segmental phonotactics

Jeremy Steffman^{1,a)}  and Megha Sundara² 

¹Linguistics and English Language, The University of Edinburgh, Edinburgh, EH8 9AD, United Kingdom

²Linguistics, University of California, Los Angeles, California 90095, USA

jeremy.steffman@ed.ac.uk, megha.sundara@humnet.ucla.edu

Abstract: This study evaluates the malleability of adults' perception of probabilistic phonotactic (biphone) probabilities, building on a body of literature on statistical phonotactic learning. It was first replicated that listeners categorize phonetic continua as sounds that create higher-probability sequences in their native language. Listeners were also exposed to skewed distributions of biphone contexts, which resulted in the enhancement or reversal of these effects. Thus, listeners dynamically update biphone probabilities (BPs) and bring this to bear on perception of ambiguous acoustic information. These effects can override long-term BP effects rooted in native language experience. © 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

[Editor: Douglas D. O'Shaughnessy]

<https://doi.org/10.1121/10.0023900>

Received: 7 July 2023 **Accepted:** 21 November 2023 **Published Online:** 12 December 2023

1. Introduction

A central property of speech perception is context-sensitivity, such as sensitivity to sound patterns in words (e.g., [Luce and Pisoni, 1998](#)). Speech perception is also highly adaptable: adults can learn many novel contingencies based on short-term experience. These include novel syllable sequencing restrictions (e.g., [Saffran et al., 1996](#)), segment sequencing restrictions (e.g., [Warker and Dell, 2006](#)) cue distributions (e.g., [Clayards et al., 2008](#)), and co-occurrence patterns between two acoustics cues (e.g., [Schertz and Clare, 2020](#)). Additionally, adults' perception of acoustic cues can be modified as a result of short-term experience with altered distributions of lexical or sentential contexts (e.g., [Bushong and Jaeger, 2019](#); [Norris et al., 2003](#)). Context effects such as these and the extent to which listeners learn from context provide a window into how and when listeners dynamically update hypotheses about what they are hearing. In this study, we evaluate the extent to which altering the distribution of phonological (segmental) contexts in which a sound occurs alters adults' perception of ambiguous acoustic information, testing if and how learning for segmental context co-occurrences is flexible.

Perhaps the most well-known context effect in the speech perception literature is the Ganong effect: listeners are biased to categorize ambiguous segments from word-nonword continua to create words (e.g., [Ganong, 1980](#)). For example, listeners are biased to perceive ambiguous voice onset time (VOT) values as /g/ in the frame _ift, creating the word "gift" (not the nonword "kift"). Likewise, in the frame _iss, listeners are biased to hear the same ambiguous stimuli as /k/, creating the word "kiss" (not "giss"). Adult listeners' categorization of ambiguous segments is also biased by the preceding and following context above the level of the word. Listeners are more likely to identify stimuli from a *tent-dent* continuum as /t/ when the target words are embedded in a sentence like "When the _ent in the forest was well camouflaged..." compared to a sentence like "When the _ent in the fender was well camouflaged..." ([Connine et al., 1991](#); see also [Borsky et al., 1998](#); [Bushong, 2020](#)). Most relevant to the present study, phonetic categorization is also biased by the preceding and following sub-lexical context: listeners categorize ambiguous acoustic information to create segment sequences that are more likely in their native language ([Pitt and McQueen, 1998](#); [Steffman and Sundara, 2023](#)).

Importantly, the extent to which listeners rely on or weight contextual information is not static. Instead, it can be updated, reflecting an adaptive perceptual system is modified by short-term exposure. A well-studied case of exposure altering lexical context-based effects is lexically guided perceptual retuning (e.g., [Cummings and Theodore, 2022](#); [Norris et al., 2003](#); [Samuel and Kraljic, 2009](#)). In the classic example, during exposure, listeners hear a sound ambiguous between [f] and [s] in words where only one segment creates a word, e.g., bookshel[?] for /fl/. After this exposure phase, listeners' perceptual boundaries are shifted: ambiguous stimuli are more likely to be perceived as the phoneme that created a word during the exposure phase (although these effects diminish with repeated categorization of postexposure stimuli; [Liu and Jaeger, 2018](#)). Exposure has also been shown to alter context effects at the sentence level ([Bushong and Jaeger, 2019](#)). [Bushong and Jaeger \(2019\)](#) demonstrate that phonetic categorization of ambiguous segments can be altered as a function

^{a)} Author to whom correspondence should be addressed.

