Lexically guided perceptual learning is robust to task-based changes in listening strategy[a)]

Julia R. Drouin[b)] and Rachel M. Theodore[c)]
Department of Speech, Language, and Hearing Sciences, University of Connecticut, 850 Bolton Road, Unit 1085, Storrs, Connecticut 06269-1085, USA

(Received 16 March 2018; revised 29 June 2018; accepted 2 July 2018; published online 30 August 2018)

Listeners use lexical information to resolve ambiguity in the speech signal, resulting in the restructuring of speech sound categories. Recent findings suggest that lexically guided perceptual learning is attenuated when listeners use a perception-focused listening strategy (that directs attention towards surface variation) compared to when listeners use a comprehension-focused listening strategy (that directs attention towards higher-level linguistic information). However, previous investigations used the word position of the ambiguity to manipulate listening strategy, raising the possibility that attenuated learning reflected decreased strength of lexical recruitment instead of a perception-oriented listening strategy. The current work tests this hypothesis. Listeners completed an exposure phase followed by a test phase. During exposure, listeners heard an ambiguous fricative embedded in word-medial lexical contexts that supported realization of the ambiguity as /ʃ/. At test, listeners categorized members of an /ɑsi/-/ɑʃi/ continuum. Listening strategy was manipulated via exposure task (experiment 1) and explicit acknowledgement of the ambiguity (experiment 2). Compared to control participants, listeners who were exposed to the ambiguity showed more /ʃ/ responses at the test; critically, the magnitude of learning did not differ across listening strategy conditions. These results suggest that given sufficient lexical context, lexically guided perceptual learning is robust to task-based changes in listening strategy.
https://doi.org/10.1121/1.5047672

[TCB] Pages: 1089–1099

I. INTRODUCTION

Listeners are exposed to wide variability in the acoustic instantiation of individual speech sounds that arises due to numerous contextual factors including accent or dialect characteristics (e.g., Bradlow and Bent, 2008), coarticulatory effects (e.g., Summerfield, 1981), and idiolect characteristics of individual talkers (e.g., Hillenbrand *et al.*, 1995; Newman *et al.*, 2001; Theodore *et al.*, 2009). Remarkably, listeners extract meaningful information from this invariant signal with relative ease. While representations for speech sound categories begin to reflect native language phonology in the first year of life (e.g., Werker and Tees, 1984), they remain flexible to allow listeners to dynamically modify speech sound representations to accommodate systematic variability throughout the lifespan (e.g., Polka and Werker, 1994; Werker *et al.*, 2007; Werker *et al.*, 2012; Lively *et al.*, 1993; Pisoni, 1993; Wang *et al.*, 1999; Theodore *et al.*, 2015).

Previous research has shown that lexical knowledge is one factor that can guide interpretation of acoustic-phonetic input (Ganong, 1980), promoting changes to phonetic categories even when lexical context is subsequently removed (Norris *et al.*, 2003). For example, Norris *et al.* (2003) exposed Dutch listeners to a fricative ambiguous between /f/ and /s/ during an exposure task. Listeners in the /f/-bias exposure group heard the ambiguous sound in the context of Dutch words that ended in /f/ (i.e., "witlof"), whereas listeners in the /s/-bias exposure group heard the same ambiguous sound in the context of Dutch words that ended in /s/ (i.e., "naaldbos"). During exposure, listeners performed a lexical decision task for the critical exposure stimuli, along with other filler word and nonword items. After the exposure phase, learning was assessed using a phoneme identification task where listeners categorized items for a nonword /ɛf/ - /ɛs/ continuum. The results showed that listeners in the /f/-bias training group categorized more continuum items as /f/ compared to listeners in the /s/-bias group, demonstrating that the mapping to speech sound categories had been modified to include the ambiguous speech sound into the category consistent with experience during exposure. To confirm that this effect was driven by lexical knowledge, they demonstrated that the learning effect did not occur when the ambiguous fricative was embedded in nonword contexts during exposure.

Lexically guided perceptual learning (also referred to as phonetic recalibration) is very robust. It has been shown for numerous speech sound contrasts (e.g., McQueen *et al.*, 2006; Sjerps and McQueen, 2010; Jesse and McQueen, 2011; Mitterer *et al.*, 2013), and has been demonstrated to persist over time (Kraljic and Samuel, 2005; Eisner and McQueen, 2006). Its effects are not limited to the boundary region; instead, this type of learning promotes a

[a)]Portions of this work were presented at the 173rd Meeting of the Acoustical Society of America, Boston, MA, USA, June 2017.
[b)]Also at: Connecticut Institute for the Brain and Cognitive Sciences, University of Connecticut, 337 Mansfield Road, Storrs, Connecticut 06269, USA.
[c)]Electronic mail: rachel.theodore@uconn.edu

comprehensive reorganization of internal category structure (Drouin *et al.*, 2016; Xie *et al.*, 2017). However, recent research points towards constraints on lexically guided perceptual learning, highlighting a role for attentional processes on the engagement of this learning mechanism. For example, Scharenborg *et al.* (2015) operationalized attention as individual differences in attention-switching control, as indexed by performance on a standardized assessment. Their results showed that adults with poorer attention-switching abilities showed *enhanced* perceptual learning compared to adults with better attention-switching control. They hypothesize that this reflects a relationship between attention-switching ability and listening strategy such that those with better attention-switching were more likely to attend to phonetic information while those with poorer attention-switching were likely to rely on lexical information. On this view, the effect of attention-switching ability can be considered a distinction in listening strategy during exposure, where increased attention to surface variation leads to decreased perceptual learning (Scharenborg *et al.*, 2015).

Support for this view comes from McAuliffe and Babel (2016) who examined whether task-driven attention during exposure would modulate the magnitude of the lexically guided perceptual learning effect. Two manipulations were used to bias listeners towards either a perception-oriented listening strategy, where attention was directed towards surface variation, or a comprehension-oriented listening strategy, where attention was directed towards higher-level linguistic information. In the first manipulation, the ambiguous sound (midway between /s/ and /ʃ/) was either presented in a word-initial position (perception-oriented) or word-medial position (comprehension-oriented). In the second manipulation, listeners were either told explicitly that the speaker produces ambiguous /s/ sounds (perception-oriented) or were not given any mention of the ambiguity (comprehension-oriented). These two manipulations were fully crossed among four groups of participants. With this design, one condition was predominantly perception-oriented (word-initial + explicit knowledge of ambiguity), one condition was predominantly comprehension-oriented (word-medial + no explicit knowledge of ambiguity), with the other two conditions intermediate to contain one perception-oriented component and one comprehension-oriented component (word-initial + no explicit knowledge of ambiguity; word-medial + explicit knowledge of ambiguity). All participants completed a lexical decision exposure phase (during which the ambiguous fricative was presented in an s-biasing context) followed by a category identification test phase using members of /s/ - /ʃ/ word continua (e.g., *sin* - *shin*). One additional group of listeners served as a control condition; these listeners only completed the category identification test phase.

