Stimulus-directed attention attenuates lexically-guided perceptual learning

Michael McAuliffe, and Molly Babel

Citation: The Journal of the Acoustical Society of America **140**, 1727 (2016); doi: 10.1121/1.4962529
View online: https://doi.org/10.1121/1.4962529
View Table of Contents: https://asa.scitation.org/toc/jas/140/3
Published by the Acoustical Society of America

Stimulus-directed attention attenuates lexically-guided perceptual learning

Michael McAuliffe[1,a)] and Molly Babel[2]
[1]*Department of Linguistics, McGill University, Montreal, Quebec, Canada*
[2]*Department of Linguistics, University of British Columbia, Vancouver, British Columbia, Canada*

(Received 8 January 2016; revised 10 August 2016; accepted 12 August 2016; published online 15 September 2016)

Studies on perceptual learning are motivated by phonetic variation that listeners encounter across speakers, items, and context. In this study, the authors investigate what control the listener has over the perceptual learning of ambiguous /s/ pronunciations through inducing changes in their attentional set. Listeners' attention is manipulated during a lexical decision exposure task such that their attention is directed at the word-level for comprehension-oriented listening or toward the signal for perception-oriented listening. In a categorization task with novel words, listeners in the condition that maximally biased listeners toward comprehension-oriented attentional sets showed the most perceptual learning. Focus on higher levels of linguistic meaning facilitated generalization to new words. These results suggest that the way in which listeners attend to the speech stream affects how linguistic categories are updated, providing insight into the qualitative differences in perceptual learning between the psychophysics and language-focused literatures.
 [http://dx.doi.org/10.1121/1.4962529]

[TB] Pages: 1727–1738

I. INTRODUCTION

Listeners are faced with a large degree of phonetic variability when interacting with their fellow language users. Speakers differ in size, gender, and sociolect, and this contributes to speech sound categories overlapping in acoustic dimensions. Despite this variation, listeners can interpret disparate and variable productions as belonging to a single word type or sound category, a phenomenon referred to as perceptual constancy (Shankweiler *et al.*, 1977; Kuhl, 1979) or recognition equivalence (Sumner and Kataoka, 2013). One of the processes for achieving this constancy is perceptual learning, whereby perceivers update a perceptual category based on contextual factors.

Perceptual learning is a well-established phenomenon in the psychophysics literature. Training can improve a perceiver's ability to discriminate in many disparate modalities [e.g., visual acuity, somatosensory spatial resolution, weight estimation, and discrimination of hue and acoustic pitch (for a review, see Goldstone, 1998 and for historical context, see Gibson, 1963)]. In the psychophysics literature, perceptual learning is generally seen as an improvement in a perceiver's ability to judge the physical characteristics of objects in the world through training that assumes attention on the task, but does not require reinforcement, correction, or reward. Here we focus on the kind of lexically-guided perceptual learning in speech perception that relates to the updating of a listener's sound categories based on exposure to a speaker's modified production of a particular category (Norris *et al.*, 2003; Vroomen *et al.*, 2007). For example, Kraljic *et al.* (2008b) showed that exposure to a speaker exhibiting /s/ productions which had been modified to sound more /ʃ/-like caused listeners to update their perceptual /s/ category to include more /ʃ/-like instances. This expanded or shifted /s/ category results in a greater willingness of the participant to categorize ambiguous /s/-/ʃ/ instances as /s/ rather than /ʃ/. Shifts in categorization functions have been demonstrated to be changes in phonetic representations and not merely shifts in post-perceptual decision biases (Clarke-Davidson *et al.*, 2008). Here, we use the term PERCEPTUAL LEARNING to refer to a shift in categorization function to novel items presented after exposure to lexical items with ambiguous pronunciations; the generalization of what was learned in the exposure phase to the novel items provides evidence that phonetic representations have been retuned. We reserve the term ADAPTATION for changes in listener performance to ambiguous items through the course of the exposure phase. Perceptual learning effects are typically evaluated as the difference between the normal categorization function and the one following exposure to a modification, although priming (Witteman *et al.*, 2013) and lexical endorsement (Maye *et al.*, 2008) have also been used to illustrate these perceptual adjustments.

While perceptual learning and adaptation in speech depends on perceivers being sensitive to the perceptual details that are to be learned, it also hinges on higher-level linguistic knowledge. Thus, lexically-guided perceptual learning is considered by many as evidence for interactive processes in speech perception (although see Norris *et al.*, 2000). TRACE is an interactive model of speech perception (McClelland and Elman, 1986; McClelland *et al.*, 2006) that provides a mechanism for lexically-guided perceptual learning. In an interactive model like TRACE, activated or predicted lexical information reaches down and activates sub-lexical phoneme representations. Retuning or perceptual learning takes place in the mapping of the auditory input to

a)Electronic mail: michael.mcauliffe@mail.mcgill.ca

the sub-lexical representations, which guide the parsing of the signal. These mechanisms rely on expectations and activation of higher-level knowledge.

While perceptual learning is a response to speaker variation and considerable literature has examined speaker characteristics in perceptual learning (e.g., Kraljic *et al.*, 2008a; Kraljic *et al.*, 2008b; Witteman *et al.*, 2013), there is also variation on the part of the listener. Listeners can introduce variability into perceptual learning via their adopted attentional sets, of which two broad types have been posited (Cutler *et al.*, 1987; Pitt and Szostak, 2012). The first is a COMPREHENSION-ORIENTED or diffuse attentional set—this is the attentional set assumed to operate during normal language use. When oriented toward comprehension, listeners are focused on understanding the intended message of the speech, and a comprehension-oriented set is promoted by tasks that focus on word identity and word recognition. The comprehension-oriented attentional set is elicited in lexically-guided perceptual learning paradigms through the use of lexical decision tasks and the embedding of modified sound categories in word tokens. There is already some correlational evidence that comprehension-oriented attention facilitates lexically-guided perceptual learning. Listeners' endorsement rates of the ambiguously pronounced critical items as words in a lexical decision exposure phase (i.e., responding "word" to an item like *castle* produced with an ambiguous fricative /kæ?sl/) correlates with perceptual learning shifts in categorization tasks (Scharenborg and Janse, 2013), which suggests that attending to the critical items at a lexical level enhances perceptual learning.