of the co-occurrence probabilities between a given step on a stop voicing (VOT) continuum from /t/ to /d/ and semantic context (of the sort in Connine, 1987). Typically, as in Connine (1987), each continuum step is presented an equal number of times in each context. We refer to this as a “flat” distribution. Bushong and Jaeger (2019) created a second condition in which listeners were presented with more contexts that favored /t/ with longer, more /t/-like, VOT. Similarly, contexts which favored /d/ appeared more often with shorter, more /d/-like, VOT. This distributional manipulation enhanced the context effect, even for ambiguous stimuli on the continuum which were presented an equal number of times in each context. In this paper, we evaluated whether adults are able to update sub-lexical context effects in perception of phonetic cues given short-term exposure in the laboratory, and using a manipulation similar to the one in Bushong and Jaeger (2019).

Research from a variety of domains shows that adults can learn novel sub-lexical regularities with short-term exposure. Phonotactic learning is well documented in speech error research, where speakers produce speech errors which respect phonotactic constraints that have been learned in the laboratory (e.g., Warker and Dell, 2006, 2015), whether these constraints are categorical (Taylor and Houghton 2005; Warker, 2013) or gradient/probabilistic (Goldrick and Larson, 2008). In the domain of speech perception, there is also some evidence that short-term exposure can modify adults’ perception of sub-lexical regularities in their native language. For example, Cutler *et al.* (2008) provide some preliminary evidence that adults may update categorical sub-lexical (phonotactic) context effects in the perception of ambiguous acoustic information. In an exposure phase, listeners heard a sound which was ambiguous between [f] and [s] in nonword frames in which it either preceded /n/ or /r/. In English, /sn/ is a licit sequence but /fn/ is not. Similarly, /fr/ is a licit sequence and /sr/ is not. Listeners used this sub-lexical information to retune their perception of an ambiguous stimulus: they were biased to categorize the stimulus as the segment that creates a licit sequence in English.

Our approach is similar to the one taken by Cutler *et al.* (2008), although, unlike them, we explore graded variation in co-occurrence probability. We experimentally manipulated distributional information to alter segment co-occurrence probabilities. The distributional manipulation preceded or followed the ambiguous segment and either enhanced or overrode English segment co-occurrence probabilities.

2. Methods

All data, models, model summaries, and code for analysis can be found hosted on the Open Science Framework (OSF).¹

2.1 Stimuli

Phonological (sequential) co-occurrences are often measured using biphone probabilities (BPs): the probability that two phones occur in sequence. This is often computed over a lexicon, representing what listeners know about a particular language. We used two continua, referred to as CV (continuum 1) and VC (continuum 2), to manipulate BP in a two-phone sequence. Table 1 shows BPs for these sequences, computed using two measures, the KU phonotactic probability calculator (Vitevitch and Luce, 2004) and the UCI phonotactic probability calculator (Mayer *et al.*, 2022). These stimuli were also designed to control for neighborhood density (shown in Table 1) with words that have previously been rated as familiar (Nusbaum *et al.*, 1984). Neighborhood density and BP measurements were computed with the KU phonotactic probability calculator and KU neighborhood density calculator (Vitevitch and Luce, 2004). For the UCI phonotactic probability calculator (Mayer *et al.*, 2022), we computed measures with the Carnegie Mellon University Pronouncing Dictionary corpus (Weide, 1998) by using the version of the dictionary which contains words with frequencies of at least one in the CELEX database (Baayen *et al.*, 1995). These methods are the same as those in Steffman and Sundara (2023), and the reader is referred to that paper and the references therein for more details. Table 1 shows the computed BP values, as well as the continuum bias: the BP of one end point of the continuum subtracted from the other. The direction of bias differences across consonant frames were consistent across both measures. For the CV continuum, BP favors /ʌ/. As depicted in Table 1, using the KU calculator, the C₁V₂ portion of the /tʊvip/ ~ /tʌvip/ continuum exhibited an /ʌ/ bias (0.0009), calculated by subtracting the BP for the critical biphone in the /ʊ/ end point from the BP in the /ʌ/ end point. The /sʊvip/ ~ /sʌvip/ continuum has an /ʌ/ bias as well (0.0056). When considering the effect of manipulating BP, the relevant metric is the *difference* in biases for the two initial consonants. This relative bias value (0.0045) indicates that the /ʌ/ bias in the /sʊvip/ ~ /sʌvip/ continuum is greater than the /ʌ/ bias in the /tʊvip/ ~ /tʌvip/. For the VC continuum, a following /v/ favors perception of /ɛ/, whereas a following /b/ creates a higher BP sequence with a preceding /ae/. Based on long-term exposure to English, we thus predict that listeners should provide relatively more /ʌ/ responses with a preceding /s/ in the CV continuum and relatively more /ɛ/ responses with a following /v/ in the VC continuum. Note that the CV continuum has a larger bias difference by both metrics than the VC continuum. This predicts a larger effect of BP for perception of this continuum, although a direct comparison across both continua is difficult because they range between different vowel endpoints and are acoustically different.