McAuliffe and Babel (2016) predicted that the comprehension-oriented listening strategies would promote lexically guided perceptual learning, while the perception-oriented listening strategies would attenuate learning. Their results partially supported these predictions. Specifically, while the intermediate conditions did not differ from control, providing evidence of attenuated perceptual learning, both the predominantly comprehension-oriented *and* predominantly perception-oriented listening strategies did show evidence of robust perceptual learning. McAuliffe and Babel (2016) reconciled these findings by suggesting that learning occurred in the predominantly comprehension-oriented condition due to attention to lexical information, and that it occurred in the predominantly perception-oriented condition due to increased encoding specificity, given the shared word position of the exposure and test fricatives (i.e., word-initial). Because the greatest magnitude of learning was observed in the most comprehension-oriented listening condition, the authors conclude that stimulus-directed attention attenuated lexically guided perceptual learning.

This finding is striking in the context of the broader lexically guided perceptual learning literature given myriad findings showing that learning is robust across exposure tasks that promote perception-focused listening strategies. For example, Samuel (2016) exposed participants to both a female voice and a distractor male voice during an exposure period. The female voice contained an ambiguous sound midway between /s/ and /ʃ/ embedded into words that would normally contain an /s/ or /ʃ/ in word-medial position and thus created a condition for lexically guided perceptual learning to occur. The male and female voices were presented dichotically on the same trial, but the male voice began 200 ms later and co-occurred with the onset of the ambiguous phoneme in the female voice. Following training, listeners categorized items for an /ɑsi/ to /ɑʃi/ continuum in the female voice heard during training. In one of a series of experiments, Samuel (2016) found that listeners showed robust phonetic recalibration even when performing a syllable counting task on the female's voice, suggesting that attention does not need to be explicitly drawn to lexical status for learning to occur. Moreover, a listener does not even need to attend to the acoustic stimuli for learning to occur, as other research has observed phonetic recalibration when listeners are asked to count each trial during the exposure phase (McQueen *et al.*, 2006).

Other work has observed lexically guided perceptual learning using visual monitoring tasks during the exposure phase. van Linden and Vroomen (2007) compared the effects of lipread and lexical information on phonetic recalibration by exposing Dutch participants to an ambiguous sound midway between /t/ and /p/. In the lipread condition, they embedded the ambiguous sound into Dutch pseudowords (i.e., /wo?/) and dubbed those pseudowords onto a video of a face articulating either a /p/ or /t/ interpretation (i.e., /wop/ or /wot/), where neither interpretation yields a real Dutch word. In the lexical condition, they embedded the ambiguous phoneme into Dutch words that ended in either a /p/ or /t/. Across conditions in the exposure phase, participants did not complete a phonetic task, but rather monitored a video for a white dot to appear either on the speaker's face (audiovisual condition) or on a black screen (lexical condition). Exposure trials were intermixed with categorization trials where they identified continuum items as either /p/ or /t/. They found statistically equivalent learning across both the lipread and lexical conditions suggesting that participants can use either source of information comparably to disambiguate the

signal. Of critical interest to the current study, they also concluded that while lipread and lexical conditions operate on different perceptual levels (i.e., bottom-up vs top-down information), both sources of information can be used to guide phonetic retuning, suggesting that attention need not be geared towards the lexicon to engage in this type of learning.

Supporting lexical context may be one reason for the discrepant findings showing that lexically guided perceptual learning is attenuated in some but not all perception-focused tasks (McAuliffe and Babel, 2016; McQueen *et al.*, 2006; Samuel, 2016; van Linden and Vroomen, 2007). McAuliffe and Babel (2016) noted that while learning was most robust in the condition that optimized comprehension-oriented listening strategy (no explicit knowledge of the ambiguity + ambiguity presented in the word-medial position), learning was attenuated to a similar degree for their intermediate conditions. It is unclear whether the observed attenuation reflects a perception-oriented listening strategy per se or the lack of surrounding lexical context (i.e., when the ambiguity was presented in the word-initial position). Research examining positional effects on phonetic recalibration has found that learning is attenuated when the ambiguous phoneme is presented in an onset position, while learning is robust when the ambiguity is presented in a coda position (e.g., Jesse and McQueen, 2011). Across many studies, learning has been observed when the ambiguity is presented in a word-medial position (e.g., Norris *et al.*, 2003; Drouin *et al.*, 2016) and word-final position (e.g., Mitterer *et al.*, 2013; Pitt and Samuel, 2006), suggesting that both of these positions offer sufficient access to lexical information. Indeed, the studies utilizing perception-focused tasks (Samuel, 2016; McQueen *et al.*, 2006; van Linden and Vroomen, 2007) never presented the ambiguity in a word-initial position, which supported access to local lexical context cues as the auditory stream unfolded. Thus, it is unclear whether diminished access to surrounding lexical context, rather than just induced listening strategy, was the putative factor for attenuated perceptual learning in McAuliffe and Babel (2016).

In addition to surrounding lexical context, another factor that may play a role in whether lexically guided perceptual learning is observed is the specific circumstances under which the listener is exposed to the talker. Previous research has shown that learning can be diminished through external factors beyond local lexical context. For example, Kraljic *et al.* (2008) suggested that listeners interact with the ambiguity differently depending on whether the production is characteristic of a speaker or not. They found that when listeners heard and simultaneously saw the speaker produce an ambiguous sound with a pen in the speaker's mouth, phonetic recalibration did not occur. However, when listeners only heard the ambiguity, but did not see the speaker with a pen in her mouth, learning did occur. They argued that if the listener has an external factor to attribute the ambiguity to (i.e., a pen in the speaker's mouth), learning does not occur since the production is not characteristic of the talker. However, if the production is considered characteristic of the speaker, the listener will restructure their phonetic categories to accommodate the variability. Moreover, other research has also shown that lexically guided perceptual learning may be talker-specific. If the voice listeners hear during exposure is different than that heard at test learning does not occur (Eisner and McQueen, 2005; Kraljic and Samuel, 2005) and fricatives may be more likely to show talker-specificity effects (Kraljic and Samuel, 2007). One way in which McAuliffe and Babel (2016) biased listeners towards a perception-orientation was explicitly telling listeners that the speaker produces ambiguous /s/ sounds. This instruction not only notifies the listener of the ambiguity, but also signals that this production is characteristic of the talker. While McAuliffe and Babel (2016) suggest that this manipulation should attenuate learning because the listener would be more likely to focus on the specific acoustic properties of the speech signal and less on lexical information (i.e., this manipulation induces a perception-oriented listening strategy), work by Kraljic *et al.* (2008) would predict that the listener would indeed learn with this manipulation because it is framed as characteristic of the talker.

Collectively, research has shown that attention towards surface level features of the signal may attenuate lexically guided perceptual learning in some but not all cases (McAuliffe and Babel, 2016, Samuel, 2016; McQueen *et al.*, 2006; van Linden and Vroomen, 2007). Moreover, the task instructions used to bias the listener towards a perception-focus may be sufficient to diminish learning (McAuliffe and Babel, 2016), but may also introduce talker-specificity effects that promote learning (Kraljic *et al.*, 2008). In the current study, we attempt to reconcile these discrepant findings by examining how lexically guided perceptual learning changes as a function of task-induced listening strategy while, critically, holding supporting lexical context constant.

