A second chance for a first impression:

Sensitivity to cumulative input statistics for lexically guided perceptual learning

Christina Y. Tzeng[a], Lynne C. Nygaard[a], and Rachel M. Theodore[b,c]

[a]Department of Psychology

Emory University

36 Eagle Row

Atlanta, GA 30322

[b]Department of Speech, Language, and Hearing Sciences

University of Connecticut

2 Alethia Drive, Unit 1085

Storrs, CT 06269-1085

[c]Connecticut Institute for the Brain and Cognitive Sciences

University of Connecticut

337 Mansfield Road, Unit 1872

Storrs, CT 06269-1872

Author to whom correspondence should be addressed:

Rachel M. Theodore

rachel.theodore@uconn.edu

Abstract

Listeners use lexical knowledge to modify the mapping from acoustics to speech sounds, but the timecourse of experience that informs lexically guided perceptual learning is unknown. Some data suggest that learning is contingent on initial exposure to atypical productions, while other data suggest that learning reflects only the most recent exposure. Here we seek to reconcile these findings by assessing the type and timecourse of exposure that promote robust lexcially guided perceptual learning. In three experiments, listeners (n = 560) heard 20 critical productions interspersed among 200 trials during an exposure phase and then categorized items from an *ashi–asi* continuum in a test phase. In experiment 1, critical productions consisted of ambiguous fricatives embedded in either /s/- or /ʃ/-biasing contexts. Learning was observed; the /s/-bias group showed more *asi* responses compared to the /ʃ/-bias group. In experiment 2, listeners heard ambiguous *and* clear productions in a consistent context. Order and lexical bias were manipulated between-subjects, and perceptual learning occurred regardless of the order in which the clear and ambiguous productions were heard. In experiment 3, listeners heard ambiguous fricatives in both /s/- and /ʃ/-biasing contexts. Order differed between two exposure groups, and no difference between groups was observed at test. Moreover, the results showed a monotonic decrease in learning across experiments, in line with decreasing exposure to stable lexically biasing contexts, and were replicated across novel stimulus sets. In contrast to previous findings showing that either initial or most recent experience are critical for lexically guided perceptual learning, the current results suggest that perceptual learning reflects cumulative experience with a talker's input over time.

Introduction

Listeners achieve constancy in speech perception despite variation in talkers' productions (e.g., Allen et al., 2003; Hillenbrand et al., 1995; Newman et al., 2001; Theodore et al., 2009). One mechanism that underlies this ability is *perceptual learning*, in which listeners restructure phonetic categories to accommodate systematic variation in speech input (e.g., Norris et al., 2003). Considerable evidence suggests that listeners accumulate distributional information about talker-specific acoustic-phonetic characteristics and use this information to dynamically adjust mappings to linguistic representations (e.g., Norris et al., 2003; Nygaard & Pisoni, 1998; Samuel & Kraljic, 2009; Theodore et al., 2015; Theodore & Miller, 2010). These findings support distributional tracking accounts of perceptual learning, which posit that listeners use statistical contingencies in the speech signal to accommodate variation (e.g., Kleinschmidt & Jaeger, 2015; Maye et al., 2008; McMurray et al., 2009).

Lexically guided perceptual learning offers a means to assess how listeners maintain tension between flexibility and stability in speech perception (Norris et al., 2003). During an exposure phase, listeners hear an ambiguous sound (e.g., a fricative with spectral energy ambiguous between /s/ and /ʃ/) embedded in a disambiguating lexical context that differs between listener groups. For example, the ambiguity replaces /s/ for some listeners (e.g., *compensate*) and /ʃ/ for other listeners (e.g., *publisher*). Following exposure, listeners categorize members along a speech sound continuum (e.g., *ashi – asi*). Given exposure to an ambiguous sound in disambiguating lexical contexts, listeners subsequently modify the perceptual boundary along a speech sound continuum in line with biasing lexical context (e.g., listeners biased to interpret the ambiguity as /s/ show more /s/ responses at test than listeners biased to interpret the ambiguity as /ʃ/). Listeners use lexical knowledge to accommodate ambiguities for a host of

acoustic-phonetic properties including those that cue fricative place of articulation (Kraljic et al., 2008; Norris et al., 2003), vowel identity (Maye et al., 2008), voicing (Kraljic & Samuel, 2006), and stop consonant place of articulation (Maye et al., 2008).

What remains unclear for theories of perceptual learning is the timecourse of experience that informs lexically guided perceptual learning. The Bayesian belief-updating model of speech adaptation (Kleinschmidt & Jaeger, 2015) predicts that learning reflects a context-dependent (e.g., talker-specific) cumulative integration of listeners' experience with speech input. Initial input from a novel talker is processed based on prior knowledge (e.g., knowledge of language-specific cue distributions). Learning occurs if the talker's input deviates from these expectations, reflecting an integration of prior knowledge and the observed new evidence. Iterative updating is predicted to occur until a new context in encountered (e.g., a change in talker), at which point priors are reset to initial expectations.

Though numerous investigations suggest that listeners use cumulative (i.e., global) experience with input statistics for adaptation in speech perception (e.g., Idemaru & Holt, 2011; Kraljic et al., 2008; Kraljic & Samuel, 2005; Theodore & Monto, 2019) and auditory perception more generally (e.g., Baese-Berk et al., 2014; McAuley & Miller, 2007), the timecourse of experience that contributes to perceptual adaptation remains unknown (Theodore & Monto, 2019; Xie et al., 2018). Indeed, findings from lexically guided perceptual learning remain equivocal on this point. Kraljic and colleagues (2008) found that perceptual recalibration for a talker's ambiguous productions only occurred if listeners had no prior experience with that talker producing clear productions. This "first impressions" effect suggests that listeners are sensitive to global experience to the degree that initial exposure affects (or blocks) learning from later exposure, but also suggests that adaptation does not simply reflect aggregated experience. In