The stimuli were recorded by a female speaker of American English in a sound-attenuated booth using a Shure SM81 Condenser Handheld Microphone and Pop Filter (Niles, IL) with a sampling rate of 44.1 kHz (32 bit). Then, the vowel continuum was resynthesized using a Praat script (Winn, 2016), which implemented linear predictive coding (LPC)-based formant resynthesis and varied F₁, F₂, and F₃ in ten equidistant Bark-spaced steps between the two endpoints. BP-manipulating consonants were cross-spliced from a production in which they preceded or followed the vowel that would

Table 1. BP and ND (neighborhood density) are displayed for the two continua. See the text for details.

	BP (KU)	BP (UCI)	ND (KU)
Continuum 1: CV	C_1V_2	CVCVC	CVCVC
/tʌvip/	0.0005	0.0021	0
/tʌvip/	0.0014	0.0065	0
Bias (positive = /ʌ/)	0.0009	0.0044	0
/sʌvip/	0.0003	0.0020	0
/sʌvip/	0.0059	0.0160	0
Bias (positive = /ʌ/)	0.0056	0.0140	0
Bias difference	0.0045	0.0096	Matched
Continuum 2: VC	V_2C_3	CVC	CVC
/maeb/	0.0026	0.0104	29.54
/mɛb/	0.0007	0.0063	17.96
Bias (positive = /ɛ/)	-0.0019	-0.0041	-11.58
/maev/	0.0019	0.0100	30.25
/mɛv/	0.0026	0.0084	17.37
Bias (positive = /ɛ/)	0.007	-0.0016	-12.88
Bias difference	0.0026	0.0025	Matched

predict the opposite of the BP effect, ensuring that any possible traces of vowel information in the consonant would not present a confound (e.g., /t/ from a pre-/ʌ/ context and /s/ from a pre-/ʌ/ context for the CV continuum). We used the same stimulus creation approach as in Steffman and Sundara (2023), which the reader is referred to for more information on the technique. Stimuli are available on the OSF repository.

2.2 Distributional manipulation

The distributional manipulation is visualized in Fig. 1(A). As in Bushong and Jaeger (2019), we varied the pairing of each continuum step with consonant frames. In the flat distributional condition, each continuum step was presented 10 times in each consonant frame (10 steps × 10 reps × 2 frames = 200 trials). The distributions of continuum steps 4–7 were identical in the flat, enhancing, and reversed conditions; the crucial difference was in how often listeners heard the steps closest to the continuum end points. In the enhancing condition [middle section of Fig. 1(A)], listeners never heard the low probability BP sequence (/tʌvip/ or /mɛb/) with the /ʌ/ or /ɛ/ endpoints. Taking /ʌ/ and /ɛ/ as step 1 on each continuum, this resulted in the following number of presentations of the “high BP” frame (/sV/ and /Vv/ in the CV and VC continuum, respectively): steps 1 and 2, 20 presentations; step 3, 15 presentations; step 4–7, 10 presentations; step 8, 5 presentations; steps 9 and 10, 0 presentations for a total of 200 trials). This is shown in terms of proportions in Fig. 1(A) for the enhancing distribution. In the reversed distribution [rightmost section of Fig. 1(A)], the pairings of frames and continuum steps were reversed such that now a frame and step pairing which has lower BP appeared most often with the /ʌ/ or /ɛ/ endpoints. In this condition, the distributional co-occurrence properties in the stimuli crucially conflict with the long-term BP effects in English.

2.3 Participants and procedure

We used a between-subjects design to test the effect of the distributional manipulation. For each of the 2 continua and 3 distributional manipulations, we recruited 36 participants (36 × 6 = 216 participants total). The data were not inspected or analyzed until all participants had completed the experiment. Each participant was a self-reported native American English speaker with normal hearing and vision. Participants were students at a large North American university and received course credit for participation. All provided informed consent. The experiment was implemented in Labvanced (Finger et al., 2017), and participants chose between one of two vowels after hearing a nonword. Only the vowel continuum endpoints were presented as visual choices during a trial, not the whole word. For the CV continuum (/ʌ/ ~ /ɛ/), the choices were orthographic representations OO and UH; for the VC continuum (/ae/ ~ /ɛ/), the choices were A and E. The instructions preceding the experiment gave examples of real words with target vowels to familiarize participants with the labels, and there were four practice trials in which each continuum end point was presented once in each consonant frame.