In experiment 1, we manipulated listening strategy using a depth of processing task manipulation. All listeners, except for a control group, were assigned to an /ʃ/-bias exposure condition; the stimuli were identical across exposure conditions. However, the task the listener was asked to perform differed across conditions. During exposure, listeners completed a loudness judgment task (perception-focus), a lexical decision task (comprehension-focus), or a syntactic decision task (comprehension-focus). All listeners completed a categorization task following exposure. Previous research has found that phonetic recalibration can occur using perception-focused tasks like syllable counting (Samuel, 2016), trial counting (McQueen *et al.*, 2006) and visual monitoring (van Linden and Vroomen, 2007), and also using comprehension-focused tasks like lexical decision (e.g., Norris *et al.*, 2003) and semantic decision (Zhang and Samuel, 2014). However, it is unclear whether the magnitude of learning differs as task changes from a more perception-focus to comprehension-focus. If perceptual learning is weakened as a consequence of a perception-focused listening strategy, then we would expect to see attenuated learning in the amplitude decision task compared to the lexical decision and syntactic decision tasks, in line with previous findings (McAuliffe and Babel, 2016). However, if access to local lexical context, rather than induced listening strategy, is sufficient to guide lexically guided perceptual learning, then we would expect to observe

no difference in learning between the perception-focused task and comprehension-focused tasks.

The loudness judgment task was selected for the perception-focus listening strategy because it is a task that can be completed solely based on a perceptual analysis of the signal; that is, listeners do not require explicit (or implicit) linguistic processing in order to perform this task. Though amplitude may inherently differ as a function of speech sound class (e.g., vowels have higher amplitude than fricatives) findings from the literature on spoken word recognition have shown that specificity effects for recognition memory that are observed for speaking rate and talker do not extend to amplitude (Bradlow *et al.*, 1999), and trial-by-trial amplitude variation does not impede word recognition as does variation in talker or speaking rate (Sommers *et al.*, 1994). The lexical and syntactic decision tasks were selected for the comprehension-focused listening strategy because, in contrast to the amplitude judgment task, these two tasks require linguistic processing. While both of these require linguistic processing and thus can be considered a means to induce comprehension-focused listening strategy, they differ with respect to depth of processing, which is greater for the syntactic decision compared to lexical decision task. These two comprehension-focused tasks were selected in order to provide a built-in replication for learning in a comprehension-focused task, and to examine whether learning would be graded as a function of depth of processing (i.e., greater learning in the syntactic compared to the lexical condition). All tasks were designed to promote the targeted listening strategy (i.e., perception-focused vs comprehension-focused) across the set of exposure items as a gestalt, and not to specifically shift listening strategy for the ambiguous phoneme specifically.

In experiment 2, a lexical decision task was used during exposure, with listening strategy manipulated through explicit knowledge of the ambiguity (perception-focus) or no knowledge of the ambiguity (comprehension-focus). These two conditions were selected for their representative listening strategy following McAuliffe and Babel (2016) who induced a perception-focused listening strategy by explicitly telling listeners that the talker may produce ambiguous sounds. If explicit knowledge of the ambiguity draws listeners focus towards the acoustic signal and weakens lexical activation as hypothesized by McAuliffe and Babel (2016), then phonetic recalibration will be attenuated compared to those who did not have explicit knowledge of the ambiguity. However, if surrounding lexical context or knowledge that an ambiguous production is characteristic of a talker is sufficient to guide lexically guided perceptual learning, then we predict no difference in learning between listeners who had knowledge of the ambiguity and those who did not (Kraljic *et al.*, 2008).

II. EXPERIMENT 1

A. Methods

1. Participants

One hundred adults between the ages of 18 and 32 years ($M = 20$, $SD = 2$, 19 males) participated in experiment 1. Two additional participants were tested in the lexical condition, described below, but excluded due to low accuracy (<75% correct) during the exposure task. All were monolingual speakers of American English with no reported history of speech, language, or reading disorders. All participants passed a pure tone hearing screening (administered at 20 dB for octave frequencies between 500 and 4000 Hz) on the day of testing. Participants were randomly assigned to one of three exposure conditions (i.e., amplitude, lexical, syntactic, $n = 20$ in each condition) or the control condition ($n = 40$). The sample size was based on that used in previous work (e.g., Norris *et al.*, 2003; McAuliffe and Babel, 2016). Participants were compensated with either partial course credit or monetary payment.

2. Stimuli

The stimuli consisted of 200 exposure items and six test items drawn from those used in Myers and Mesite (2014) and Drouin *et al.* (2016), to which the reader is referred for comprehensive details on stimulus creation. The exposure stimuli consisted of 100 auditory words and 100 auditory nonwords produced by a native female speaker of American English. The 100 auditory words were divided into three classes: 20 ambiguous /ʃ/ words (e.g., *publi?er*), 20 clear /s/ words (e.g., *pencil*), and 60 filler words that contained no instance of /s/ or /ʃ/ (e.g., *napkin*). The ambiguous /ʃ/ words were created by recording both an /s/ and /ʃ/ version of the word (e.g., *publiser* and *publisher*) and excising the fricative portion (at zero crossings) of each version. The fricatives were then equated on duration by trimming the longer fricative to match the length of the shorter one. The two fricatives were blended using waveform averaging in Praat and the 50% blend was inserted back into the /s/-frame production (e.g., *publiser*). Using this procedure, a unique 50% blend was created for each ambiguous /ʃ/ word. Using the Praat software, amplitude for half of the tokens for each of the four item types (i.e., ambiguous /ʃ/ words, clear /s/ words, filler words, nonwords) was set to 60 dB and amplitude for the other half of the tokens was set to 70 dB.

The test stimuli consisted of a six-step nonword continuum produced by the same talker as for the exposure stimuli. The continuum perceptually ranged from /ɑsi/ to /ɑʃi/. The test continuum was created following a similar protocol for the exposure stimuli. First, recordings were made of naturally produced /ɑʃi/ and /ɑsi/ tokens. The fricatives in each token were excised and waveform averaging was used to create six unique fricative blends. In terms of proportion /ʃ/ energy, the blends included 20%, 30%, 40%, 50%, 60%, and 70%. The blends were then inserted back into the /s/ frame (i.e., /ɑsi/) to create six unique tokens. The amplitude of the six test tokens was set to 65 dB using Pratt.

3. Procedure

All participants in the three exposure conditions (i.e., amplitude, lexical, syntax) completed an exposure phase followed by a test phase. The exposure stimuli were identical among the three exposure conditions; the ambiguous fricative was embedded in a word-medial, /ʃ/-biasing lexical

context in all cases. Following McAuliffe and Babel (2016), participants in the control condition completed only the test phase. With this design, learning was assessed by comparing performance in the exposure conditions to the control group. This method of assessing learning is different from the standard lexically guided perceptual learning paradigm where two biasing groups are compared to each other (e.g., /ʃ/-bias versus /s/-bias). We adopted the same methodology as McAuliffe and Babel (2016) who used this design in light of the previously reported instances of asymmetrical learning in the lexically guided perceptual learning paradigm (Zhang and Samuel, 2014; Drouin *et al.*, 2016; Samuel, 2016).

All testing took place individually in a sound-attenuated booth. The stimuli were presented via headphones at a comfortable listening level that was held constant across participants. Stimuli presentation and data collection were controlled using the SuperLab software (version 4.5) on a Mac OS X operating system. Responses were collected using a button box. Within each condition, button assignment was counterbalanced by dominant hand for both exposure (YES/NO) and test (ASI/ASHI) responses.

a. Exposure. During exposure, participants heard one randomization of the 200 exposure items. On each trial, participants were asked to make a two-alternative forced-choice decision according to their exposure condition. Participants in the amplitude condition were asked to indicate whether the item was loud (YES) or not (NO). Participants in the lexical condition were asked to indicate whether the item was a word (YES) or not (NO). Participants in the syntactic condition were asked to indicate whether the item was a noun (YES) or not (NO). In all conditions, participants were encouraged to respond as quickly as possible without sacrificing accuracy. The interstimulus interval was 2000 ms, timed from the participant's response. If a participant failed to respond within 5000 ms of stimulus onset, then no response was recorded and the experiment advanced to the next trial.

b. Test. During test, participants completed a category identification task in which they heard six randomizations of the six test tokens. On each trial, they were instructed to categorize the token as either ASI or ASHI by pressing the appropriate button on the button box. The interstimulus interval was 2000 ms, timed from the participant's response. As for the exposure phase, failure to respond within 5000 ms resulted in the experiment advancing to the next trial and participants were encouraged to respond as quickly and accurately as possible.