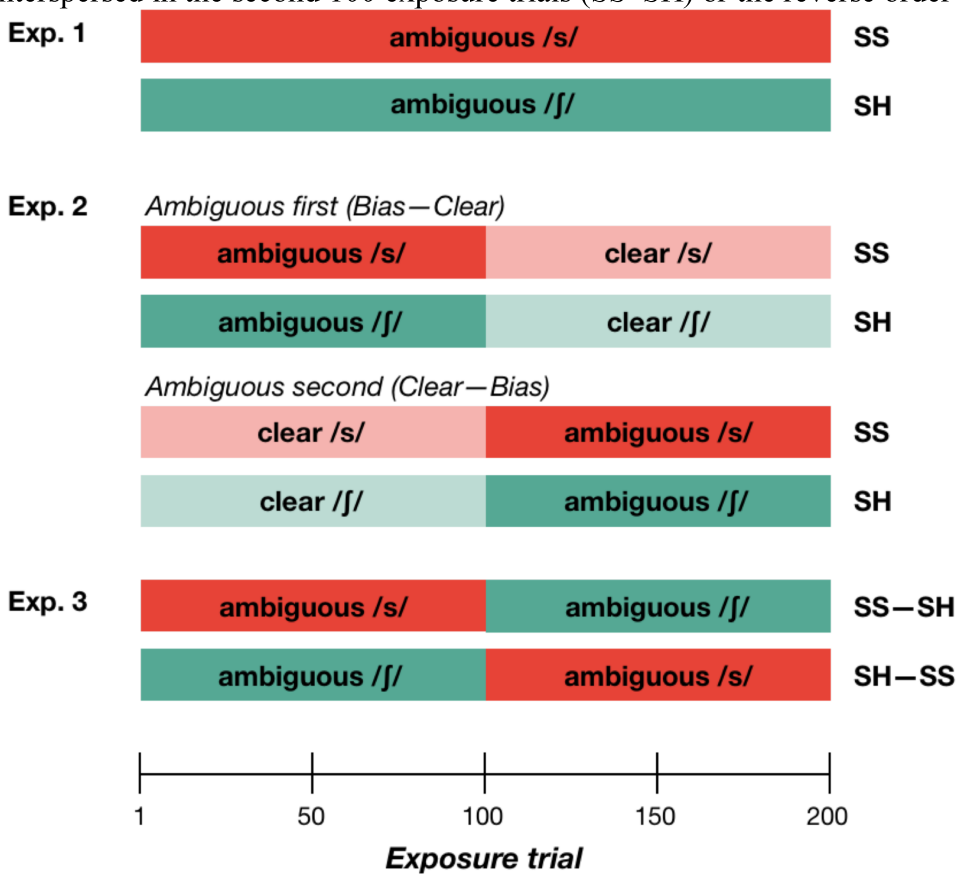
contrast, Saltzman and Myers (2018) suggested that perceptual learning reflects sensitivity to recent (i.e., local) input statistics. Listeners were biased to perceive an ambiguous fricative as both /s/ and /ʃ/ in separate exposure-test blocks and block order was manipulated. A learning effect of similar magnitude was observed in each block, suggesting that perceptual recalibration reflects sensitivity to the most recent statistical cues in the input.[1] Disparate results regarding listeners' reliance on local versus global input statistics preclude drawing definitive conclusions about the learning mechanism.

Investigations to date also do not afford a specific test of a cumulative tracking tenet of the belief-updating model of speech adaptation (Kleinschmidt & Jaeger, 2015); namely, that the magnitude of learning should reflect the *consistency* of a talker's input. In Kraljic et al. (2008), the magnitude of learning resulting from exposure to 10 ambiguous and 10 clear productions of the a given biasing context was not directly compared to learning that occurs from exposure to 20 ambiguous productions in the same biasing context. In Saltzman and Myers (2018), listeners were given exposure to ambiguous productions in two different biasing contexts, and learning was assessed after each biasing context. As such, although biasing context was inconsistent, the learning assay itself may have triggered a return to prior knowledge (Kleinschmidt & Jaeger, 2015).

Here we test predictions of the local and global statistics hypotheses – and the extent to which consistency in exposure promotes learning – by manipulating the type and timing of critical productions while holding exposure "dose" constant (Figure 1).

[1] A retraction note (Saltzman & Myers, 2020) for this study was issued after the initial submission of the current manuscript. Because the results presented in Saltzman and Myers (2018) contributed to the scientific premise of the current work, we present them here so that the introduction is a veridical representation of our understanding of the scientific record as this study was developed.

Figure 1. Distribution of critical productions for each bias group (labeled in bold, at right) during the exposure phase for each experiment. In experiment 1, the 20 critical productions consisted of ambiguous fricatives consistently presented in either an /s/- or /ʃ/-biasing context (labeled as SS and SH, respectively); in both cases, the 20 critical productions appeared randomly throughout the 200 exposure trials. In experiment 2, the 20 critical productions consisted of 10 ambiguous productions (dark) and 10 clear productions (light) of the same category. Order in which the ambiguous and clear productions were encountered was manipulated between two order groups such that listeners heard 10 ambiguous productions randomly interspersed in the first 100 exposure trials followed by 10 clear productions randomly interspersed in the second 100 exposure trials (Bias–Clear) or the reverse order (Clear–Bias). In experiment 3, critical productions consisted of 10 ambiguous fricatives presented in an /s/-biasing context *and* 10 ambiguous productions presented in an /ʃ/-biasing context. Order of the biasing contexts was manipulated such that listeners heard 10 ambiguous /s/ productions randomly interspersed in the first 100 exposure trials followed by 10 ambiguous /ʃ/ productions randomly interspersed in the second 100 exposure trials (SS–SH) or the reverse order (SH–SS).



The standard dose in the lexically guided perceptual learning paradigm is 20 critical productions that are randomly distributed across 200 exposure trials. Experiment 1 is a replication of the standard paradigm; critical productions were uniformly ambiguous and presented in a consistent biasing context. Following Kraljic et al. (2008), critical productions in Experiment 2 consisted of

10 ambiguous and 10 clear productions in a consistent context, and we manipulated the order in which the ambiguous and clear productions were encountered. In Experiment 3, critical productions were uniformly ambiguous, but lexical context was inconsistent, as in Saltzman and Myers (2018). Listeners heard 10 productions in each of the two biasing contexts, and we manipulated the order in which each context was encountered. Finally, within each experiment, we conducted parallel examinations for two stimulus sets to assess replicability and generalizability of the results. That is, each experiment was conducted twice (e.g., 1A and 1B for experiment 1), one for stimuli produced by a female talker (i.e., 1A, 2A, 3A) and one for stimuli produced by a male talker (i.e., 1B, 2B, 3B).

If local input statistics are the putative determinant of perceptual learning, then learning will be observed in experiments 1 and 3, and for the "Ambiguous second" conditions in experiment 2. If perceptual learning is contingent on initial exposure to *ambiguous* productions, then learning will be observed in experiments 1 and 3, and for the "Ambiguous first" conditions in experiment 2. In contrast, the global statistics hypothesis predicts that (1) learning will be observed in experiments 1 and 2 but not experiment 3, (2) learning in experiment 2 will not depend on the order in which clear and ambiguous productions are encountered, and (3) the magnitude of learning will decrease across experiments in line within diminishing consistency between ambiguous input and lexical context. We present each experiment in turn, and then present analyses that compare performance across experiments.

Experiment 1

Methods

Participants. All participants reported in this manuscript were recruited from the Prolific participant pool (https://www.prolific.co). Participants were monolingual, native speakers of

American English between 18 and 35 years of age currently residing in the United States with no history of language-related disorders per self-report. Each participant only participated once across the experiments reported here. All passed the headphone screen of Woods et al. (2017) at the time of testing, achieved ≥ 70% lexical decision accuracy for all four item types presented during exposure, and showed a logistic response function at test.[2] Experiments 1A and 1B each included 70 participants; within each experiment, 35 participants were randomly assigned to the SS exposure group and 35 participants were randomly assigned to the SH exposure group. Demographic information for the participants in each experiment is shown in Table 1; all were paid $3.33 for their participation.

Table 1. Demographic characteristics of participants in each experiment.