A second kind of attentional set is a PERCEPTION-ORIENTED attentional set, where a listener is focused more on the low-level signal properties of the speech rather than the message. The perception-oriented attentional set is promoted by tasks such as phoneme/syllable monitoring, mispronunciation detection, or gender detection, and can even be elicited through task monotony (Cutler *et al.*, 1987; Norris and Cutler, 1988; Pitt and Samuel, 1990). The tasks used in visually-guided perceptual learning and perceptual learning within the psychophysics literature can be thought of as eliciting this attentional set, due to their focus on stimuli that are devoid of linguistic meaning or relevance outside of the experiment.

While task difficulty or task instructions may elicit different kinds of listening styles (McLennan and Luce, 2005; Theodore *et al.*, 2015), individuals may habitually differ in their use of attentional sets and there is evidence in the literature that individual's attentional sets may impact perceptual learning. Older adults with poorer attention-switching abilities showed more perceptual learning than older adults with better attention-switching abilities (Scharenborg *et al.*, 2015). Scharenborg and colleagues reason that listeners with poorer attention-switching skills may rely more on lexical content, which facilitates the process of lexically-guided perceptual learning. Those with better attention-switching abilities may attend more to the phonetic details, noticing a less than perfect match between the perceived signal and a lexical representation, thus perceptually learning less.

Attentional set effects have been argued to be evidence for autonomous models of perception (MERGE, Norris *et al.*, 2000). However, Mirman *et al.* (2008) demonstrate that TRACE can account for listeners' focus on perception-oriented attentional sets through the implementation of a negative bias parameter that reduces lexical activation, thus attenuating lexical effects on phoneme-level. An assumption of this approach is that listeners are generally engaged in comprehension-oriented listening and benefit from maximal lexical activation (Mirman *et al.*, 2008). While listeners are not presented with nonword stimuli in day-to-day interaction, there are occasions where a listener's ear is turned toward more perception-oriented listening. An unfamiliar accent, a non-native language, or an odd (mis)pronunciation (Pitt and Szostak, 2012) can naturally shift listeners to be more signal-oriented. Word structure plays a role here as well, as listeners are less inclined to have fully activated a lexical prediction at the onset of a word. Lexical information exhibits less of an effect on word-initial positions as compared to later positions (Pitt and Samuel, 2006) and lexical bias effects and phoneme restoration patterns are stronger when the ambiguous sound is in word-medial position than in word-initial position (Pitt and Samuel, 2006; Samuel, 1981). Thus, it is not surprising that one study that tested for perceptual learning with exposure to ambiguous pronunciations in word-initial positions did not find evidence for perceptual learning (Jesse and McQueen, 2011).

In this paper we seek to connect these different levels of perceptual attention to the degree of generalization in the perceptual learning task. The idea we introduce is that activation of lexical information facilitates generalization across the lexicon, whereas more focused perception-oriented attention will show smaller degrees of generalization, as the locus of what is learned will not benefit from the same strength of lexical prediction. These predictions align with TRACE and also follow the predictive coding model, a hierarchical generative Bayesian framework for perception (Clark, 2013). This framework uses a hierarchical generative model that aims to minimize prediction error between bottom-up sensory inputs and top-down expectations. Mismatches between the top-down expectations and the bottom-up signals generate error signals that are used to modify future expectations. Perceptual learning then is the result of modifying expectations to match learned input and reduce future error signals. Within TRACE-like terms, boosts in lexical activation initiate phoneme-level units that predict a sensory experience, the feature-level in TRACE. Experiencing a noncanonical /s/ sound generates an error signal which feeds back up to the lexical level to modify the distribution for future experience with this item. The feedback from the error signal updates distributions associated with lexical and phoneme-level representations. Attenuating lexical activation by directing focused perception-oriented attention decreases the lexical activation and predictions, and retuned sensory-to-phoneme connections do not propagate as efficiently up to the lexical level, thus reducing the magnitude of generalization.

To this end we use a lexically-guided perceptual learning paradigm to test the role of attention type—perception- or comprehension-oriented attentional sets—on perceptual

TABLE I. The two word positions (Word-medial and Word-initial) and the Attention manipulation (direction to the speaker's /s/ and no explicit instructions) combine to create four experimental conditions.

		Attention	
		"Speaker has an ambiguous 's' "	No explicit attention instructions
Word position	Word-medial, e.g., *castle*	Word-medial/ Attention *Prediction*: less perceptual learning	Word-medial/ No Attention *Prediction*: most perceptual learning
	Word-initial, e.g., *silver*	Word-initial/ Attention *Prediction*: less perceptual learning	Word-initial/ No Attention *Prediction*: less perceptual learning

learning. Listeners are exposed to ambiguous productions of words containing a single instance of /s/, where the /s/ has been modified to sound more /ʃ/-like. Exposure comes in the guise of a lexical decision task. There are four conditions. In one group, the critical words have an /s/ in word-initial position (*cement*, /səmɛnt/), with no /ʃ/ neighbor (**shement*, /ʃəmɛnt/); this is referred to as the WORD-INITIAL condition. In the other group, the critical words have an /s/ in word-medial position (*tassel*, /tæsl/) with no /ʃ/ neighbor (**tashel*, /tæʃl/); this is referred to as the WORD-MEDIAL condition. In addition, half of each group is given instructions that the speaker has an ambiguous /s/ and to listen carefully, following Pitt and Szostak (2012). Table I presents the four experimental conditions to which listeners are assigned.

We predict that directing listeners' attention to phonetic ambiguity with the Attention instructions and with the ambiguous fricative in the more salient onset position will engage more perception-oriented attentional sets. Specifically, we predict that the adoption of more perception-oriented attentional sets will result in lower word endorsement rates (Pitt and Szostak, 2012) and faster response times in the lexical decision task. Given the suggested reliance of perceptual learning on lexical scaffolding, this lower acceptance rate should lead to a smaller perceptual learning effect for listeners in perception-oriented tasks (e.g., groups with Word-initial stimuli and Attention instructions) as compared to the listeners in the more comprehension oriented attentional set group (e.g., the group with Word-medial with No Attention instructions). These predictions are supported by previous work indicating lower word endorsement rates to items with ambiguous fricatives in onset (Pitt and Szostak, 2012) and evidence that perceptual learning is inhibited by word-initial ambiguity (Jesse and McQueen, 2011). Directing listeners' attention to ambiguity in the stimuli either through explicit instruction or word position is predicted to amount to equivalent levels of attenuated perceptual learning; we offer no predictions about differences in performance between these two groups. In contrast, listeners exposed to Word-medial ambiguity with no attention instructions are predicted to show the most perceptual learning, as they are hypothesized to show higher word endorsement rates with slower response times.