The entire procedure lasted approximately 30 min for individuals in the three exposure conditions (i.e., exposure phase followed by test phase) and approximately 5 min for individuals in the control condition (i.e., only test phase).

B. Analysis and results

The exposure and test phases were analyzed separately. Performance for exposure was measured in terms of accuracy and reaction time for correct responses (RT); reaction time for a given trial was measured relative to the onset of the auditory stimulus. For the mixed-effects model described below, RTs were log-transformed to more closely approximate a normal distribution. Log RTs deviating more than three standard deviations from each participant's mean log RT were deemed outliers and excluded from further analysis; these outlier RTs accounted for less than 1% of the data (101 trials out of 10 566 correct responses). For each subject, mean accuracy and RT were calculated separately for each of the four item types. One participant in the amplitude condition was removed from further analysis because accuracy during exposure (49.79%) was near chance. Table I shows mean accuracy across participants in each of the three exposure conditions for the four item types; reaction time performance is shown in Table II. Accuracy was near ceiling in all cases, except for nonwords for those in the syntax condition. We had expected performance to be near ceiling here too, given that because these are nonwords, the correct response of "no" in the noun decision task should have been straightforward; however, it appears that participants in this condition were attempting to attribute morphosyntactic properties to the nonword items, consistent with results from investigations using "Jabberwocky" speech (e.g., Johnson and Goldberg, 2013).

Mean accuracy and RT for the three exposure conditions (collapsing across item type) are shown in Fig. 1. To compare performance during exposure among the three exposure conditions, accuracy and RT were analyzed in separate mixed-effects models. For accuracy, individual trial responses (0 = incorrect, 1 = correct) were fit to a generalized linear mixed-effects model using the glmer() function with the binomial response family as implemented in the lme4 package (Bates *et al.*, 2015) in R (http://www.r-project.org); all test statistics and p-values represent the calculations from the glmer() function. The model contained exposure as a fixed effect, which was entered as two

TABLE I. Mean accuracy (proportion correct) and standard error of the mean (in parentheses) during exposure for each exposure condition and each of the four item types. The "eLexical" label refers to participants who were explicitly told of the speaker's ambiguous /ʃ/ productions.

		Item Type			
Experiment	Exposure	/ʃ/	/s/	Filler	Nonword
1	Amplitude	0.947 (0.015)	0.939 (0.014)	0.943 (0.014)	0.932 (0.012)
	Lexical	0.955 (0.013)	0.990 (0.005)	0.940 (0.010)	0.924 (0.014)
	Syntax	0.865 (0.017)	0.927 (0.012)	0.891 (0.017)	0.769 (0.051)
2	eLexical	0.925 (0.023)	0.993 (0.004)	0.940 (0.008)	0.915 (0.019)
	Lexical	0.963 (0.011)	0.995 (0.003)	0.922 (0.008)	0.937 (0.008)

TABLE II. Mean reaction time (in milliseconds) and standard error of the mean (in parentheses) to correct responses during exposure for each exposure condition and each of the four item types in experiment 2. The "eLexical" label refers to participants who were explicitly told of the speaker's ambiguous /ʃ/ productions.

		Item Type			
Experiment	Exposure	/ʃ/	/s/	Filler	Nonword
1	Amplitude	812 (49)	830 (48)	826 (50)	816 (48)
	Lexical	923 (86)	929 (87)	906 (83)	1062 (100)
	Syntax	1469 (59)	1439 (74)	1498 (65)	1519 (79)
2	eLexical	1047 (52)	991 (32)	1028 (36)	1182 (51)
	Lexical	976 (34)	957 (30)	985 (32)	1103 (45)

orthogonal contrasts, one for Lexical vs Amplitude (Amplitude = –1/2, Lexical = 1/2, Syntax = 0) and one for Syntax vs Lexical and Amplitude (Amplitude = –1/3, Lexical = –1/3, Syntax = 2/3). The model also contained random intercepts by subject and item. The model showed no significant effect for the lexical vs amplitude contrast ($\beta = -0.160$, $SE = 0.268$, $z = -0.598$, $p = 0.550$), but revealed that accuracy in the syntax condition was significantly lower compared to the lexical and amplitude conditions ($\beta = -1.148$, $SE = 0.224$, $z = -5.131$, $p < 0.001$).

Trial-level, log-transformed response times were fit to a mixed-effects model using the lmer() function from the lme4 package (Bates *et al.*, 2015) in R; the Satterthwaite approximation of the degrees of freedom was used to calculate p-values as implemented using the lmerTest package (Kuznetsova *et al.*, 2015). The fixed and random effects structure was identical to that described above for the accuracy model. The model showed no significant difference in RT between the Lexical and Amplitude conditions ($\beta = 0.086$, $SE = 0.142$, $t = 0.608$, $p = 0.546$), though RT in the Syntax condition was significantly slower compared to the Lexical and Amplitude conditions ($\beta = 0.570$, $SE = 0.122$, $t = 4.680$, $p < 0.001$).

Performance at test was measured in terms of ASHI responses. Figure 1, panel (c) shows mean ASHI responses as a function of continuum step for the three exposure conditions and the control condition; continuum is plotted in terms of percent /ʃ/ energy in the fricative blend for each continuum step. Prior to calculating mean ASHI responses for the three exposure conditions, mean proportion ASHI responses was calculated separately for each participant for each step of the test continuum. Visual inspection suggests that all three exposure groups show evidence of perceptual learning in that there are more ASHI responses for these conditions compared to the control group; no robust differences among the three exposure conditions are readily apparent. To examine this pattern statistically, individual trial level responses (0 = ASI, 1 = ASHI) were fit to a generalized linear mixed-effects model using the glmer() function from the lme4 package in R with the binomial response family. Exposure was specified as a fixed effect by three orthogonal contrasts, one that examined performance between the control condition and the three exposure conditions (Control = −3/4,