		Gender		Age	
Experiment	*n*	Women	Men	Range	Mean (SD)
1A	70	37	33	18 – 35	27.1 (5.0)
1B	70	38	32	18 – 35	27.0 (5.0)
2A	140	65	75	18 – 35	26.8 (4.9)
2B	140	71	69	18 – 35	26.2 (4.9)
3A	70	41	29	18 – 35	26.1 (5.1)
3B	70	29	41	18 – 35	26.5 (4.8)

Stimuli. Two native speakers of American English (one female, one male) recorded the stimuli from Kraljic and Samuel (2005) for the lexical decision (exposure) task and the phonetic categorization (test) task. Stimuli for the lexical decision task consisted of 20 critical /s/ words, 20 critical /ʃ/ words, 60 filler words, and 100 filler nonwords. The 40 critical words ranged in length from two to four syllables, with the critical /s/ or /ʃ/ sound occurring relatively late in the

[2] In addition to the 560 participants reported here, an additional 32 participants were tested but excluded from the study because they showed lexical decision accuracy < 70% for at least one of the item types presented during the exposure phase (n = 24) or did not show a logistic response function at test (n = 8).

word. Half of the critical words contained a single instance of /s/ and no occurrences of /ʃ/, and the other half contained a single /ʃ/ and no /s/. Both sets of critical words were matched in mean syllable length and word frequency. The 60 filler words had no instance of /s/ or /ʃ/ and were matched to the critical words in stress pattern, number of syllables, and word frequency. Filler nonwords contained no /s/ or /ʃ/ phonemes (see Kraljic & Samuel, 2005 for details).

Both talkers produced a second version of each of the 40 critical words, replacing the critical phoneme with its counterpart phoneme (e.g., *compensate* and *compenshate*). We created an ambiguous s-ʃ mixture for each critical word in Praat (Boersma & Weenink, 2018). The /s/ and /ʃ/ phonemes in each critical word pair were mixed together with seven equidistant weightings from 80% /s/ - 20% /ʃ/ to 20% /s/ - 80% /ʃ/ (i.e., 80-20, 70-30, 60-40, 50-50, 40-60, 30-70, and 20-80). Each mixture was inserted into the /s/ word frame and saved as an independent file. Two native speakers of American English listened to each of the seven mixtures and independently judged which was most ambiguous for each item. If the two listeners disagreed by more than one step, then the midpoint was selected as most ambiguous. If the two listeners disagreed by a single step, then a new mixture was created that was intermediate between the two steps. The specific mixtures for each exposure stimulus are listed in the OSF repository for this manuscript as identified in the Open Practices Statement.

Stimuli for the phonetic categorization task consisted of nine items on a continuum that ranged from /ɑʃi/ to /ɑsi/, recorded by the same two talkers who recorded the lexical decision stimuli. Items on the /ɑʃi/–/ɑsi/ continuum ranged from 100% /ɑʃi/ - 0% /ɑsi/ to 0% /ɑʃi/ - 100% /ɑsi/. The procedure for creating the seven intermediate items on the continuum was identical to that for creating the ambiguous critical words in the lexical decision task such that the fricatives in each of the continuum endpoints were mixed together with the same weightings (i.e., 80-20,

70-30, 60-40, 50-50, 40-60, 30-70, and 20-80) and then reinserted into the /ɑsi/ frame to create seven equidistant mixtures.

Procedure. Stimuli from the female talker (f1) were used in 1A and stimuli from the male talker (m2) were used in 1B. All experiments presented in this manuscript were web-based studies hosted on the Gorilla platform (Anwyl-Irvine et al., 2019). After providing informed consent, participants completed a headphone screen, exposure phase, and test phase. The headphone screen followed the protocol of Woods and colleagues, which is designed to ensure compliance with headphone use for web-based experiments (Woods et al., 2017). During the exposure phase, the 200 items appropriate for each exposure condition were presented in randomized order. For listeners in the SS groups, stimuli consisted of 20 tokens with ambiguous fricatives embedded in /s/-biasing contexts, 20 tokens with clear /ʃ/ productions, 60 filler words, and 100 nonwords. For listeners in the SH groups, stimuli consisted of 20 tokens with clear /s/ productions, 20 tokens with ambiguous fricatives embedded in /ʃ/-biasing contexts, 60 filler words, and 100 nonwords. On each exposure trial, participants indicated whether the item was a word or not by pressing one of two keys on the keyboard.

During the test phase, the nine test stimuli were presented in eight cycles, each consisting of a random ordering of the nine continuum steps, for a total of 72 test trials. On each trial, participants identified each item as either *asi* or *ashi* by pressing one of two keys on the keyboard. For both the training and test phases, trials were separated by 1000 ms, timed from the participant's response. The entire procedure lasted approximately 20 minutes.

Statistical analysis. Trial-level data and an analysis script for all experiments reported here can be retrieved at https://osf.io/wa7m3/. Trial-level responses (0 = *ashi*, 1 = *asi*) were submitted to a generalized linear mixed effects model (GLMM) with the binomial response

family as implemented in lme4 (Bates et al., 2015); the Satterthwaite approximation of degrees of freedom was used to evaluate statistical significance using lmerTest (Kuznetsova et al., 2017). The 95% confidence interval for model coefficients was calculated using the summ() function of the jtools package in R (Long, 2020). The model included continuum step, bias, and their interaction as fixed effects. Continuum step was entered into the model as a scaled/centered continuous variable; bias was sum-coded (SH = -0.5, SS = 0.5). The random effects structure consisted of random intercepts by subject and random slopes for continuum step by subject, which reflects the maximal random effects structure for the experimental design.

Results

Experiment 1A. Performance during the exposure phase was near ceiling for all experiments and is presented in Table 2. Figure 2A displays mean proportion *asi* responses at test. Visual inspection suggests a robust learning effect, reflecting more *asi* responses in the SS bias group compared to the SH bias group. Model results are shown in Table 3. As expected, there was a main effect of continuum step ($p < 0.001$), indicating that *asi* responses increased with percent /s/ energy in the continuum. There was also a main effect of bias ($p < 0.001$), with more *asi* responses in the SS compared to the SH exposure group. The interaction between continuum step and bias was not reliable ($p = 0.410$). The main effect of bias was confirmed using a likelihood ratio test that compared the omnibus model to a simpler model in which bias was removed as a fixed effect; there was a significant improvement to goodness of fit when bias was included in the model ($\chi^2(2) = 34.435$, $p < 0.001$).

Table 2. Mean lexical decision accuracy in each experiment for the four item types presented during exposure. Means reflect grand means calculated over by-subject means; values in parentheses indicate standard deviation.