While many previous lexically-guided perceptual learning studies compared conditions with a manipulated /s/ to a manipulated /ʃ/, we chose to only manipulate /s/ and compare performance to a control group who had not been exposed to the speaker previously. The rationale for this decision was based on evidence of some unexplained asymmetries in perceptual learning such that the learning is more robust for /s/ (Zhang and Samuel, 2014). Thus, to focus our inquiry on attentional sets in perceptual learning, we compare perceptual learning only for /s/ to a control group.

II. METHODS

A. Participants

A total of 124 native speakers of English from the UBC population completed the experiment and were compensated with either $10 CAD or course credit. The data from two individuals who reported speech or hearing disorders were excluded from the analyses. This left data from 122 participants for analysis. Twenty additional native English speakers participated in a pretest to determine the most ambiguous-sounding tokens.

B. Materials

English words (128) and 100 phonologically-legal nonwords were used as exposure materials. The set of words consisted of 40 critical items, 20 control items, and 60 filler words. These items are listed in Appendix A. Half of the critical items had an /s/ as the onset of the first syllable (WORD-INITIAL) and half had an /s/ as the onset of the final syllable (WORD-MEDIAL). All critical tokens formed nonwords if their /s/ was replaced with /ʃ/. Half the control items had an /ʃ/ as the onset of the first syllable and half had an /ʃ/ as the onset of the final syllable. Each critical item and control item contained just the one sibilant, with no other /s z ʃ ʒ tʃ dʒ/. Filler words and nonwords did not contain any sibilants.

Four monosyllabic minimal pairs were selected as test items for categorization. These minimal pairs differed only in the voiceless sibilant at the beginning of the word (*sack-shack*, *sigh-shy*, *sin-shin*, and *sock-shock*). Two of the pairs had a higher log frequency per million words from SUBTLEXus (Brysbaert and New, 2009) for the /s/ word (*sack* = 1.11, *shack* = 0.75; *sin* = 1.2, *shin* = 0.48) and two had higher frequencies for the /ʃ/ word (*sigh* = 0.53, *shy* = 1.26; *sock* = 0.95, *shock* = 1.46).

All words and nonwords were recorded by a male Vancouver English speaker in a quiet room. Critical words for the lexical decision exposure phase were recorded in pairs, once normally and once with the sibilant swapped forming a nonword (see Appendix B). The speaker was

instructed to produce both forms with comparable speech rate, speech style, and prosody.

The critical /s/ words and the categorization test items required digital manipulation. For each critical item, the word (e.g., *castle* /kæsl/) and associated nonword (/kæʃl/) versions were morphed together in an 11-step continuum (0%–100% of the nonword /ʃ/ recording, in steps of 10%) using STRAIGHT (Kawahara *et al.*, 2008). A similar process was done for the minimal pair categorization items (e.g., *sack-shack*). Prior to morphing, the word and nonword versions (word and word versions in the case of the categorization items) were time aligned based on acoustic landmarks, such as stop bursts, onset of $F2$, nasalization or frication, etc. All control items and filler words were processed and resynthesized by STRAIGHT to ensure a consistent quality across items.

To determine which step of each continuum for the critical /s/ words and the categorization items was most ambiguous, an initial pretest was conducted on a group of participants ($n = 20$) who did not complete the main experiment. These participants were presented with each of the 11 steps of the word-nonword continua (e.g., morphings of /kæsl/ to /kæʃl/) and each categorization minimal pair continuum (e.g., /sæk/ to /ʃæk/), resulting in 484 trials (40 critical /s/ words plus four minimal pairs by 11 steps). These items were separated into two blocks. Participants completed a lexical decision task for the word–nonword critical item continua, responding with either word or nonword to each step. For the word-word categorization continua, participants identified the first sound as either "s" or "sh."

The proportion of word responses for the critical items and /s/-responses for categorization items at each step of each continuum was calculated and the most ambiguous step was chosen for each item. The threshold for selecting the most ambiguous step was when the percentage of /s/-response dropped near 50%. Due to experimenter error, the continuum for *seedling* was not included in the stimuli list for this pre-test, so the chosen step for *seedling* was the average chosen step for the /s/-initial words. The average step chosen for Word-initial /s/ words was 6.8 [standard deviation (SD) = 0.5], and for Word-medial /s/ words the average step was 7.7 (SD = 0.8). For the categorization minimal pairs, six steps surrounding the 50% cross-over point were selected for use in the phonetic categorization task.

C. Procedure

The experiment consisted of an exposure phase, where participants completed a lexical decision task, and a categorization test, where participants categorized minimal pair continua. Twenty-five participants were assigned to a control group and only completed the categorization task.

Participants in the experimental conditions were assigned to one of four groups from a 2×2 between-subject factorial design for the exposure lexical decision phase, as outlined in Table I. The first factor was the position of the ambiguous sibilant in the exposure words (Word Position: WORD-INITIAL versus WORD-MEDIAL) and the second factor was whether participants were given additional instructions about the sibilant (Attention: ATTENTION versus NO ATTENTION). Thus, the four experimental conditions were Word-initial/Attention ($n = 24$), Word-initial/No Attention ($n = 25$), Word-medial/Attention ($n = 25$), and Word-medial/No Attention ($n = 25$).

Listeners in Word-initial /s/ words conditions were presented with 20 critical word-initial ambiguous /s/ words, the 20 control /ʃ/ words, 60 filler words, and 100 filler nonwords. Listeners in the Word-medial condition were presented with 20 critical word-medial ambiguous /s/ words, 20 control /ʃ/ words, 60 filler words, and 100 filler nonwords. Thus, all exposure phases had a consistent 200 trials with identical control and filler items for all participants. Participants in the Attention conditions received additional instructions. Specifically, they were told "this speaker's 's' sound is sometimes ambiguous" and instructed to "listen carefully so as to choose the correct response."

The instructions for all participants in the lexical decision exposure phase were to respond with either word if they thought what they heard was a word or "nonword" if they did not think it was a word. The buttons corresponding to word and nonword were counterbalanced across participants. Trial order was pseudorandom, with no critical or control items appearing in the first six trials and no critical or control trials in a row (following Reinisch *et al.*, 2013). For each trial, a blank screen was shown for 500 ms, and then the two responses and their corresponding buttons on the button box were shown (i.e., word and response button 1 were associated on one side of the screen and nonword and response button 5 on the other side of the screen). The auditory stimulus was played 500 ms following the presentation of the response options. Participants had 3000 ms from the onset of the auditory stimulus to respond. Feedback about whether a response was detected was given once a button was pressed or the 3000 ms had elapsed and was shown for 500 ms before the following trial began. Every 50 trials participants were given a break and the next trial did not start until the participant pressed a button.