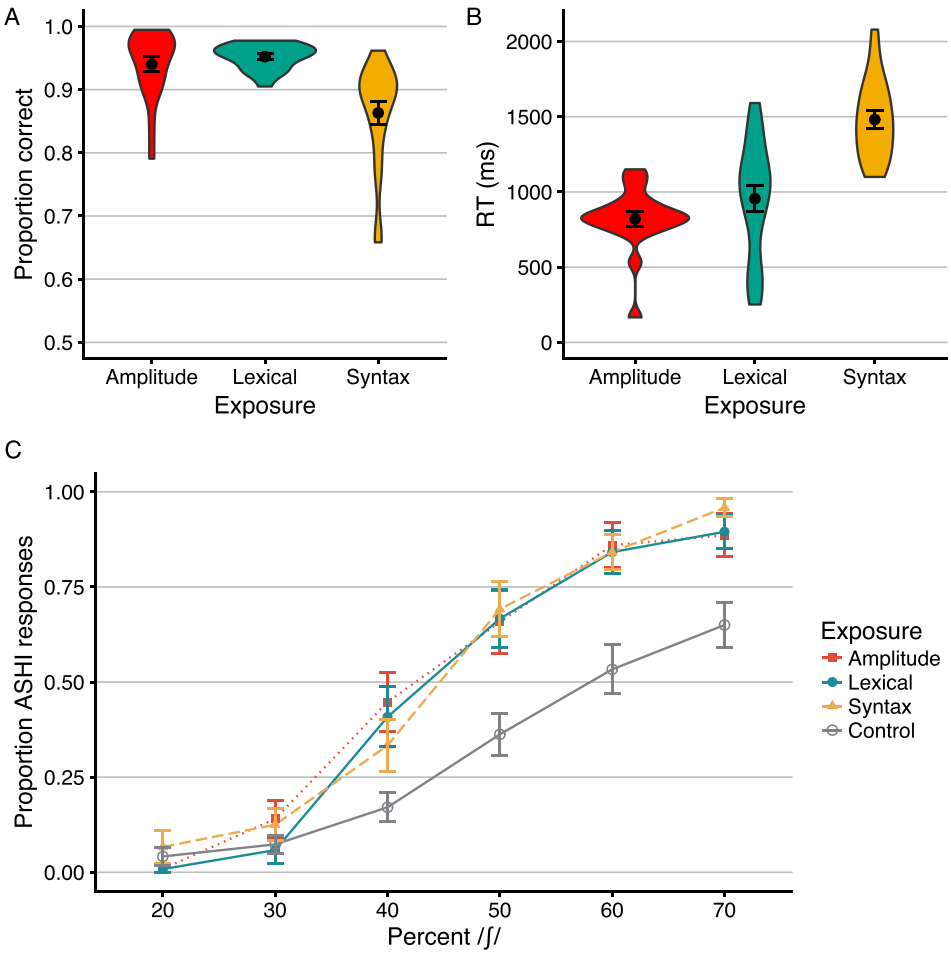

FIG. 1. (Color online) Results for experiment 1. Panel (A) shows mean proportion correct responses during exposure for each of the exposure conditions; panel (B) shows mean reaction time (RT, in ms) to correct responses during exposure. Panel (C) shows mean proportion ASHI responses during test as a function of percent /ʃ/ energy in the test continuum for each exposure condition and the control condition. In panels (A) and (B), the violin length indicates the response range across participants. The violin width shows a kernel density estimation to illustrate the distribution of responses across participants; wider regions indicate greater density of participant performance. Error bars indicate standard error of the mean.

Amplitude = 1/4, Lexical = 1/4, Syntax = 1/4), one that examined the comprehension-oriented conditions to the perception-oriented condition (Control = 0, Amplitude = −2/3, Lexical = 1/3, Syntax = 1/3), and one that examined performance for the Syntax compared to the Lexical condition (Control = 0, Amplitude = 0, Lexical = −1/2, Syntax = 1/2). The model also contained Continuum Step (scaled and centered around the mean, specified in terms of percent /ʃ/ energy) and the interaction between Continuum Step and Exposure as fixed effects, with random intercepts by subjects and random slopes by subject for Continuum Step. The model showed a main effect of Continuum Step ($\beta = 3.375$, $SE = 0.235$, $z = 14.362$, $p < 0.001$), indicating that ASHI responses increased as did percent /ʃ/ energy in the test continuum. There were more ASHI responses in the exposure conditions compared to the control condition ($\beta = 2.002$, $SE = 0.446$, $z = 4.493$, $p < 0.001$), which is evidence of perceptual learning. There was no significant difference in ASHI responses for the comprehension-oriented exposure conditions compared to the perception-oriented exposure condition ($\beta = -0.139$, $SE = 0.595$, $z = -0.233$, $p = 0.816$), nor was there a significant difference in ASHI responses between the Syntax and Lexical conditions ($\beta = -0.019$, $SE = 0.675$, $z = -0.029$, $p = 0.977$). The model showed a significant interaction between Continuum Step and Exposure for the exposure vs control contrast ($\beta = 1.371$, $SE = 0.403$, $z = 3.398$, $p < 0.001$); indicating that the learning effect was not the same at all continuum steps. As can be viewed in Fig. 1, panel (c), mean ASHI responses for three exposure groups separate from the control group near the midpoint of the test continuum, extending towards the /ɑʃi/ end of the continuum. The interaction between Continuum Step and Exposure was not reliable for the contrast between perception- and comprehension-oriented exposure conditions ($\beta = 0.240$, $SE = 0.557$, $z = 0.431$, $p = 0.666$) or for the contrast between the Lexical and Syntax exposure conditions ($\beta = -0.440$, $SE = 0.642$, $z = -0.684$, $p = 0.494$).

III. EXPERIMENT 2

In experiment 1, we observed processing differences during exposure among the three conditions; increased processing demands of the task resulted in increased reaction times, in line with depth of processing findings (Craik and Tulving, 1975). We also observed robust phonetic recalibration across all exposure conditions. All three experimental groups differed from the control group at test, and the magnitude of learning did not statistically differ among the exposure conditions, suggesting that attending to different aspects of the stimuli during exposure did not influence the robustness of lexically guided perceptual learning. The data from experiment 1 suggest that supporting lexical context is sufficient to guide phonetic recalibration, independently of listening strategy.

Findings from experiment 1 suggest that listening strategy did not influence lexically guided perceptual learning. However, listening strategy can be manipulated in a number of different ways. While experiment 1 used an exposure task to direct a perception- or comprehension-focused listening strategy, it has been argued that listening strategy can also be manipulated through pragmatic knowledge about the speaker. McAuliffe and Babel (2016) invoked a perception-oriented listening strategy by giving listeners explicit knowledge of potentially ambiguous input during exposure and found diminished perceptual learning compared to listeners who were not explicitly told of the ambiguity. Here we provide an additional test of the role of listening strategy on perceptual learning. Two groups of listeners were exposed to an ambiguous fricative using the lexical decision exposure task of experiment 1. One group of listeners was made explicitly aware of the anomalous input; the other group was not. If explicit knowledge of the ambiguity creates a barrier to perceptual learning, then the magnitude of learning observed at test will differ between the two exposure conditions. However, conclusions from Kraljic *et al.* (2008) support an alternative prediction; namely, that if listeners treat the explicit knowledge of the ambiguity as informing a characteristic of the speaker, then listeners will show robust lexically guided perceptual learning.

A. Methods

1. Participants

Forty participants between the ages of 18 and 31 years (M = 20, SD = 2; 22 males) participated in experiment 2 and were compensated with course credit or monetary payment. Participants were randomly assigned the explicit Lexical (eLexical, n = 20) or Lexical (n = 20) exposure conditions. The same participants (n = 40) for the control group in experiment 1 were also used as the control group in experiment 2. All participants were monolingual speakers of American English with no reported history of speech, language, or reading disorders. All participants passed a pure tone hearing screening (administered at 20 dB for octave frequencies between 500 and 4000 Hz) on the day of testing.