2.4 Analysis

Data were analyzed using Bayesian mixed-effect logistic regression implemented with *brms* (Bürkner 2017). Because we are not interested in comparing effects across continua, we analyzed each continuum separately. The analysis approach was based on that in Steffman and Sundara (2023), although analysis protocols were not preregistered. Binary responses

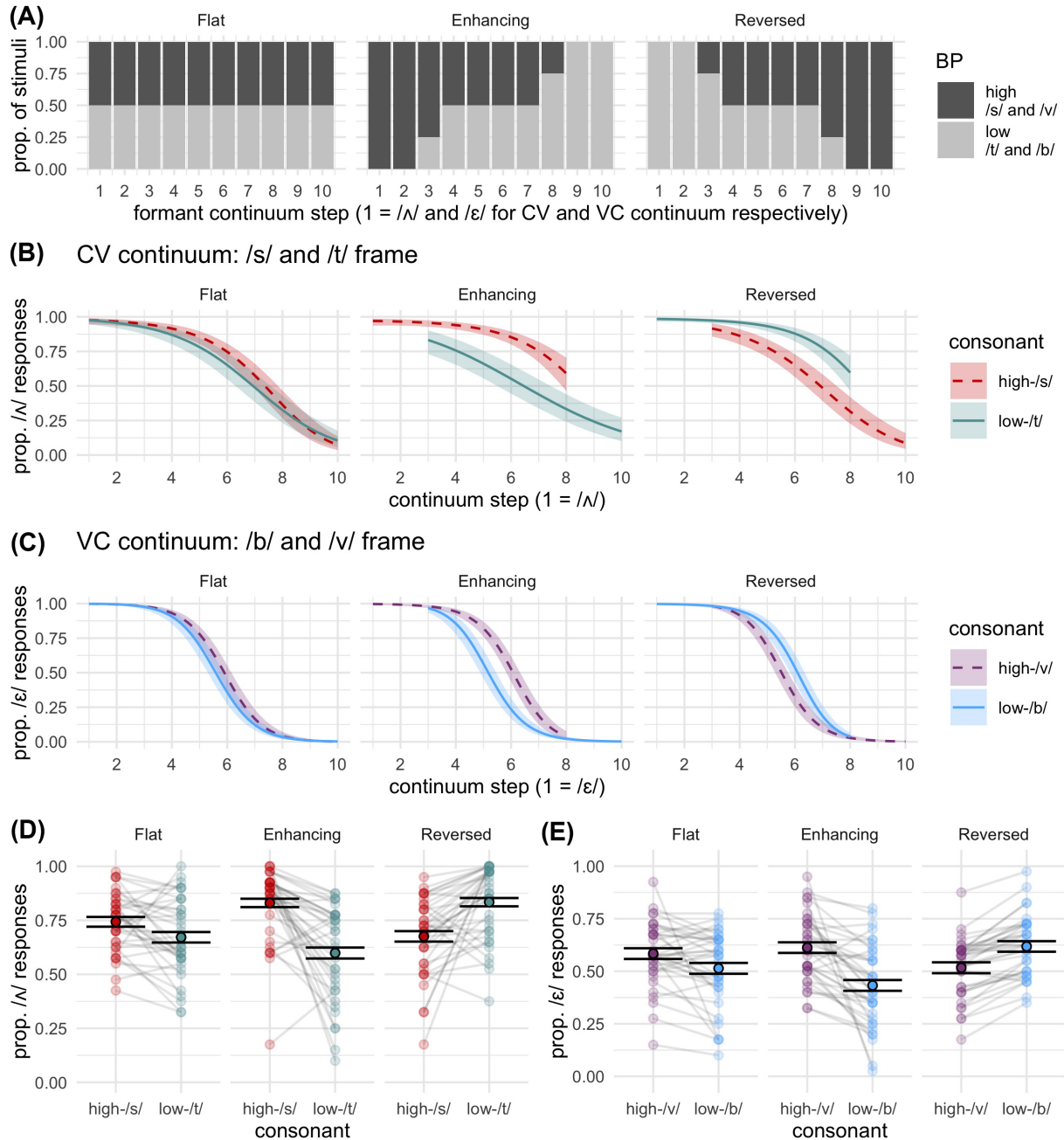


Fig. 1. (A) Visualization of the distributional conditions (see the text). (B) and (C) Model fits for categorization across the continua (x axis) with responses on the y axis (/s/ for the CV continuum and /v/ for the VC continuum); ribbons show 95% CrI from the model. (D) and (E) Responses aggregated across steps 4-7 (which did not vary based on distributional condition), split by consonant frame. Lighter points show individual participants' responses while darker points show group means computed from the raw data, where 95% CIs are shown as error bars.