All participants in the experimental conditions and those in the control condition ($n = 25$) completed a categorization task. Participants heard an auditory stimulus and had to categorize it as one of two words, differing only in the onset sibilant, e.g., *sin* or *shin*. Buttons corresponding to the words were counterbalanced across participants. The six most ambiguous steps of the minimal pair continua as determined by the pre-test were used with 7 repetitions each, giving a total of 168 trials.

Participants were given oral instructions explaining both tasks at the beginning of the experiment to remove experimenter interaction and avoid the potential use of additional /s/ and /ʃ/ sounds between exposure and categorization, as there is some reported evidence of cross-speaker generalization for sibilants (Kraljic and Samuel, 2005). Written instructions were presented to participants at the beginning of each task as well.

III. ANALYSIS AND RESULTS

We analyze adaptation to the ambiguous pronunciations in the exposure phase and generalization of perceptual learning in the categorization phase. Prior to analysis, the 50%

cross-over points for all subjects were calculated according to the methods described in Kleber *et al.* (2012). Two subjects in the Word-medial/No Attention (leaving a total $n = 23$ analyzed) were removed because their 50% cross-over point fell outside of the range of steps presented, as they categorized almost all of the continua steps as /s/ rather than /ʃ/.

A. Exposure

1. Word recognition accuracy

Performance on the exposure task was high overall: 92% of the filler words were correctly accepted and 89% of nonwords were correctly rejected. Trials with nonword stimuli and responses faster than 200 ms or slower than 2500 ms from the onset of the trial were excluded from further analysis. Participants were expected to show adaptation effects across exposure, such that accuracy (endorsement as a word) would improve over time for words with the modified /s/ category. Unmodified words should show no such adaptation effects. Trial Number is included in the model as a continuous variable to examine adaptation across the exposure task.

A logistic mixed-effects model with accuracy as the dependent variable was fit with fixed effects[1] for Trial Number (0–200), Trial Type (Filler, /s/, and /ʃ/), Attention (No Attention and Attention), Word Position (Word-Initial and Word-Medial), and all possible interactions. The random effect structure was as maximally specified as possible with random intercepts for Subject and Word. Random slopes by Subject for Trial Number, Trial Type, and their interactions were coded as well. The full statistical model is presented in Appendix C.

A significant fixed effect was found for Trial Type of /s/ versus Filler [$B = -2.01$, standard error (SE) $= 0.31$, $z = -6.56$, $p < 0.01$], as participants were less likely to endorse words containing the modified /s/ category as words compared to filler words. However, there was a significant interaction between Trial Number and Trial Type of /s/ versus Filler ($B = 0.41$, SE $= 0.12$, $z = 3.36$, $p < 0.01$), indicating that participants adapted to the speaker's /s/ and endorsed more of these items as words over the course of the experiment. Trial Type of /ʃ/ did not differ significantly from Filler, and did not interact with Trial Number. Figure 1 shows within-subject mean accuracy across exposure, with Trial Number presented in four blocks. A clear learning effect across the experiment can be seen with the /s/ items, with a greater likelihood of word responses to these items later in the course of the experiment.

2. Response time

Response time was collected from the onset of each item. To normalize for the varying durations of the exposure stimuli, each item's duration was subtracted from the response time for each associated trial. A linear mixed-effects model with this normalized response time as the dependent variable was fit with an identical fixed effect and random effect structure as the logistic model for accuracy. The full statistical model is presented in Appendix C. Significant effects were found for Trial Type of /s/ versus Filler ($B = 0.37$, SE $= 0.09$, $t = 4.12$, $p < 0.01$), the interaction between Trial Number and Trial Type of /s/ versus Filler ($B = -0.08$, SE $= 0.027$, t $= -3.05$, $p < 0.01$), and the interaction between Trial Type of /s/ versus Filler and Word Position ($B = -0.38$, SE $= 0.15$, $t = -2.61$, $p < 0.01$). Trial Type of /ʃ/ was not significantly different than Filler, and did not significantly interact with Trial Number or Word Position, unlike Trial Type of /s/. These effects generally follow the pattern found in the accuracy model. Participants begin with slower response times to words with a modified /s/, but over time the difference between these words and filler words lessens. This declination in response times is larger for Word-medial exposure items. Figure 2 shows within-subject mean response time across exposure, with Trial again in four blocks, separated by Word Position. Listeners are slower to respond word to items with ambiguous /s/ in

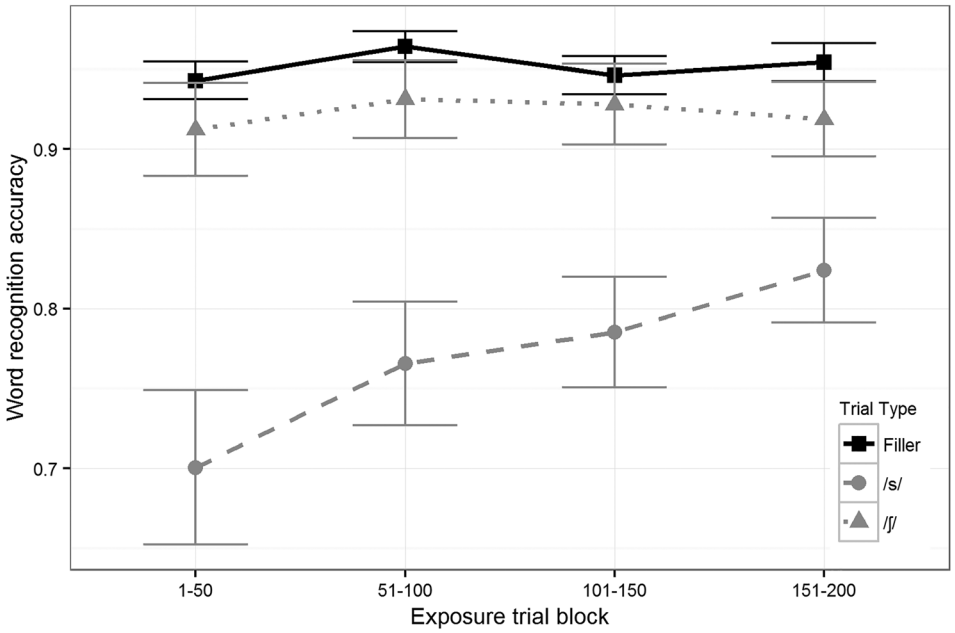

FIG. 1. Within-subject mean accuracy for words in the exposure phase for filler words, /s/-words (with ambiguous /s/ realizations), and /ʃ/ words, collapsed across Attention and No Attention conditions. Error bars represent 95% confidence intervals. Trial Number is shown blocked by groups of 50 to allow for within-subject visualization.