2. Stimuli

The exposure and test stimuli were identical to those used in experiment 1

3. Procedure

For participants in the lexical group, the procedure was identical to that used for the participants in lexical exposure condition of experiment 1. For participants in the explicit lexical group (i.e., the eLexical condition), the procedure was the same with one exception: these participants were given the following extra set of instructions during exposure, "be aware that this speaker's /ʃ/ sounds are sometimes ambiguous, or sound funny, so listen carefully so as to choose the correct response." This extra instruction was used to orient this group of listeners towards a perception-focused listening strategy following the instructions provided by McAuliffe and Babel (2016) in their perception-oriented manipulation.

B. Analysis and results

The exposure and test phases were analyzed separately. Performance during exposure was measured in terms of

accuracy and reaction time to correct responses (RT) as described for experiment 1. Reaction times were log-transformed to better approximate a normal distribution, and outlier RTs (defined as log RTs exceeding three standard deviations of each participant's mean log RT) were excluded from further analysis; these outliers consisted of 78 RTs out of 7486 correct responses, representing 1% of the total RT data. For each subject, mean accuracy and RT were calculated separately for each of the four item types; mean accuracy and RT across participants in each exposure group is shown in Tables I and II, respectively.

Performance between the two exposure conditions for accuracy and RT was compared in separate mixed-effects models. For accuracy, individual trial responses (0 = incorrect, 1 = correct) were fit to a generalized linear mixed-effects model using glmer() in R with exposure as a fixed effect (contrast-coded; eLexical = −0.5, Lexical = 0.5) and random intercepts by subject and by item. The effect of exposure was not reliable ($\beta = 0.125$, $SE = 0.201$, $z = 0.624$, $p = 0.533$). For RT, trial-level log-transformed reaction times to correct trials (excluding outlier log RTs) were fit to a linear mixed-effects model using lme4() in R with exposure as a fixed effect (contrast-coded; eLexical = −0.5, Lexical = 0.5) and random intercepts by subject and by item; the model showed no main effect of exposure ($\beta = -0.040$, $SE = 0.047$, $t = -0.844$, $p = 0.404$).

Performance at test was measured in terms of ASHI responses. Figure 2, panel (c) shows mean ASHI responses as a function of continuum step for the two exposure conditions and the control condition; note that the control condition here is the same group of participants used as the control condition in experiment 1. Prior to calculating mean ASHI responses for the two exposure conditions, mean proportion ASHI responses was calculated separately for each participant for each step of the test continuum. To analyze performance among the conditions, individual trial-level responses (0 = ASI, 1 = ASHI) were fit to a generalized linear mixed-effects model using glmer() in R, as described for experiment 1. Exposure was entered as a fixed effect specified by two orthogonal contrasts, one that examined performance in the two exposure conditions to the control condition (Control = −2/3, eLexical = 1/3, Lexical = 1/3) and one that examined performance in the perception-oriented exposure condition to the comprehension-oriented exposure condition (Control = 0, eLexical = −0.5, Lexical = 0.5). The model also contained Continuum Step as a fixed effect (scaled and centered around the mean, specified in terms of percent /ʃ/ energy), random intercepts by subjects, and random slopes by subject for Continuum Step. The model revealed a main effect of Continuum Step ($\beta = 3.322$, $SE = 0.246$, $z = 13.516$, $p < 0.001$), indicating more ASHI responses were provided as /ʃ/ energy increased in the test continuum. There were more ASHI responses in the exposure conditions compared to the control condition ($\beta = 1.737$, $SE = 0.511$, $z = 3.397$, $p < 0.001$), but there was no significant difference in ASHI responses between the eLexical and Lexical exposure

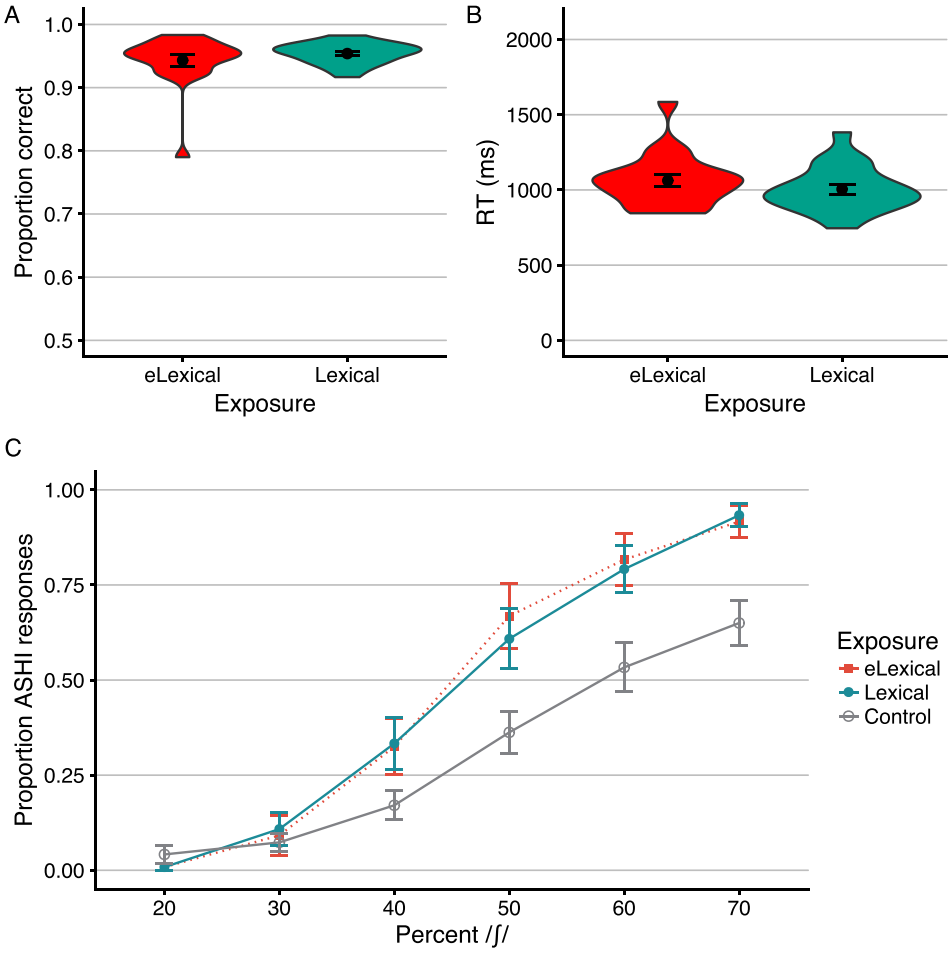

FIG. 2. (Color online) Results for experiment 2. Panel (A) shows mean proportion correct responses during exposure for each of the exposure conditions; panel (B) shows mean reaction time (RT, in ms) to correct responses during exposure. The "eLexical" label refers to participants who were explicitly told of the speaker's ambiguous /ʃ/ productions. Panel (C) shows mean proportion ASHI responses during test as a function of percent /ʃ/ energy in the test continuum for each exposure condition and the control condition. In panels (A) and (B), the violin length indicates the response range across participants. The violin width shows a kernel density estimation to illustrate the distribution of responses across participants; wider regions indicate greater density of participant performance. Error bars indicate standard error of the mean.

conditions ($\beta = -0.211$, $SE = 0.712$, $z = -0.296$, $p = 0.767$). There was no interaction between Continuum Step and Exposure for the Lexical vs eLexical contrast ($\beta = -0.582$, $SE = 0.600$, $z = -0.970$, $p = 0.332$), but there was a reliable interaction between Exposure and Continuum Step for the contrast comparing the exposure conditions (eLexical and Lexical) to the control group ($\beta = 1.515$, $SE = 0.410$, $z = 3.694$, $p < 0.001$). This latter interaction confirms that the learning effect for the exposure conditions was not equivalent across the test continuum; as shown in Fig. 2, panel (c), mean ASHI responses for the two exposure conditions are similar to the control condition at the /s/ end of the test continuum, with performance separating near the midpoint of the continuum.