were predicted as a function of the consonant frame, continuum step, distributional condition, and all interactions. Consonant frame was coded with two levels (/s/ and /v/ mapped to 0.5, and /t/ and /b/ mapped to -0.5 for the CV and VC continuum, respectively). Continuum step was coded as continuous and centered. We included two terms for the continuum: a Gelman-scaled linear term (Gelman, 2008) and a quadratic term (the quadratic term enabling the model to capture a potentially larger effect of the consonant frame in the middle ambiguous region of the continuum when interacted with the frame variable). In coding the distributional manipulation, we used one unit sliding difference coding, which allowed the comparison of (1) the enhancing distribution versus the flat distribution, and (2) the flat distribution versus the reversed distribution. Our crucial interest given this coding scheme is the interaction term between consonant frame

and distributional condition terms. For the comparison between enhanced versus flat distributions, a credible interaction with consonant frame would indicate that the effect of consonant frame differs between these two conditions as predicted if listeners are sensitive to the distributional manipulation. Likewise, for the term comparing flat versus reversed distributions, a credible interaction would indicate that the consonant frame effect differed between these two conditions. These interactions will be our focus in reporting the results. Random effects included by-participant intercepts and slopes for continuum step terms, consonant frame, and their interaction (excluding the distributional manipulation as it was not within participant). Weakly informative priors (student t priors with three degrees of freedom centered on zero: student- $t(3,0,2.5)$) were used for the intercept and fixed effects. The random effect standard deviation (sd) priors were kept at the default half student t priors (with, otherwise, the same parameters as for fixed effect priors). Priors for random effect correlation were the default LKJ(1) priors. In the results, we report the median for an estimate's posterior and 95% credible intervals (CrI). When the CrI excludes the value of zero, this is taken as compelling evidence for an effect. We also report the probability of direction (pd), computed using the *R* package *bayestestR* (Makowski *et al.*, 2019), which gives the percentage of a posterior distribution with a given sign. When 95% CrI exclude zero, $pd > 97.5\%$.

3. Results

Categorization across the continua is presented in Figs. 1(B) and 1(C) (CV continuum and VC continuum, respectively) with vowel responses on the y axis, split by consonant frame (line coloration) and distributional condition (panels). In Figs. 1(D) and 1(E) (CV continuum and VC continuum, respectively) the effects are collapsed across the continuum, plotting just the effect of the consonant frame at steps 4–7 of the continuum, which were presented an equal number of times in each distributional condition. The interactions of interest between the distributional manipulation and consonant frame were credible for the comparison of enhancing versus flat (CV continuum, $\hat{\beta} = -1.11$, 95CrI = $[-1.80, -0.42]$, $pd = 100$; VC continuum, $\hat{\beta} = -0.86$, 95CrI = $[-1.34, -0.39]$, $pd = 100$) and the comparison of flat versus reversed (CV continuum, $\hat{\beta} = -1.86$, 95CrI = $[-2.57, -1.17]$, $pd = 100$; VC continuum, $\hat{\beta} = -1.32$, 95CrI = $[-1.79, -0.85]$, $pd = 100$).

To further examine the interactions, we used *emmeans* (Lenth, 2021) to extract marginal estimates of the consonant frame effects in each distribution condition. For both continua, there was a credible effect of the consonant frame in the flat condition whereby /s/ favored perception of /ʌ/ in the CV continuum ($\hat{\beta} = 0.49$, 95CrI = $[0.02, 0.98]$, $pd = 98$) and /v/ favored perception of /e/ in the VC continuum ($\hat{\beta} = 0.44$, 95CrI = $[0.13, 0.79]$, $pd = 100$), replicating the effects in Steffman and Sundara (2023). The effect showed the same directionality in the enhancing condition, although it was larger in magnitude (CV continuum, $\hat{\beta} = 1.61$, 95CrI = $[1.09, 2.10]$, $pd = 100$; VC continuum, $\hat{\beta} = 1.30$, 95CrI = $[0.97, 1.65]$, $pd = 100$). The consonant frame effect was also credible in the reversed condition, however, the directionality was reversed (CV continuum, $\hat{\beta} = -1.38$, 95CrI = $[-1.88, -0.85]$, $pd = 100$; VC continuum, $\hat{\beta} = -0.87$, 95CrI = $[-1.21, -0.54]$, $pd = 100$). In summary, we found a credible effect of the BP-manipulating frame in each of the distributional conditions, where the effect in the enhancing condition is larger than that in the flat condition, and the reversed condition shows a total reversal of the effect.