			Item type			
Experiment	Bias	Order	/s/	/ʃ/	Filler word	Nonword
1A	SS	n/a	98 (4)	99 (3)	94 (6)	96 (4)
	SH	n/a	99 (4)	99 (2)	95 (3)	95 (5)
1B	SS	n/a	93 (7)	98 (4)	95 (4)	94 (6)
	SH	n/a	99 (3)	98 (4)	96 (3)	95 (3)
2A	SS	Bias – Clear	96 (5)	99 (2)	96 (4)	96 (5)
	SH	Bias – Clear	98 (4)	100 (2)	94 (6)	95 (6)
	SS	Clear – Bias	99 (2)	100 (1)	96 (3)	96 (3)
	SH	Clear – Bias	99 (2)	99 (3)	95 (3)	95 (7)
2B	SS	Bias – Clear	94 (8)	99 (2)	96 (3)	94 (5)
	SH	Bias – Clear	98 (4)	99 (3)	94 (5)	95 (5)
	SS	Clear – Bias	97 (4)	98 (3)	96 (4)	93 (5)
	SH	Clear – Bias	97 (4)	99 (4)	95 (5)	94 (6)
3A	SH – SS	SH – SS	95 (7)	98 (6)	95 (5)	93 (7)
	SS – SH	SS – SH	97 (5)	97 (5)	96 (3)	95 (6)
3B	SH – SS	SH – SS	94 (6)	97 (5)	94 (5)	92 (6)
	SS – SH	SS – SH	94 (8)	98 (2)	96 (4)	94 (5)

Figure 2. Mean proportion *asi* responses as a function of continuum step for each bias condition in experiment 1A (panel A) and experiment 1B (panel B). Continuum step is presented in terms of percent /s/ energy in each step of the test continuum. Means reflect grand means calculated over by-subject means; error bars indicate standard error.

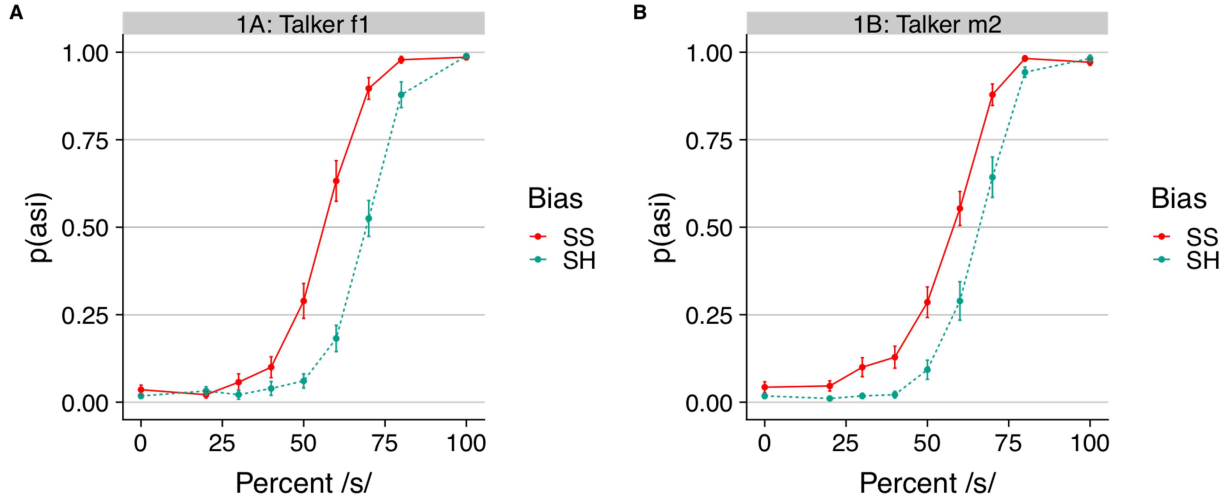


Table 3. Results of the generalized linear mixed effects model for each experiment. The models for experiments 1A, 1B, 3A, and 3B each contained 5,040 observations total across 70 participants. The models for experiments 2A and 2B each contained 10,080 observations total across 140 participants.

Experiment	Fixed effect	$\hat{\beta}$	*SE*	95% CI		*z*	*p*
				Lower	*Upper*		
1A	(Intercept)	-2.110	0.228	-2.557	-1.663	-9.252	< 0.001
	Step	5.185	0.317	4.564	5.806	16.367	< 0.001
	Bias	2.025	0.437	1.169	2.881	4.636	< 0.001
	Step * Bias	0.464	0.563	-0.640	1.568	0.824	0.410
1B	(Intercept)	-1.907	0.197	-2.293	-1.522	-9.689	< 0.001
	Step	4.798	0.292	4.225	5.370	16.426	< 0.001
	Bias	2.143	0.380	1.397	2.888	5.632	< 0.001
	Step * Bias	-1.735	0.532	-2.778	-0.691	-3.258	0.001
2A	(Intercept)	-2.651	0.170	-2.985	-2.318	-15.592	< 0.001
	Step	5.776	0.266	5.255	6.297	21.741	< 0.001
	Bias	0.606	0.315	-0.011	1.223	1.924	0.054
	Order	-0.315	0.314	-0.930	0.301	-1.001	0.317
	Step * Bias	1.601	0.469	0.683	2.520	3.418	0.001
	Step * Order	-0.357	0.465	-1.268	0.555	-0.767	0.443
	Bias * Order	-0.104	0.628	-1.334	1.126	-0.166	0.868
	Step * Bias * Order	-1.053	0.930	-2.875	0.769	-1.133	0.257
2B	(Intercept)	-1.973	0.147	-2.261	-1.685	-13.440	< 0.001
	Step	5.162	0.223	4.726	5.599	23.167	< 0.001
	Bias	0.786	0.279	0.240	1.333	2.820	0.005
	Order	0.047	0.278	-0.498	0.592	0.169	0.866
	Step * Bias	-0.352	0.394	-1.124	0.420	-0.894	0.371
	Step * Order	0.000	0.393	-0.771	0.771	0.000	1.000
	Bias * Order	0.783	0.555	-0.304	1.871	1.412	0.158
	Step * Bias * Order	-0.319	0.785	-1.857	1.219	-0.407	0.684
3A	(Intercept)	-2.851	0.305	-3.449	-2.254	-9.353	< 0.001
	Step	5.912	0.421	5.088	6.737	14.057	< 0.001
	Bias	0.310	0.575	-0.817	1.437	0.539	0.590
	Step * Bias	0.030	0.761	-1.461	1.521	0.040	0.968
3B	(Intercept)	-1.381	0.158	-1.691	-1.072	-8.740	< 0.001
	Step	4.344	0.250	3.853	4.834	17.362	< 0.001
	Bias	0.400	0.306	-0.200	0.999	1.307	0.191
	Step * Bias	-0.379	0.462	-1.286	0.527	-0.820	0.412

Experiment 1B. Figure 2B shows performance at test. The model revealed a main effect of continuum step ($p < 0.001$), a main effect of bias ($p < 0.001$), and an interaction between continuum step and bias ($p = 0.001$), indicating that the learning effect (i.e., more *asi* responses in the SS compared to the SH condition) differed across continuum steps. As for 1A, the effect of bias was confirmed using a likelihood ratio test showing a significant improvement to goodness of fit when bias was included in the model ($\chi^2(2) = 29.780$, $p < 0.001$).