IV. DISCUSSION

The goal of the current set of experiments was to examine the influence of listening strategy on lexically guided perceptual learning. Previous research has shown equivocal results. On the one hand, some findings suggest that learning may be attenuated when listeners use a perception-oriented listening strategy during exposure (McAuliffe and Babel, 2016; Scharenborg *et al.*, 2015). On the other hand, some findings document robust lexically guided perceptual learning even when the exposure task directs attention towards surface features of the stimulus (Samuel, 2016; McQueen *et al.*, 2006; van Linden and Vroomen, 2007). Here, we directly compared the magnitude of the lexically guided perceptual learning effect as a consequence of exposure tasks that were designed to differentially promote perception or comprehension-oriented listening strategies. Critically, the exposure (and test) stimuli were identical across exposure conditions, and thus listening strategy was manipulated independently of lexical context.

In experiment 1, there was no evidence that the magnitude of perceptual learning differed depending on whether attention was directed towards amplitude, lexical, or syntactic properties of the stimuli during exposure. In experiment 2, there was no evidence indicating that knowledge of the speaker's atypical productions diminished learning compared to those were not explicitly told of the ambiguity. Thus, the results of both experiments converge to provide no evidence that learning was attenuated for perception-oriented compared to comprehension-oriented listening strategies. This finding contributes to a wide body of literature demonstrating the robustness and automaticity associated with phonetic retuning. Previous research has shown that lexically guided perceptual learning can create changes to speech sound categories that are resistant to time delay (Eisner and McQueen, 2006) and that generalize across word position (Jesse and McQueen, 2011), speakers (Kraljic and Samuel, 2006, 2007), speech sounds (e.g., Kraljic and Samuel, 2006), and languages (Reinisch *et al.*, 2013). Lexically guided perceptual learning may also occur regardless of whether the listener is asked to engage with the auditory input explicitly (e.g., McQueen *et al.*, 2006). Indeed, previous research has suggested that local lexical information may act as a teaching signal during phonetic categorization, providing direct feedback as to how that speech sound should be classified, and may be a process that is automatically initiated without explicit attentional resources (e.g., Norris *et al.*, 2003; Samuel, 2016). Our work is in line with these findings and demonstrates that for the perception-focused tasks employed here, supporting lexical context is sufficient to allow phonetic recalibration to occur.

Our results of robust phonetic recalibration regardless of listening strategy differ from conclusions of previous work that observed attenuated perceptual learning (McAuliffe and Babel, 2016). McAuliffe and Babel (2016) suggested that listening strategy underlies why perceptual learning was attenuated under some of their perception-oriented listening manipulations. An alternative explanation could be the lexical context in which the ambiguity was presented. In the current experiments, supporting lexical context was held constant during exposure, with the ambiguity always presented in a word-medial position. This position allows the listener access to surrounding lexical cues as the auditory stream unfolds, which is weakened when the ambiguity is presented in a word-initial position. Previous research examining positional effects on lexically guided perceptual learning has demonstrated attenuated perceptual learning when the ambiguity is placed in a word-initial position (Jesse and McQueen, 2011). Another distinction between current and previous work is the nature of the test continuum. Learning was assessed in McAuliffe and Babel (2016) for word continua (e.g., *sin - shin*), with the critical fricative presented in the word-initial position. The current study assessed learning in a nonword context, in line with the standard lexically guided perceptual learning paradigm (e.g., Kraljic and Samuel, 2005; Myers and Mesite, 2014; Samuel, 2016; Drouin *et al.*, 2016). Using a lexical context at test may have implicitly affected how learning was assessed given findings demonstrating influences of lexical status and word frequency on phonetic categorization (Ganong, 1980; Fox, 1984). Moreover, the position of the critical fricative (i.e., word-medial) was held constant between exposure and test in the current work. It may be the case that if the test continuum presented the critical fricative in a novel position (i.e., word-initial), then differences in listening strategy during exposure may be observed with respect to generalization of learning to a novel word position (in addition to novel lexical items).

In the current work, a perception-focused listening strategy was induced in two ways: (1) by explicitly alerting listeners to the talker's ambiguous speech, as in McAuliffe and Babel (2016), and (2) through completion of a task during exposure (i.e., loudness decision) that could be performed without a linguistic analysis of the signal. These are but two of many different tasks that could promote attention to surface variation, and thus induce a perception-oriented listening strategy, and future work is needed to determine whether learning for other perception-oriented listening strategies will show the same patterns as the current results. Of note, the loudness judgment task represents a global listening strategy that could be performed equivalently for all items during exposure. That is, this task was designed to shift attention to amplitude during exposure as a gestalt, and not to differentially affect any particular stimulus. An alternative approach consistent with a perception-focused listening

strategy would be a task that requires explicitly processing surface detail for the critical phoneme, such a goodness judgment task of the critical fricatives during exposure. This type of task would link the perception-focused listening strategy to a linguistic context, in contrast to the non-linguistic context used presently, and may yield differential effects on perceptual learning as a consequence.

One challenge for future work that examines the role of listening strategy on perceptual learning is to *a priori* operationalize tasks that induce perception- vs comprehension-focused listening strategies, and to be able to measure whether the intended listening strategy was in fact achieved. In experiment 1, there is some evidence of differential listening strategies during exposure to the extent that reaction time in the perception-oriented, loudness judgment task was faster compared to the two comprehension-oriented tasks. This difference in processing time is consistent with differential depth of processing. In experiment 2, there is no way to objectively confirm that a different listening strategy was used between listeners who were or were not explicitly aware of the ambiguity in that no differences in performance during exposure were observed. This latter point reflects a limitation of the current study. Additional investigations that use tasks for which the intended listening strategy can be confirmed will advance an understanding of how differential engagement with the exposure stimuli interacts with perceptual learning. Another important avenue for future research is to better specify the mechanisms by which listening strategy influences learning. In both previous work and the current work, it is presumed that a perception-oriented listening strategy shifts attention to surface variation (i.e., non-linguistic aspects of the stimuli) whereas a comprehension-focused listening strategy shifts attention to the semantic content of the stimulus. As elegantly outlined in McAuliffe and Babel (2016), such shifts in selective attention may serve to influence the degree to which top-down lexical feedback is used for perceptual learning or the degree to which error-driven learning signals may be localized to earlier levels of processing.

To conclude, future research is aimed at understanding how external factors beyond local lexical context influence the learning mechanism. Recall that in experiment 2, half of our experimental participants received an extra instruction to alert them of the ambiguity in the signal. We observed robust perceptual learning with this manipulation, which converges with McAuliffe and Babel's (2016) most perception-oriented condition (word-initial + explicit knowledge of ambiguity), but differs from one of their intermediate condition (word medial + explicit knowledge of ambiguity). The degree to which lexical context and knowledge about the speaker's production interact to affect the learning mechanism requires further research in order to explicate which factors receive the greatest weight in determining changes in phonetic category structure, which could shed light on the differences in learning outcomes using perception-oriented manipulations.