4. Discussion

We tested whether listeners update their use of BP information on the basis of laboratory exposure to a skewed distribution of contexts. First, we replicated previously attested long-term BP effects, that is, those based on experience with English. We then showed that with 10–15 min of short-term exposure, BP effects from native language experience can be credibly boosted or completely overridden. Notably, listeners generalized these effects across continuum steps whose distribution was flat in all distributional conditions (steps 4–7). Further, these effects were observed whether the critical context preceded (CV continuum) or followed (VC continuum) the ambiguous vowel. Whether these adaptive effects are temporary or have long-term consequences remains to be determined. Overall, our results show that adults can rapidly update not just categorical restrictions (Cutler *et al.*, 2008) but also gradient phonotactic probabilities based on distributional evidence.

The effects in the enhancing condition are consistent with findings from Bushong and Jaeger (2019), where listeners adapted to the changes in distribution of a following sentence context to modulate phonetic categorization. The enhancement effect is also consistent with perceptual retuning of ambiguous segments as shown in lexically guided perceptual retuning experiments (a different method and experimental design than our own). Retuning occurs whether the supporting lexical context precedes ambiguous segmental information (Jesse and McQueen, 2011; Charoy and Samuel, 2023) or follows it (Charoy and Samuel, 2023; McAuliffe and Babel, 2016). The robust effects in CV and VC continua here, thus, comport with the view that the perceptual system maintains uncertainty about fine-grained aspects of the speech signal prior to commitment to a particular percept (e.g., Bushong, 2020; McMurray *et al.*, 2009; Samuel, 2016).

The complete reversal of long-term, native language segment co-occurrence probability effects demonstrated in this study, however, is surprising as well as novel. In the literature on adaptation, short-term experience has, at best, been shown to neutralize the effects of long-term experience. For example, Idemaru and Holt (2011) find that listener's down-weight F_0 as a cue to stop voicing when its correlation with VOT (the primary cue) is reversed. However, listeners do not reverse the F_0 effect, suggesting a lingering influence of long-term experience with these cues. The present study presents

a clearly different pattern whereby a complete reversal is achieved relatively quickly, showing that the present context effects are more malleable than learning cue correlations that signal categories.

In this study, although we showed enhancement and reversal for the CV and VC continua, the two were not identical in terms of effect size and shape of categorization function. Differences across continua may be due to the different vowel contrasts tested in each, and/or to the different sizes of biases, where a larger bias difference in the CV continuum may have led to greater change in the reweighting of consonant contexts and continuum acoustics (i.e., less-steep categorization functions). What remains an open question is whether qualitatively different patterns of categorization can arise based on the same adaptive mechanism or must be modeled differently.

Xie *et al.* (2023) demonstrate the feasibility of evaluating mechanistic hypotheses against experimental results on lexically guided perceptual retuning. They implement models of adaptive speech perception with three mechanisms: low-level cue normalization, changes in linguistic category representations, or changes in decision-making biases. The authors show, rather strikingly, that each mechanism in the model can effectively account for the empirical data. Thus, even qualitatively different categorization functions may result from the same mechanisms. Such an approach, as applied further to the learning of context effects generally, and more specifically, the effects observed in this study, could be quite useful to identify the mechanism(s) underlying the flexibility observed here.

More broadly, there is a rich literature on the kinds of information listeners can encode and update based on distributional evidence (for a review, see Aslin and Newport, 2014). The argument goes as follows: of the multitude of statistics available in their input, human learners encode only a subset; and distributional learning experiments can help us identify the statistics humans encode. We have shown here that adults update biphone co-occurrence restrictions and, therefore, must encode them. If human listeners are able to update higher-order triphone probabilities in response to distributional evidence favoring it, it would provide evidence that listeners encode such higher-order probabilities (cf. discussion of this point in Norris *et al.*, 2000; Newman, 2000). Further, gradient BP effects of the sort evaluated here vary in the degree to which they are supported by native language experience. It is possible then that because of stronger support from native language experience, some BP effects are more resistant to manipulation from short-term exposure in the laboratory. If this is correct, then effects of manipulating the distributional evidence to reverse existing BP effects can provide a new window into the strength of listeners' representations. The limits on learning from distributional input are also informative to understand how humans learn about the sound system of their native language (e.g., Nevins, 2010). When humans underlearn or overlearn from distributional data, we find evidence for biases (e.g., Moreton and Pater, 2012). Changes in phonetic categorization in response to distributional differences in the input, thus, provide another implicit learning paradigm for future investigations of learning biases.