Experiment 2

Experiment 1 confirms that perceptual learning in the standard lexically guided perceptual learning paradigm was elicited for both stimulus sets in our web-based paradigm. Experiment 2 consisted of two replications of Kraljic et al., 2008, one for each of the two stimulus sets used in experiment 1. Listeners heard 10 ambiguous *and* 10 clear fricatives for the 20 critical items during the exposure block. The order in which listeners encountered ambiguous and clear productions for the same sound was manipulated between listener groups. The "first impressions" account (Kraljic et al., 2008) predicts that learning will only occur for listeners who hear the ambiguous productions first, and makes no specific predictions regarding the magnitude of learning in experiment 2 compared to experiment 1. The global statistics hypothesis predicts that learning (as tested here, in a single session that follows all exposure) will not depend on the order in which clear and ambiguous productions are encountered and that the magnitude of learning will be smaller in experiment 2 compared to experiment 1.

Methods

Participants. Experiments 2A and 2B each tested 140 participants; within each experiment, 35 participants were randomly assigned to one of the four between-subjects cells formed by crossing bias (SS vs. SH) and order (Bias–Clear vs. Clear–Bias), as illustrated in

Figure 1. All participants met the inclusion criteria described for experiment 1 and were paid $3.33 for their participation.

Stimuli. Experiment 2 used the same stimuli as described for experiment 1.

Procedure. Stimuli from talker f1 were used in 2A and stimuli from talker m2 were used in 2B. As described for experiment 1, the study consisted of a headphone screen, an exposure phase, and a test phase that was completed online using the web-based Gorilla platform. The procedure was a direct replication of that outlined for the "Audio-only" conditions of Kraljic et al. (2008). All listeners completed one lexical decision exposure block consisting of 200 trials. The 200 exposure items described for experiment 1 were randomly assigned to either the first or second half of the exposure block so that the first 100 trials and the second 100 trials each contained 10 critical /s/ words, 10 critical /ʃ/ words, 30 filler words, and 50 nonwords. Trials within each half of the exposure block were presented randomly for each participant. For those in the Bias–Clear conditions, ambiguous fricatives appeared in a biasing context in the first half of the exposure block and no ambiguous fricatives were heard in the second half of the block. For those in the Clear–Bias conditions, clear fricatives were heard in the first half of the exposure block followed by ambiguous fricatives in the second half of the block. For example, listeners assigned to the SS bias condition in the Bias–Clear order heard 10 ambiguous fricatives in /s/-biasing contexts (and 10 clear /ʃ/ items) interspersed in the first 100 exposure trials, and then heard 10 clear /s/ items (and 10 clear /ʃ/ items) interspersed in the second 100 exposure trials (Figure 1). On each exposure trial, participants indicated whether the item was a word or not by pressing one of two keys on the keyboard.

The test phase was identical to that described for experiment 1. For both the training and test phases, trials were separated by 1000 ms, timed from the participant's response. The entire

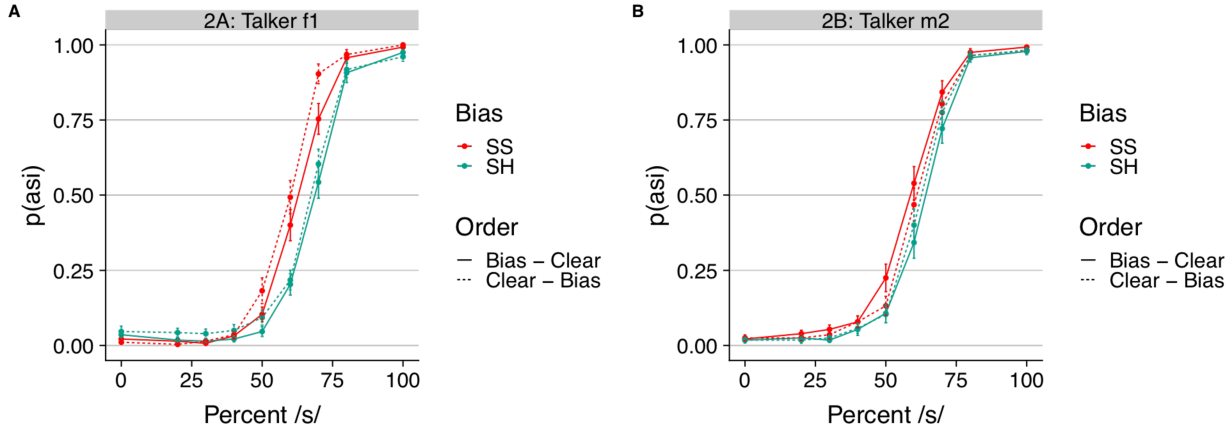
procedure lasted approximately 20 minutes.

Statistical analysis. Trial-level responses (0 = *ashi*, 1 = *asi*) were submitted to a GLMM with the fixed effects of continuum step (entered as a scaled/centered continuous variable), bias (SH = -0.5, SS = 0.5), order (Clear–Bias = -0.5, Bias–Clear = 0.5), and all interactions among the three factors. The random effects structure consisted of random intercepts by subject and random slopes by subject for continuum step, reflecting the maximal random effects structure given the experimental design.

Results

Experiment 2A. Performance at test is shown in Figure 3A. Model results, shown in Table 3, revealed a main effect of continuum step ($p < 0.001$) and an interaction between continuum step and bias ($p = 0.001$), the latter indicating the presence of a learning effect that varied in magnitude across the test continuum. No other main effects or interactions were reliable, including the interaction between bias and order ($p = 0.868$).

Figure 3. Mean proportion *asi* responses as a function of continuum step for each bias and order condition in experiment 2A (panel A) and experiment 2B (panel B). Continuum step is presented in terms of percent /s/ energy in each step of the test continuum. Means reflect grand means calculated over by-subject means; error bars indicate standard error.



Lexically guided perceptual learning was observed in experiment 2A, but learning was not influenced by the order in which ambiguous productions were encountered. To confirm this

interpretation, likelihood ratio tests were used to compare the omnibus model described above to a simpler model in which bias and order were successively removed as fixed effects. There was a significant change to goodness of fit when bias was included as a fixed effect ($\chi^2(2) = 39.876$, $p < 0.001$); however, there was no significant change to goodness of fit when order was further included in the model ($\chi^2(4) = 7.071$, $p = 0.132$).