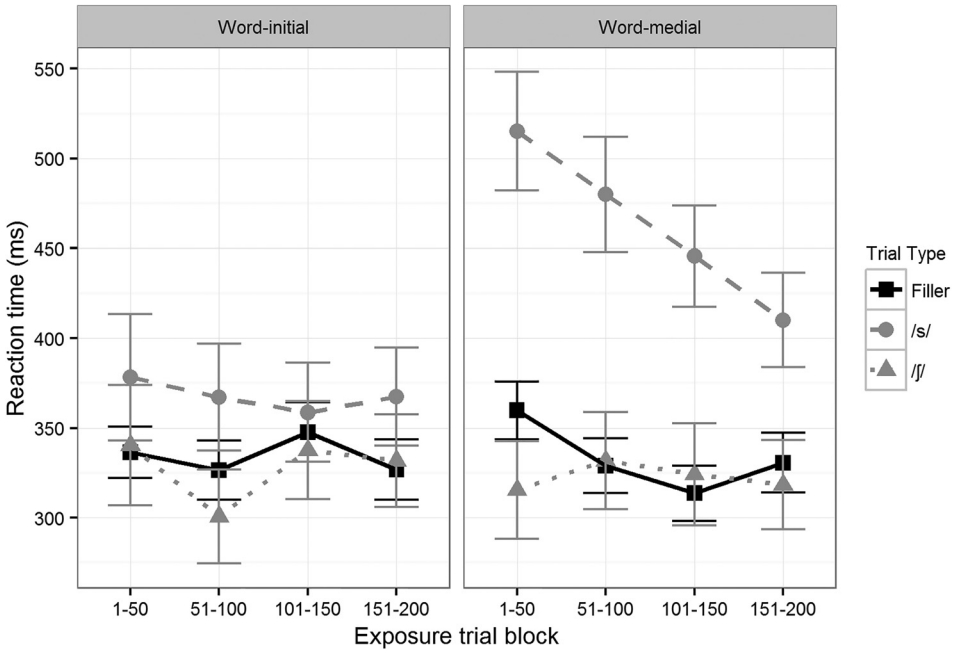

FIG. 2. Within-subject mean response time to words in the exposure phase for filler words, /s/-words (with ambiguous /s/ realizations), and /ʃ/ words by Word-initial or Word-medial exposure lists. Error bars represent 95% confidence intervals. Trial Number is shown blocked by groups of 50 to allow for within-subject visualization.

Word-medial position than those items in Word-initial position, and while response time decreases with exposure in Word-initial and Word-medial conditions, that decrease in response times is larger for the Word-medial listeners.

B. Categorization

Responses were coded as 1 for /s/ responses and 0 for /ʃ/ responses. Positive significant estimates therefore indicate a higher likelihood of /s/ response in categorization. Thus, positive significant effects are indicative of perceptual learning, as a higher likelihood of /s/ response is associated with an expanded /s/ category.

As in the analysis of the exposure phase, responses with response times less than 200 ms or greater than 2500 ms were excluded from analyses, following Reinisch *et al.* (2013).

1. Control

A logistic mixed-effects model was fit for the control group with Subject and Continuum as random effects[2] and Step[3] as a fixed effect with by-Subject and by-Continuum random slopes for Step. The intercept was not significant ($B = 0.43$, SE $= 0.29$, $z = 1.5$, $p = 0.13$), indicating that control participants did not differ significantly from the pretest participants who determined the most ambiguous steps. Step was significant ($B = -2.61$, SE $= 0.28$, $z = -9.1$, $p < 0.01$), with higher steps (more /ʃ/-like) responded to more as /ʃ/ words.

Results from the control experiment are included in the analyses with the experimental conditions below and are shown in Fig. 3 alongside the categorization results from the experimental conditions.

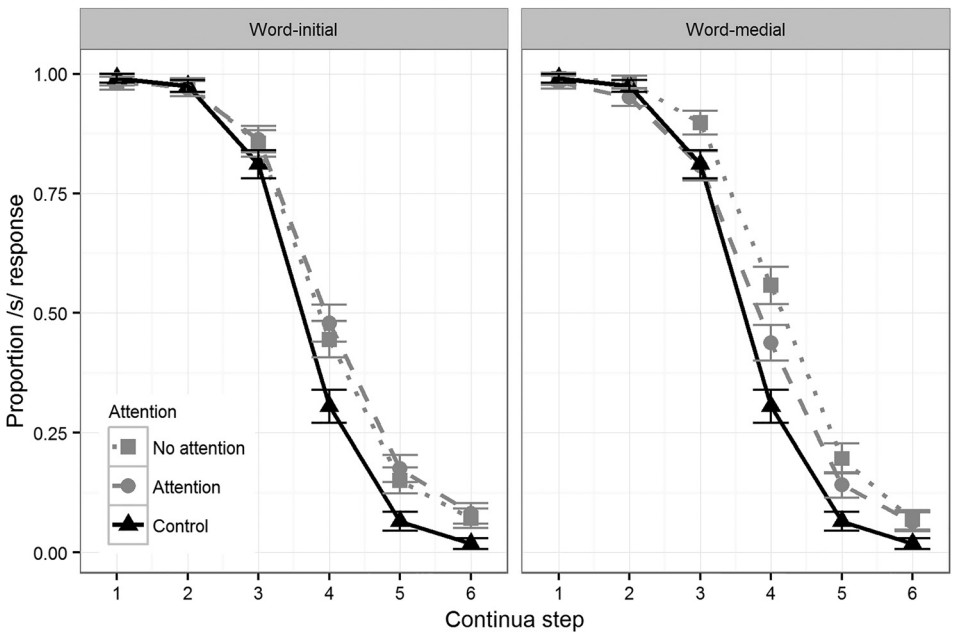

FIG. 3. Proportion /s/-word responses 6-step continua for Attention and No Attention conditions. The categorization function for control participants is included for visual comparison. Error bars represent 95% confidence intervals.

2. Experimental

We conduct two analyses of the categorization data from the experimental groups. In the first analysis we compare the four experimental groups. The second analysis compares the performance of the experimental groups to the control group to assess learning relative to baseline. The full details of both models are presented in Appendix C.