ACKNOWLEDGMENTS

This research was supported by a seed grant from the Connecticut Institute for the Brain and Cognitive Sciences and by NIH NIDCD grant R21DC016141 to R.M.T. The views expressed here reflect those of the authors and not the NIH or the NIDCD. We extend gratitude to Nicholas Monto for helpful comments on an earlier version of this manuscript. We also extend gratitude to Jacqueline Ose for assistance with data collection.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Soft. **67**(1), 1–48.
Bradlow, A. R., and Bent, T. (**2008**). "Perceptual adaptation to non-native speech," Cognition **106**(2), 707–729.
Bradlow, A. R., Nygaard, L. C., and Pisoni, D. B. (**1999**). "Effects of talker, rate, and amplitude variation on recognition memory for spoken words," Percept., Psychophys. **61**(2), 206–219.
Craik, F. I., and Tulving, E. (**1975**). "Depth of processing and the retention of words in episodic memory," J. Exp. Psychol. Gen. **104**(3), 268–294.
Drouin, J. R., Theodore, R. M., and Myers, E. B. (**2016**). "Lexically guided perceptual tuning of internal phonetic category structure," J. Acoust. Soc. Am. **140**(4), EL307–EL313.
Eisner, F., and McQueen, J. M. (**2005**). "The specificity of perceptual learning in speech processing," Percept., Psychophys. **67**(2), 224–238.
Eisner, F., and McQueen, J. M. (**2006**). "Perceptual learning in speech: Stability over time," J. Acoust. Soc. Am. **119**(4), 1950–1953.
Fox, R. A. (**1984**). "Effect of lexical status on phonetic categorization," J. Exp. Psychol., Human Percept. Perform. **10**(4), 526–540.
Ganong, W. F. (**1980**). "Phonetic categorization in auditory word perception," J. Exp. Psychol., Human Percept. Perform. **6**(1), 110–125.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**(5), 3099–3111.
Jesse, A., and McQueen, J. M. (**2011**). "Positional effects in the lexical retuning of speech Perception," Psychonom. Bull. Rev. **18**(5), 943–950.
Johnson, M. A., and Goldberg, A. E. (**2013**). "Evidence for automatic accessing of constructional meaning: Jabberwocky sentences prime associated verbs," Lang. Cogit. Proc. **28**(10), 1439–1452.
Kraljic, T., and Samuel, A. G. (**2005**). "Perceptual learning for speech: Is there a return to normal?," Cognit. Psychol. **51**(2), 141–178.
Kraljic, T., and Samuel, A. G. (**2006**). "Generalization in perceptual learning for speech," Psychonom. Bull. Rev. **13**(2), 262–268.
Kraljic, T., and Samuel, A. G. (**2007**). "Perceptual adjustments to multiple speakers," J. Mem. Lang. **56**(1), 1–15.
Kraljic, T., Samuel, A. G., and Brennan, S. E. (**2008**). "First impressions and last resorts: How listeners adjust to speaker variability," Psychol. Sci. **19**(4), 332–338.
Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (**2015**). "lmerTest: Tests in linear mixed effects models," R package version 2(0).
Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). "Training Japanese listeners to identify English /r/ and /l/ II: The role of phonetic environment and talker variability in learning new perceptual categories," J. Acoust. Soc. Am. **94**(3), 1242–1255.
McAuliffe, M., and Babel, M. (**2016**). "Stimulus-directed attention attenuates lexically-guided perceptual learning," J. Acoust. Soc. Am. **140**(3), 1727–1738.
McQueen, J. M., Norris, D., and Cutler, A. (**2006**). "The dynamic nature of speech perception," Lang. Speech **49**(1), 101–112.
Mitterer, H., Scharenborg, O., and McQueen, J. M. (**2013**). "Phonological abstraction without phonemes in speech perception," Cognition **129**(2), 356–361.
Myers, E. B., and Mesite, L. M. (**2014**). "Neural systems underlying perceptual adjustment to non- standard speech tokens," J. Mem. Lang. **76**, 80–93.
Newman, R. S., Clouse, S. A., and Burnham, J. L. (**2001**). "The perceptual consequences of within-talker variability in fricative production," J. Acoust. Soc. Am. **109**(3), 1181–1196.
Norris, D., McQueen, J. M., and Cutler, A. (**2003**). "Perceptual learning in speech," Cognit. Psychol. **47**(2), 204–238.
Pisoni, D. B. (**1993**). "Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning," Speech Commun. **13**(1), 109–125.
Pitt, M. A., and Samuel, A. G. (**2006**). "Word length and lexical activation: Longer is better," J. Exp. Psychol., Human Percept. Perform. **32**(5), 1120–1135.

Polka, L., and Werker, J. F. (**1994**). "Developmental changes in perception of nonnative vowel contrasts," J. Exp. Psychol., Human Percept. Perform. **20**(2), 421–435.
Reinisch, E., Weber, A., and Mitterer, H. (**2013**). "Listeners retune phoneme categories across languages," J. Exp. Psychol., Human Percept. Perform. **39**(1), 75–86.
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.
Scharenborg, O., Weber, A., and Janse, E. (**2015**). "The role of attentional abilities in lexically guided perceptual learning by older listeners," Atten., Percept., Psychophys. **77**(2), 493–507.
Sjerps, M. J., and McQueen, J. M. (**2010**). "The bounds on flexibility in speech perception," J. Exp. Psychol., Human Percept. Perform. **36**(1), 195–211.
Sommers, M. S., Nygaard, L. C., and Pisoni, D. B. (**1994**). "Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude," J. Acoust. Soc. Am. **96**(3), 1314–1324.
Summerfield, Q. (**1981**). "Articulatory rate and perceptual constancy in phonetic perception," J. Exp. Psychol., Human Percept. Perform. **7**(5), 1074–1095.
Theodore, R. M., Miller, J. L., and DeSteno, D. (**2009**). "Individual talker differences in voice-onset-time: Contextual influences," J. Acoust. Soc. Am. **125**(6), 3974–3982.
Theodore, R. M., Myers, E. B., and Lomibao, J. A. (**2015**). "Talker-specific influences on phonetic category structure," J. Acoust. Soc. Am. **138**(2), 1068–1078.
van Linden, S., and Vroomen, J. (**2007**). "Recalibration of phonetic categories by lipread speech versus lexical information," J. Exp. Psychol., Human Percept. Perform. **33**(6), 1483–1494.
Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A. (**1999**). "Training American listeners to perceive Mandarin tones," J. Acoust. Soc. Am. **106**(6), 3649–3658.
Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., and Amano, S. (**2007**). "Infant-directed speech supports phonetic category learning in English and Japanese," Cognition **103**(1), 147–162.
Werker, J. F., and Tees, R. C. (**1984**). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," Infant Behav. Dev. **7**(1), 49–63.
Werker, J. F., Yeung, H. H., and Yoshida, K. A. (**2012**). "How do infants become experts at native-speech perception?," Curr. Dir. Psychol. Sci. **21**(4), 221–226.
Xie, X., Theodore, R. M., and Myers, E. B. (**2017**). "More than a boundary shift: Perceptual adaptation to foreign-accented speech reshapes the internal structure of phonetic categories," J. Exp. Psychol. Human Percept. Perform. **43**(1), 206–217.
Zhang, X., and Samuel, A. G. (**2014**). "Perceptual learning of speech under optimal and adverse conditions," J. Exp. Psychol., Human Percept. Perform. **40**(1), 200–217.