Experiment 2B. Figure 3B shows performance at test; model results are shown in Table 3. There was a main effect of continuum step ($p < 0.001$) and a main effect of bias ($p = 0.005$). No other main effect or interaction was reliable, including the interaction between bias and order ($p = 0.158$). Likelihood ratio tests were used to compare the omnibus model to a simpler model in which bias and order were successively removed as fixed effects. There was a significant change to goodness of fit when bias was included as a fixed effect ($\chi^2(2) = 9.142$, $p = 0.010$); however, there was no significant change to goodness of fit when order was further included in the model ($\chi^2(4) = 2.473$, $p = 0.649$).[3]

The results of experiments 2A and 2B converged to show no evidence that learning that was contingent on the order in which ambiguous and clear productions were encountered. However, inspection of the beta coefficients for the bias by order interactions (shown in Table 3) reveals a considerable difference in the effect size estimates for the two talkers. Given the

[3] In Kraljic et al. (2008), performance at test was analyzed using ANOVA after first collapsing across steps of the test continuum. Parallel analyses were performed for experiment 2 in order to examine whether the different pattern of results could be attributed to the different analysis approaches. For experiment 2A, mean proportion *asi* responses was calculated for each participant by collapsing across continuum step. These values were submitted to an ANOVA with the between-subjects factors of bias and order. The ANOVA showed a main effect of bias [$F(1,136) = 23.871$, $p < 0.001$] and no interaction between bias and order [$F(1,136) = 0.263$, $p = 0.609$]. The same procedure was used for experiment 2B. The ANOVA showed a main effect of bias [$F(1,136) = 7.706$, $p = 0.006$] and no interaction between bias and order [$F(1,136) = 2.103$, $p = 0.149$]. The results of these ANOVAs converge with the GLMMs reported in the main text.

contrast-coding structure (i.e., -0.5 vs. 0.5 for each level of bias and order), the effect size of the interaction can be derived by dividing the beta coefficient by two; likewise, the uncertainty of the beta estimate can be derived by dividing the standard error by two. The effect size for the bias by order interaction was -0.052 (*SE* = 0.314) in 2A and 0.392 (*SE* = 0.278) in 2B; both effect sizes show considerable uncertainty. To increase power for detecting potential order effects, data from 2A and 2B were analyzed together (see Supplementary Material), and the bias by order interaction was not significant in this model ($\hat{\beta}$ = 0.358, *SE* = 0.420, *z* = 0.852, *p* = 0.394). The corresponding effect size for the bias by order interaction in this model was 0.179 (*SE* = 0.210), which falls intermediate to the effect sizes observed in the individual models and has slightly greater precision as indexed by a smaller standard error.[4]

Experiment 3

In contrast to Kraljic et al. (2008), the results of experiment 2 provided no evidence of a "first impressions" effect; perceptual learning occurred regardless of the order in which the atypical productions were encountered. In experiment 3, exposure was inconsistent throughout the exposure block, as in experiment 2, but listeners heard ambiguous fricatives in *both* biasing contexts and we manipulated the order in which the biasing contexts were encountered. Across conditions, listeners were exposed to either /s/- and then /ʃ/-biasing contexts (the SS–SH group) or the reverse order (the SH–SS group) to examine whether recent exposure or cumulative

[4] The Supplementary Material presents three additional sets of analyses to complement those presented in the main text. The first set collates data across the A/B versions of each experiment by including random intercepts by talker to models that are identical to the fixed effect structure described in the main text. The second set also collates data across the A/B versions of each experiment, but talker (and all interactions with talker) are included as additional fixed effects to the models described in the main text. The third set of analyses are identical to those presented in the main text except that the data are limited to trials in the first half of the test block (described further in the Discussion). In all cases, the models presented in the Supplementary Material converge with those presented in the main text.

exposure determine the extent of perceptual learning. If the most recent exposure is the putative factor for lexically guided perceptual learning, then listeners in the SH–SS condition should show more *asi* responses at test compared to those in the SS–SH condition. The global statistics hypothesis predicts no difference at test between the two exposure groups, given that cumulative experience to biasing contexts is equivalent.

Methods

Participants. Experiments 3A and 3B each tested 70 participants; within each experiment, 35 participants were randomly assigned to one of the two bias/order conditions (SH-SS vs. SS-SH). All participants met the inclusion criteria described for experiment 1 and were paid $3.33 for their participation.

Stimuli. The stimuli described in experiment 1 were used in experiment 3.

Procedure. Stimuli from talker f1 were used in experiment 3A and stimuli from talker m2 were used in experiment 3B. As described previously, the study consisted of a headphone screen, an exposure phase, and a test phase that was completed online using the web-based Gorilla platform. All listeners completed one lexical decision exposure block consisting of 200 trials. The 200 exposure items described for experiment 1 were randomly assigned to either the first or second half of the exposure block so that the first 100 trials and the second 100 trials each contained 10 critical /s/ words, 10 critical /ʃ/ words, 30 filler words, and 50 nonwords. Trials within each half of the exposure block were presented randomly for each participant. For those in the SH-SS conditions, the first half of the exposure block contained ambiguous fricatives in /ʃ/-biasing contexts and clear /s/ tokens; the second half of the exposure block contained ambiguous fricatives in /s/-biasing contexts and clear /ʃ/ tokens. Listeners in the SS-SH conditions heard the same tokens but in the opposite order. On each exposure trial, participants indicated whether the

item was a word or not by pressing one of two keys on the keyboard.

The test phase was identical to that described for experiments 1. For both the training and test phases, trials were separated by 1000 ms, timed from the participant's response. The entire procedure lasted approximately 20 minutes.

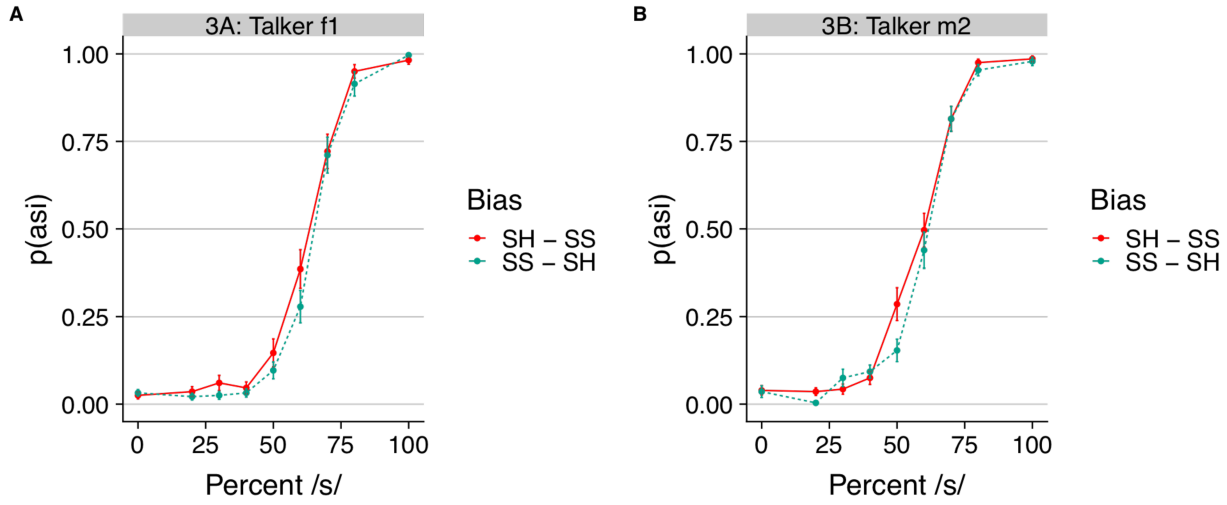
Statistical analysis. Trial-level responses (0 = *ashi*, 1 = *asi*) were submitted to a GLMM with the fixed effects of continuum step (entered as a continuous variable), bias (SS–SH = -0.5, SH–SS = 0.5), and the interaction between step and bias. The maximal random effects structure was used, consisting of random intercepts by subject and random slopes by subject for continuum step.