A logistic mixed-effects model was constructed with Subject and Continuum as random effects and a by-Subject random slope for Step and by-Continuum random slopes for Step, Attention, Word Position, and their interactions. Fixed effects for the model were Step, Word Position, Attention, and their interactions.

There was a significant effect for the intercept ($B = 1.03$, SE $= 0.25$, $z = 4.05$, $p < 0.01$), indicating that participants categorized more of the continua as /s/ in general.

This significant positive intercept is basic evidence of perceptual learning. There was also a significant main effect of Step ($B = -2.14$, SE $= 0.16$, $z = -13.63$, $p < 0.01$), and a significant interaction between Word Position and Attention ($B = -0.99$, SE $= 0.46$, $z = -2.17$, $p = 0.03$). These results are visualized in Fig. 3. When exposed to a modified /s/ category at the beginning of words, participants show a general expansion of the /s/ category with no difference in behavior induced by the attention manipulation. However, when listeners are exposed to ambiguous /s/ tokens in word medial positions, we see differences in behavior beyond the general /s/ category expansion. Participants not warned of the speaker's ambiguous tokens categorized more of the continua as /s/ compared to those who were warned of the speaker's ambiguous /s/ productions. Diverting listeners' resources away from the lexical level either by overtly noting ambiguous /s/ productions or by the initial position of /s/ in the word inhibits generalization of lexically-guided perceptual learning. Note that despite these continuum steps being the six most ambiguous steps from the original 11 step continua, listeners categorized the endpoints nearly categorically as /s/ and /ʃ/.

A second logistic mixed-effects model was created with participants from both the experimental conditions and the control experiment. A new five-level factor was created (Control and each of the four combinations of the levels of Word Position and Attention), with Control as the reference level. Using Control as the reference level allows for comparison of each experimental group to the behavior of participants' who completed the same categorization task without previous exposure to the speaker's modified /s/ category. Two experimental groups showed significant perceptual learning effects as compared to the Control group: the Word-initial/Attention participants ($B = 0.63$, SE $= 0.31$, $z = 2.05$, $p = 0.04$) and the Word-medial/No Attention participants ($B = 1.05$, SE $= 0.31$, $z = 3.37$, $p < 0.01$). The other two experimental conditions were not significantly different than the Control group (Word-initial/No Attention: $B = 0.42$, SE $= 0.30$, $z = 1.34$, $p = 0.16$; Word-medial/Attention: $B = 0.25$, SE $= 0.30$, $z = 0.84$, $p = 0.40$), but the effects were in the positive direction across all groups. The results of this model complement those of the previous model, showing that while there was perceptual learning overall on the pretested continua, only two conditions were outside of the range of variation present in the Control participants. It is important to note that the non-significance of the other two conditions does not indicate a complete lack of perceptual learning, but rather that the size of the effect requires more statistical power to detect a difference than the current experiment offers. Furthermore, the participants in the Word-Medial/No Attention condition showed a larger difference from Control participants than any other condition.

IV. DISCUSSION

The numerically largest perceptual learning effect was found in the condition that was most biased toward a comprehension-oriented attentional set: Listeners exposed to the modified /s/ category in the middle of words where they have the strongest lexical bias and with no explicit instructions about /s/ had larger perceptual learning effects than the other conditions. Directing attention to the signal with instructions or initial word position showed roughly equivalent sizes of perceptual learning in our first analysis, suggesting that there was not a compounding effect of explicit attention and word position. That is, the comprehension-oriented nature of the primary task still exerts an effect on attentional set selection, and a significant perceptual learning effect was found on novel words.

These findings partially replicate the lack of perceptual learning for fricatives in an initial position in Jesse and McQueen (2011). The group most similar to the Word-initial experiment in Jesse and McQueen was found to not be significantly different from control participants. However, the statistical trend found here is stronger than that found in Jesse and McQueen, perhaps a result of the two categories used (/s/-/f/ versus /s/-/ʃ/). The two groups that were found to behave significantly different from control participants were the ones that maximized attention (through both instructions and initial word position) and minimized attention. We can speculate that this horseshoe effect of attention on perceptual learning suggests that increasing attention can increase perceptual adaptation of positionally salient items, but it will still not approach the adaptation to not overtly attended items. The condition that maximized attention was coincidentally the most similar to the test items (word-initial sibilants), and exposure specificity has been shown previously to play a large role in perceptual learning (Kraljic and Samuel, 2005; Eisner and McQueen, 2005; and others). Increasing attention may also increase exposure specificity, which is a prediction that falls out of our use of the predictive coding framework (Clark, 2013).

In addition to the learning effects of the categorization phase, listeners also demonstrated adaptation over the course of exposure in the lexical decision task. In the initial trials, words with a modified /s/ were responded to more slowly and less accurately (in that they were less likely to be endorsed as words), but over the course of exposure to these items, both response times and accuracy approach those of filler and unmodified /ʃ/ words. Listeners exposed to items with word-initial ambiguity and thus engaged in more perception-oriented attentional sets were faster at responding

to critical items as words than those exposed to ambiguous fricatives in word-medial position. Response times to word-medial ambiguous items decreased through the course of the experiment, indicating that identifying these items as words became easier, but they were always substantially longer than listeners' responses to /ʃ/ and filler items.

Precisely why this response time profile is only present for word-medial /s/ is unclear. Indexical and item-specific effects are typically observable in retrieval tasks when processing is slow and effortful (McLennan and Luce, 2005), although it has been recently argued that specificity effects are related to attention differences in encoding (Theodore *et al.*, 2015), which may be similar to the perceptual consequences of comprehension- and perception-oriented attentional sets. Given that perceptual learning of fricatives is generally speaker-specific often showing little cross-talker generalization, we could imagine that perceptual learning necessarily involves this more slow and effortful indexical processing. The longer response times to words with word-medial /s/ may be linked to the greater perceptual learning effect found in categorization for participants not told of the ambiguous /s/. However, it may be that the processing is identical across words with medial /s/ and initial /s/. In this case, the slower indexical processing would be hidden by the time course of the primary lexical decision task for word-initial /s/, because that process is started sooner (at the onset of the stimuli). The current study was not explicitly designed to test hypotheses related to response time, as there was no control or manipulation of the duration of stimuli (the average number of syllables were controlled so that they did not differ across stimulus type). Any interpretation of the difference in response time profiles is naturally speculative and should be explored further in future work.