In conclusion, our findings demonstrate that adults' sensitivity to segment probability restrictions that underlie native speakers' knowledge of phonotactics are malleable. This malleability provides another window to evaluate models of spoken word recognition that embody different input representations. If these learned effects can be generalized to other tasks and/or shown to be long-term, they could be used to develop approaches to improve second language learning: tracking and updating the probabilistic co-occurrence of phones may help with adapting to a novel talker, dialect, or language. Future work will, therefore, benefit from extending this paradigm to cross-talker and cross-language contexts.

Author Declarations

Conflict of Interest

The authors do not have any conflicts of interest to disclose.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. All data, models, model summaries, and code for analysis can be found at <https://osf.io/9uwsk/>.

References

¹See <https://osf.io/9uwsk/>.

- Aslin, R. N., and Newport, E. L. (2014). "Distributional language learning: Mechanisms and models of category formation," *Lang. Learn.* 64(s2), 86–105.
- Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). *CELEX2 LDC96L14. Web Download.* (Linguistic Data Consortium, Philadelphia).
- Borsky, S., Tuller, B., and Shapiro, L. P. (1998). "How to milk a coat: The effects of semantic and acoustic information on phoneme categorization," *J. Acoust. Soc. Am.* 103(5), 2670–2676.
- Bürkner, P. C. (2017). "brms: An R package for Bayesian multilevel models using Stan," *J. Stat. Soft.* 80(1), 1–28.
- Bushong, W. (2020). *Maintenance of Subcategorical Information in Spoken Word Recognition* (University of Rochester, Rochester, NY).
- Bushong, W., and Jaeger, T. F. (2019). "Dynamic re-weighting of acoustic and contextual cues in spoken word recognition," *J. Acoust. Soc. Am.* 146(2), EL135–EL140.
- Charoy, J., and Samuel, A. G. (2023). "Bad maps may not always get you lost: Lexically driven perceptual recalibration for substituted phonemes," *Atten. Percept. Psychophys.* 85, 2437–2422.