Results

Experiment 3A. Figure 4A shows performance at test; model results are shown in Table 3. The model revealed a main effect of continuum step ($p < 0.001$). There was no main effect of bias ($p = 0.590$) nor an interaction between step and bias ($p = 0.968$). A likelihood ratio test showed no change in goodness of fit between the omnibus model and a simpler model in which bias was removed as a fixed effect ($\chi^2(2) = 0.769$, $p = 0.681$).

Experiment 3B. Figure 4B shows performance at test; model results are shown in Table 3. The model revealed a main effect of continuum step ($p < 0.001$). There was no effect of bias ($p = 0.191$) and no interaction between step and bias ($p = 0.412$). A likelihood ratio test showed no significant change in goodness of fit between the omnibus model and a simpler model in which bias was removed as a fixed effect ($\chi^2(2) = 1.679$, $p = 0.432$).

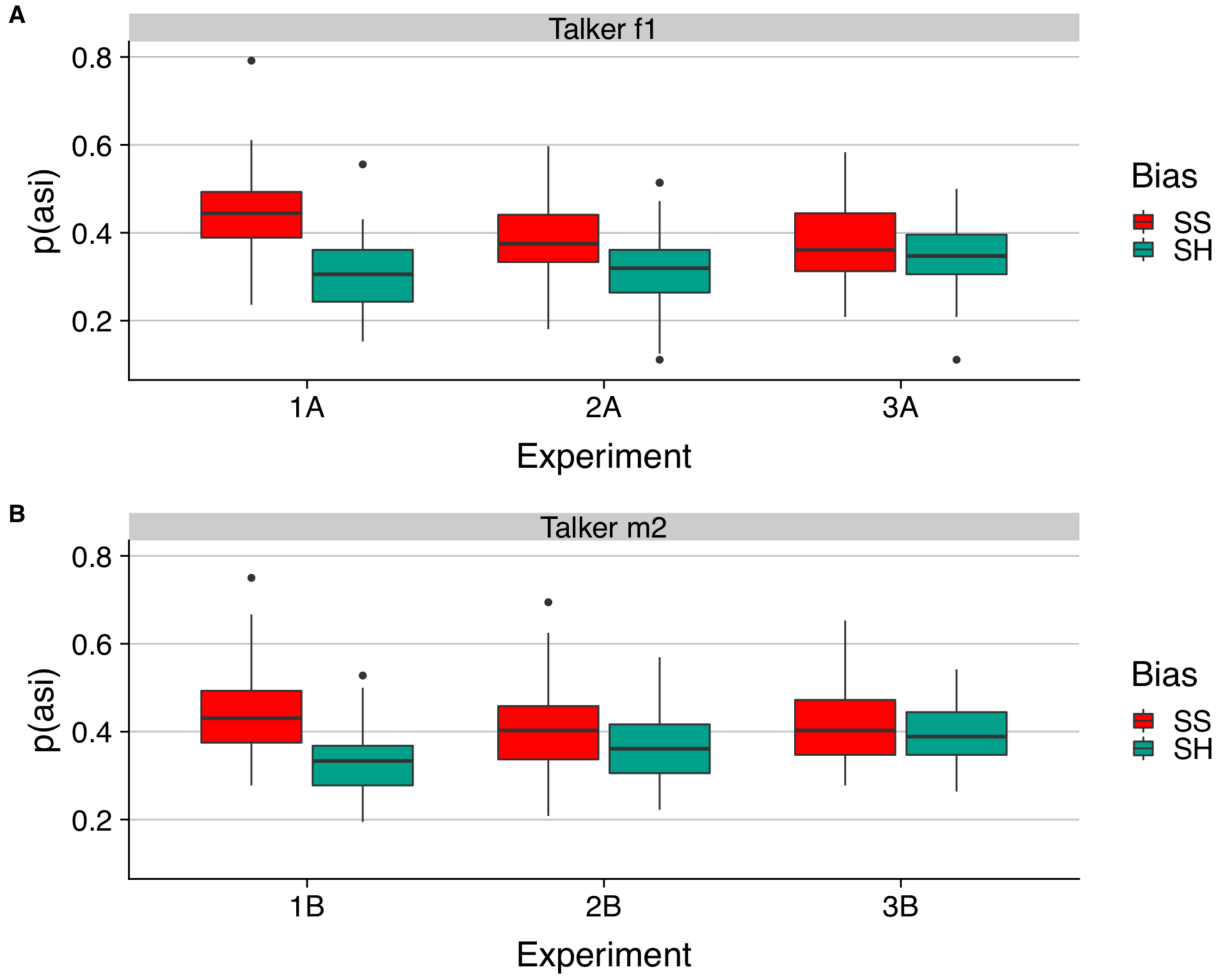
Figure 4. Mean proportion *asi* responses as a function of continuum step for each bias/order condition in experiment 3A (panel A) and experiment 3B (panel B). Continuum step is presented in terms of percent /s/ energy in each step of the test continuum. Means reflect grand means calculated over by-subject means; error bars indicate standard error.



Comparisons across experiments. A final analysis was conducted to compare performance across the three experiments. To do so, we collapsed across order conditions in experiment 2 (given no evidence that learning in experiment 2 was influenced by order). Bias in experiment 3 was coded to reflect the most recent bias. Figure 5 shows the distribution of proportion *asi* responses at test across participants (collapsing across continuum step) in each bias condition for each experiment. Visual inspection suggests a monotonic decrease in the magnitude of the learning effect across experiments, consistent with decreased exposure to regularity in ambiguous productions in the putative lexical context. This interpretation is also supported by examination of the effect sizes for the bias effect in each experiment (shown in Table 3), which are approximately halved from experiment 1 to experiment 2 and from experiment 2 to experiment 3.

To examine this pattern statistically, trial-level responses (0 = *ashi*, 1 = *asi*) were fit to a GLMM with the fixed effects of bias, experiment, and their interaction. Bias was sum-coded (SH = -0.5, SS = 0.5). Experiment was entered into the model as two sliding contrasts, one that

Figure 5. Boxplots for participants' proportion *asi* responses in each bias condition for each experiment. Panel A shows performance for the talker f1 stimulus set; panel B shows performance for the talker m2 stimulus set. As described in the main text, performance for 2A and 2B is shown collapsed across order conditions and performance for 3A and 3B is coded to reflect the most recent bias (i.e., those in the SH–SS bias/order condition are shown as SS; those in the SS–SH bias/order condition are shown as SH).



compared performance between experiment 2 and experiment 1 (E1 = -2/3, E2 = 1/3, E3 = 1/3), and one that compared performance between experiment 3 and experiment 2 (E1 = -1/3, -E2 = -1/3, E3 = 2/3). Contrasts are listed in terms of the generalized inverse of the matrix used in the contrasts() function in R, as specified by contr.sdif(3) in the MASS package (Venables & Ripley, 2002). The random effects structure included random intercepts by subject, random slopes for continuum step by subject, and random intercepts by talker. The results of this model showed a

smaller learning effect in experiment 2 compared to experiment 1 ($\hat{\beta}$ = -0.853, *SE* = 0.291, *z* = -2.935, *p* = 0.003; 95% CI = -1.423, -0.283) and a smaller learning effect in experiment 3 compared to experiment 2 ($\hat{\beta}$ = -0.767, *SE* = 0.289, *z* = -2.667, *p* = 0.008; 95% CI = -1.330, -0.203). These results confirm a monotonic decrease in learning across experiments, in line with decreasing exposure to stable lexically biasing contexts.