Contrary to our initial predictions, the position of /s/ within the word did not significantly influence accuracy in the exposure task, attention instructions did not affect the word endorsement rates of the ambiguous /s/ items, and there were no effects of attention instructions on response time. Despite this lack of findings in the exposure task, differences emerge in the categorization phase which tests generalization in perceptual learning to unheard items. The lack of synced performance in the exposure and test phases not only runs counter to our predictions, but runs counter to the findings of Scharenborg and colleagues, who demonstrate a positive correlation between rates of lexical endorsement and perceptual learning in categorization. It may be the case that we have put too much stock in the overt decisions logged by participants in a lexical decision task. Lexical decision tasks offer the opportunity for post-perceptual decisions and meta-linguistic judgment, e.g., a participant may recognize /kæʔsl/ as a potential pronunciation of *castle* but reject it as a word because it is not the canonical or correct pronunciation. While lexically-guided perceptual learning certainly relies on lexical activation, the assumption that word endorsement in a lexical decision task is the ultimate evidence of lexical activation may be too strong. Better evidence for lexical activation comes from priming experiments. In terms of ambiguity, Connine *et al.* (1993), for example, demonstrate that a one feature switch (e.g., *formal* /foɹml/ to /voɹml/) results in successful cross-modal priming. A multi-feature switch (e.g., *formal* /foɹml/ to /ɡoɹml/) does not show priming, indicating that phonetic similarity is necessary for sufficient lexical activation. The ambiguity introduced in our manipulations is more in line with one feature switches, with the /s/ becoming more [+ anterior] in its acquisition of acoustic features more similar to /ʃ/. While perceptual learning has been argued to be truly demonstrative of perceptual changes (Clarke-Davidson *et al.*, 2008), it is very plausible that *some* aspects or features of perceptual learning are indicative of shifts in criterion or response bias. Delimiting what is perceptual and what is post-perceptual in adaptation and perceptual learning will be crucial in subsequent research in this area.

The adaptation observed in the exposure phase complements the perceptual learning probed in the categorization phase and illustrate different facets of perceptual retuning processes. Listeners became more accepting of ambiguous /s/ productions throughout the exposure phase as they endorsed more of these items as words, thereby demonstrating adaptation regardless of word position. However, it seems that the updated knowledge about /s/ distributions generalizes more readily when the noncanonical /s/ surfaces in medial position where it is supported by additional lexical activation.

Attention to the ambiguous fricative equalized the perceptual learning effects across word positions, showing attenuated levels in these cases. Understanding generalization in perceptual learning in light of a listener's perceptual focus may provide an explanation as to why lexically-guided perceptual learning generalizes more readily, while repetitive (Idemaru and Holt, 2014) or visually-guided approaches (Reinisch *et al.*, 2014) do not. Idemaru and Holt (2014), for example, do not show generalization in perceptual learning when listeners are presented with a task that exclusively presents /b/ and /p/ initial words (e.g., *beer*/*pier*) and includes /d/ and /t/ initial words (e.g., *dear*/*tear*) at test. Audio-visual presentation of VCV sequences for perceptual learning in Reinisch and colleagues' work also failed to show generalization to untrained items. It may be the case that these modes of stimuli presentation are too monotonous and direct perceivers to the signal in ways that engage in perception-oriented listening (Cutler *et al.*, 1987), inhibiting generalization. Perceptual learning in the psychophysics literature has shown a large degree of exposure-specificity, where observers show learning effects only on the same or very similar stimuli as those they were trained on. As such, perceptual learning has been argued to reside in or affect the early sensory pathways, where stimuli are represented with the greatest detail (Gilbert *et al.*, 2001). Visually-guided perceptual learning has also shown a large degree of exposure-specificity, where participants do not generalize cues across contexts (Reinisch *et al.*, 2014), while lexically-guided perceptual learning does (Norris *et al.*, 2003; Kraljic and Samuel, 2005; Kraljic and Samuel, 2007). Crucially, lexically-guided perceptual learning in speech has shown a greater degree of generalization than would be expected from a purely psychophysical standpoint. The testing stimuli

are in many ways quite different from the exposure stimuli, with participants typically trained on multisyllabic words ending in an ambiguous sound and then tested on monosyllabic words (Reinisch *et al.*, 2013) and nonwords (Norris *et al.*, 2003; Kraljic and Samuel, 2005). However, exposure-specificity has been found when exposure and testing use different positional allophones (Mitterer *et al.*, 2013).

Why is lexically-guided perceptual learning more context-general? We provide evidence that this context-generality may be related to a listener's attentional set, which can be influenced by linguistic and task-instruction properties. A comprehension-oriented attentional set, where a listener's goal is to understand the meaning of speech, fosters linguistic predictions, promoting generalization and leading to greater perceptual learning. A purely perception-oriented attentional set, where a listener's goal is to perceive specific qualities of a signal, does not promote generalization. A lexically-guided perceptual learning paradigm uses tasks oriented toward comprehension, so generalization is to be expected in general, but the more perception-oriented the attentional set, then less perceptual learning should be observed [see also Scharenborg and Janse (2013) and Scharenborg *et al.* (2015)].

This finding can be easily incorporated into existing models like TRACE. The noncanonical /s/ productions heard in the exposure phase modify future predictions at the lexical and phonemic levels. The activation of lexical information facilitates generalization across the lexicon, whereas more focused perception-oriented attention shows smaller degrees of generalization, as the locus of what is learned does not benefit from the same strength of interactive lexical prediction.

Recently, proposed frameworks intended to provide mechanisms for perceptual learning have begun to incorporate Bayesian reasoning (Clark, 2013; Kleinschmidt and Jaeger, 2015; Norris and McQueen, 2008). One such framework is the predictive coding model from Clark (2013), which is a domain-general hierarchical generative Bayesian framework for perception. This model aims to minimize prediction error between bottom-up sensory inputs and top-down expectations. Mismatches between the top-down expectations and the bottom-up signals generate error signals that are used to modify future expectations. Perceptual learning is the result of modifying expectations to match learned input and reduce future error signals. In terms of this predictive coding model, a more perception-oriented attentional set would keep error propagation more local, resulting in the exposure-specificity seen more in the psychophysics literature and visually-guided perceptual learning paradigms. A more comprehension-oriented attentional set would propagate errors farther to more abstract representations. In both cases, errors would propagate to where attention is focused, but more abstract representations would be more applicable to novel contexts, leading to the observed context-general perceptual learning. Kleinschmidt and Jaeger's (2015) ideal adapter framework fits within the same family of cognitive models as the predictive coding framework, but does not have any explicit hierarchy either within a linguistic level (i.e., allophones or more abstract phonemes) or between linguistic levels (i.e., phones within words). Hierarchical linguistic structure, however, is necessary as lexical biases and predictions feed the activation patterns of the sub-lexical units. Related to this word-level focus, our results are also in line with the Network Feedback Model (Wedel, 2012) where using lexical units as the focus of perceptual or phonological processes allows sub-lexical generalizations to follow.