- Clayards, M., Tanenhaus, M. K., Aslin, R. N., and Jacobs, R. A. (2008). "Perception of speech reflects optimal use of probabilistic speech cues," *Cognition* 108(3), 804–809.
- Connine, C. M. (1987). "Constraints on interactive processes in auditory word recognition: The role of sentence context," *J. Mem. Lang.* 26(5), 527–538.
- Connine, C. M., Blasko, D. G., and Hall, M. (1991). "Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints," *J. Mem. Lang.* 30(2), 234–250.
- Cummings, S. N., and Theodore, R. M. (2022). "Perceptual learning of multiple talkers: Determinants, characteristics, and limitations," *Atten. Percept. Psychophys.* 84(7), 2335–2359.
- Cutler, A., McQueen, J. M., Butterfield, S., and Norris, D. (2008). "Prelexically-driven perceptual retuning of phoneme boundaries," in *Proceedings of INTERSPEECH 2008* (ICSA, Brisbane, Australia), pp. 2056.
- Finger, H., Goeke, C., Diekamp, D., Standvoß, K., and König, P. (2017). "LabVanced: A unified JavaScript framework for online studies," in *International Conference on Computational Social Science*, Cologne, Germany.
- Ganong, W. F. (1980). "Phonetic categorization in auditory word perception," *J. Exp. Psychol.: Human Percept. Perform.* 6(1), 110–125.
- Gelman, A. (2008). "Scaling regression inputs by dividing by two standard deviations," *Stat. Med.* 27(15), 2865–2873.
- Goldrick, M., and Larson, M. (2008). "Phonotactic probability influences speech production," *Cognition* 107(3), 1155–1164.
- Idemaru, K., and Holt, L. L. (2011). "Word recognition reflects dimension-based statistical learning," *J. Exp. Psychol.: Human Percept. Perform.* 37(6), 1939–1956.
- Jesse, A., and McQueen, J. M. (2011). "Positional effects in the lexical retuning of speech perception," *Psychon. Bull. Rev.* 18, 943–950.
- Lenth, R. (2021). "emmeans: Estimated marginal means, aka least-squares means, R package (version 1.7.1-1)," available at <https://CRAN.R-project.org/package=emmeans> (Last viewed December 1, 2022).
- Liu, L., and Jaeger, T. F. (2018). "Inferring causes during speech perception," *Cognition* 174, 55–70.
- Luce, P. A., and Pisoni, D. B. (1998). "Recognizing spoken words: The neighborhood activation model," *Ear Hear.* 19(1), 1–36.
- Makowski, D., Ben-Shachar, M. S., and Lüdtke, D. (2019). "bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework," *J. Open Source Softw.* 4(40), 1541.
- Mayer, C., Kondur, A., and Sundara, M. (2022). "UCI Phonotactic Calculator (version 0.1.0) [computer software]," available at <https://doi.org/10.5281/zenodo.7443706> (Last viewed March 1, 2023).
- McAuliffe, M., and Babel, M. (2016). "Stimulus-directed attention attenuates lexically-guided perceptual learning," *J. Acoust. Soc. Am.* 140(3), 1727–1738.
- McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (2009). "Within-category VOT affects recovery from 'lexical' garden-paths: Evidence against phoneme-level inhibition," *J. Mem. Lang.* 60(1), 65–91.
- Moreton, E., and Pater, J. (2012). "Structure and substance in artificial-phonology learning, Part I: Structure," *Lang. Linguist. Compass* 6(11), 686–701.
- Nevins, A. (2010). "Two case studies in phonological universals: A view from artificial grammars," *Biolinguistics* 4(2-3), 218–233.
- Newman, R. S. (2000). "Not all neighborhood effects are created equal," *Behav. Brain Sci.* 23(3), 343–343.
- Norris, D., McQueen, J. M., and Cutler, A. (2000). "Merging information in speech recognition: Feedback is never necessary," *Behav. Brain Sci.* 23(3), 299–325.
- Norris, D., McQueen, J. M., and Cutler, A. (2003). "Perceptual learning in speech," *Cognit. Psychol.* 47, 204–238.
- Nusbaum, H. C., Pisoni, D. B., and Davis, C. K. (1984). "Sizing up the Hoosier mental lexicon," *Research on Spoken Language Processing Report* (Bloomington, IN), Vol. 10, Iss. 3, pp. 357–376.
- Pitt, M. A., and McQueen, J. M. (1998). "Is compensation for coarticulation mediated by the lexicon?," *J. Mem. Lang.* 39(3), 347–370.
- Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996). "Word segmentation: The role of distributional cues," *J. Mem. Lang.* 35(4), 606–621.
- Samuel, A. G. (2016). "Lexical representations are malleable for about one second: Evidence for the non-automaticity of perceptual recalibration," *Cognit. Psychol.* 88, 88–114.
- Samuel, A. G., and Kraljic, T. (2009). "Perceptual learning for speech," *Atten. Percept. Psychophys.* 71(6), 1207–1218.
- Schertz, J., and Clare, E. J. (2020). "Phonetic cue weighting in perception and production," *Wiley Interdiscip. Rev.: Cognit. Sci.* 11(2), e1521.
- Steffman, J., and Sundara, M. (2023). "Disentangling the role of biphone probability from neighborhood density in the perception of non-words," *Lang. Speech* (published online).
- Taylor, C. F., and Houghton, G. (2005). "Learning artificial phonotactic constraints: Time course, durability, and relationship to natural constraints," *J. Exp. Psychol.: Learn., Mem., Cognit.* 31, 1398–1416.
- Vitevitch, M. S., and Luce, P. A. (2004). "A web-based interface to calculate phonotactic probability for words and nonwords in English," *Behav. Res. Methods, Instrum., Comput.* 36(3), 481–487.
- Warker, J. A. (2013). "Investigating the retention and time course of phonotactic constraint learning from production experience," *J. Exp. Psychol.: Learn., Mem., Cognit.* 39, 96–109.
- Warker, J. A., and Dell, G. S. (2006). "Speech errors reflect newly learned phonotactic constraints," *J. Exp. Psychol.: Learn., Mem., Cognit.* 32(2), 387–398.
- Warker, J. A., and Dell, G. S. (2015). "New phonotactic constraints learned implicitly by producing syllable strings generalize to the production of new syllables," *J. Exp. Psychol.: Learn., Mem., Cognit.* 41(6), 1902–1910.
- Weide, R. L. (1998). The Carnegie Mellon pronouncing dictionary, release 0.6, available at www.cs.cmu.edu (Last viewed December 1, 2022).
- Winn, M. (2016). "Praat script: Make formant continuum [computer software]," available at <http://www.mattwinn.com/praat.html> (Last viewed January 15, 2018).
- Xie, X., Jaeger, T. F., and Kurumada, C. (2023). "What we do (not) know about the mechanisms underlying adaptive speech perception: A computational framework and review," *Cortex* 166, 377–424.