Discussion

This study investigated timecourse and exposure characteristics that lead to robust lexically guided perceptual learning. The results provide support for cumulative registration of talker-dependent variation in the acoustic speech signal, consistent with the global statistics hypothesis. Robust perceptual learning was observed in experiment 1, where listeners heard 20 ambiguous productions in a consistent lexically-biased context during the exposure phase. In experiment 2, learning was again observed even though listeners heard only 10 ambiguous productions in a consistent lexically-biased context (along with 10 clear productions in the same context). Moreover, there was no evidence indicating that learning was influenced by the order in which the ambiguous vs. clear productions were encountered. No evidence of learning was observed in experiment 3 where listeners heard 10 ambiguous productions in each of the two lexically-biased contexts (i.e., s-bias *and* ʃ-bias). These results held across two different talkers' idiosyncratic productions, suggesting that these experiments indexed general properties of adaptation and learning in speech perception. English was the language examined here, and future research is needed in order to examine whether these patterns will generalize to other languages (e.g., Burchfield et al., 2017; Chan et al., 2020; Norris et al., 2003).

Across experiments, perceptual learning was dependent on cumulative and consistent exposure to ambiguous tokens in lexically biasing contexts. This pattern of results provides

evidence that listeners track detailed input statistics of their listening experience over time and use that experience to adjust acoustic-phonetic category structure to reflect cumulative summary distributions of pronunciation variants. These findings are broadly consistent with the Bayesian belief updating model of speech adaptation (Kleinschmidt & Jaeger, 2015) and other accounts (e.g., Goldinger, 1998) that posit dynamic sensitivity to shifting acoustic-phonetic instantiations of individual talkers as a mechanism for resolving extensive variation in spoken language.

That learning was cumulative contrasts with previous findings (Kraljic et al., 2008); listeners in the current work did not privilege either initial or most recent exposure, but rather registered variation across exposure. Regarding the lack of an order effect in experiment 2, it may be the case that the current design was insufficiently powered to detect the "first impressions" effect reported in Kraljic et al. (2008), even though the sample size was comparable between the two studies.[5] That listeners did not exhibit a reliance on local statistics points to constraints on perceptual learning, which may be contingent on the degree to which the learning assay promotes a "reset" in the registration of cumulative statistics. In the current assay, listeners were exposed first to one and then to the other biasing context, and learning was assessed at the end of the entire exposure. If learning is assessed at the end of exposure for each biasing context, then sensitivity to more local input may emerge.

In the current study, learning was assessed for conditions that differed in the consistency of the mappings between critical productions and biasing contexts. Diminished perceptual

[5] Given recent evidence that learning in this paradigm may become attenuated during the test session (Liu & Jaeger, 2018, 2019), it is also possible that the order by bias interaction may have emerged given a shorter test period. To examine this possibility, all models presented in the main text were re-run, limiting data to the first half of the test session. These models showed the same qualitative patterns that are reported in the main text and are presented in detail in the Supplementary Material.

learning was found for conditions with fewer and less consistent ambiguous productions, leading to a monotonic decrease in learning across the three experiments, in line with diminished consistent exposure to the talker's idiosyncratic productions. These results indicate that lexically guided perceptual learning is not binary, but rather a graded outcome that is tightly linked to input statistics. This reliance on cumulative registration of acoustic-phonetic variation mirrors findings in other auditory domains, perhaps suggesting a general principle of auditory and perhaps perceptual processing more generally (e.g., Baese-Berk et al., 2014; McAuley & Miller, 2007). Future research should assess under what conditions listeners reset statistical tracking and how task-related factors influence the time course and extent of perceptual learning.

Acknowledgements

This research was supported by National Science Foundation (BCS) grant 1827591 to RMT. The views expressed here reflect those of the authors and not the National Science Foundation.

Open Practices Statement

Trial-level data and analysis scripts (in R) are available at https://osf.io/gz8jn/. The scripts operate on trial-level data to reproduce all statistics reported in the main text and in the Supplementary Material.

References

Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America*, *113*(1), 544–552.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2019). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 1–20. https://doi.org/10.3758/s13428-019-01237-x

Baese-Berk, M. M., Heffner, C. C., Dilley, L. C., Pitt, M. A., Morrill, T. H., & McAuley, J. D. (2014). Long-term temporal tracking of speech rate affects spoken-word recognition. *Psychological Science*, *25*(8), 1546–1553.

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

Bertelson, P., Vroomen, J., & De Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science*, *14*(6), 592–597.

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

Chan, L., Johnson, K., & Babel, M. (2020). Lexically-guided perceptual learning in early Cantonese-English bilinguals. *The Journal of the Acoustical Society of America*, *147*(3), EL277–EL282.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099–3111.

Idemaru, K., & Holt, L. L. (2011). Word recognition reflects dimension-based statistical learning. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(6), 1939.

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.

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., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, *19*(4), 332–338.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

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

Liu, L., & Jaeger, T. F. (2019). Talker-specific pronunciation or speech error? Discounting (or not) atypical pronunciations during speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(12), 1562.

Long, J. A. (2020). *jtools: Analysis and presentation of social scientific data* (R package version 2.0.0) [Computer software]. https://cran.r-project.org/package=jtools

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, *32*(3), 543–562.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*(1), 122–134.

McAuley, J. D., & Miller, N. S. (2007). Picking up the pace: Effects of global temporal context on sensitivity to the tempo of auditory sequences. *Perception & Psychophysics*, *69*(5), 709–718.

McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science*, *12*(3), 369–378.

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

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Attention, Perception, & Psychophysics*, *60*(3), 355–376.

Saltzman, D., & Myers, E. (2020). Retraction Note: Listeners are maximally flexible in updating phonetic beliefs over time. *Psychonomic Bulletin & Review*, *27*, 819. https://doi.org/10.3758/s13423-020-01765-0

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

Theodore, R. M., & Miller, J. L. (2010). Characteristics of listener sensitivity to talker-specific phonetic detail. *The Journal of the Acoustical Society of America*, *128*(4), 2090–2099.

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

Theodore, R. M., & Monto, N. R. (2019). Distributional learning for speech reflects cumulative exposure to a talker's phonetic distributions. *Psychonomic Bulletin & Review*, *26*(3), 985–992.

Theodore, R. M., Myers, E. B., & Lomibao, J. A. (2015). Talker-specific influences on phonetic category structure. *The Journal of the Acoustical Society of America*, *138*(2), 1068–1078.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (4th ed.). Springer.

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