The modified category in this study was designed to be an idiolectal feature devoid of social meaning. However, the role of attention in the perceptual learning of sound categories has implications for other subfields of linguistics that incorporate attention and segmental variation. For instance, in sociolinguistic theory, there are three categories of linguistic variables: INDICATORS, MARKERS, and STEREOTYPES (Labov, 1972). These are concepts used in the sociolinguistics literature to describe linguistic, most often phonological, variation in terms of whether a linguistic variable is subject to variation in usage in different social environments for a single speaker, and they largely map onto different levels of speaker (or listener) awareness. Indicators do not show variability within a speaker across social contexts and generally below speakers' level of awareness, while variables termed markers or stereotypes show intraspeaker variability. The role of attention proposed here would predict progressively less perceptual learning as awareness increases. Salient social variants (e.g., r-lessness in dialects of North American English) have been found to not be encoded as robustly as canonical productions (Sumner and Samuel, 2009). We predict that less socially salient variants would generalize to new forms more readily. If patterns of phonetic accommodation can be viewed as a type of perceptual learning, then there is evidence that less salient dialect differences are spontaneously imitated more than salient and stereotyped dialect features (Babel, 2010).

V. CONCLUSION

These results suggest that attentional sets are crucial to the generalization of perceptual learning to new contexts. These results provide additional support for recent advancements in models of speech perception where linguistic representations are treated as a balance of both more abstract elements and more fine-detailed elements that also incorporate aspects of social representations (Clark 2013; Sumner and Kataoka, 2013; Kleinschmidt and Jaeger, 2015). Understanding this balance in terms of attentional sets reintroduces an insightful perspective (Cutler *et al.*, 1987). Listeners have some control over how they listen.

ACKNOWLEDGMENTS

Thanks to Jamie Russell, Graham Haber, Jobie Hui, and Michelle Chan for their assistance with data collection. This work was funded by the University of British Columbia Arts Graduate Student Research Award and SSHRC No. 435-2014-1673.

APPENDIX A: CRITICAL /s/ WORDS, CONTROL /ʃ/ WORDS, AND FILLER WORDS USED IN THE LEXICAL DECISION TASK. PARTICIPANTS IN THE WORD-INITIAL CONDITION WERE PRESENTED WITH THE /s/-INITIAL WORDS, THE /ʃ/ WORDS, AND THE FILLER WORDS. THOSE IN THE WORD-MEDIAL CONDITION RECEIVED THE /s/-MEDIAL WORDS, THE /ʃ/ WORDS, AND THE FILLER WORDS

/s/-initial	/s/-medial	/ʃ/ words	Filler words		
ceiling	carousel	auction	acorn	doorbell	movie
celery	castle	brochure	acrobat	dryer	mural
cement	concert	cashier	antenna	elephant	napkin
ceremony	croissant	chandelier	apple	feather	omelet
saddle	currency	cushion	balloon	fingerprint	painter
safari	cursor	eruption	bamboo	garlic	piano
sailboat	curtsy	hibernation	buckle	goalie	ponytail
satellite	dancer	parachute	butterfly	gondola	popcorn
sector	dinosaur	patient	cabin	graffiti	referee
seedling	faucet	shadow	calendar	helicopter	table
seminar	fossil	shampoo	camel	ladder	tadpole
settlement	galaxy	shareholder	campfire	ladle	teapot
sidewalk	medicine	shelter	candy	librarian	theatre
silver	missile	shiny	cockpit	lightning	tire
socket	monsoon	shoplifter	collar	lumber	tortilla
sofa	pencil	shoulder	cowboy	mannequin	tractor
submarine	pharmacy	shovel	cradle	meadow	traffic
sunroof	tassel	sugar	cutlery	microwave	tunnel
surfboard	taxi	tissue	darkroom	minivan	umbrella
syrup	whistle	usher	diamond	motel	weatherman

APPENDIX B: NONWORDS USED IN THE LEXICAL DECISION TASK. PARTICIPANTS IN ALL EXPERIMENTAL CONDITIONS RECEIVED THIS LIST OF NONWORDS

Nonwords				
apolm	giptern	kowack	poara	tepple
arafimp	gittle	lefeloo	poltira	teygot
arnuff	glaple	lindel	pomto	theely
balrop	golthin	mogmet	potha	theerheb
bambany	goming	mopial	prickpor	thorkwift
bawapeet	gompy	motpem	prithet	thragkole
bettle	gorder	namittle	radadub	timmer
bimobel	hagrant	nartomy	rigloriem	tingora
bipar	hammertrent	nepow	rinbel	tinogail
blial	hintarber	neproyave	rindner	tirack
brahata	hovear	nidol	ripnem	tirrenper
danoor	iddle	noler	roggel	tovey
darnat	iglopad	nometin	roppet	toygaw
deoma	igoldion	nonifem	rudle	tuckib
follipocktel	impomo	omplero	talell	tuddom
foter	inoret	pammin	talot	tutrewy
gallmit	kempel	peltlon	tankfole	wapteep
gamtee	kimmer	pickpat	tayade	wekker
ganla	kire	pidbar	teerell	wogim
gippelfraw	klogodar	pluepelai	tello	yovernon

APPENDIX C: STATISTICAL MODELS

1. Exposure word recognition accuracy

Note: Random slope for Attention by Word removed due to near-zero variance causing convergence warnings, and random effect structure was specified as uncorrelated. Word Position could not be a random slope of Word, because the /s/ words are dependent on the Word Position condition.

Random effects:

Groups	Name	Variance	Standard deviation
Word	Intercept	1.31	1.15
Subject	Intercept	0.98	0.99
	Trial Number	0.03	0.18
	Trial Type = /s/	2.20	1.48
	Trial Type = /ʃ/	0.45	0.67
	Trial Number * Trial Type = /s/	0.35	0.59
	Trial Number * Trial Type = /ʃ/	0.08	0.27

Fixed effects:

Predictor	Estimate	St. Error	*z*-value	*p*-value
Intercept	3.95	0.20	19.34	<0.01
Trial Number	0.0	0.7	0.01	0.99
Trial Type = /s/	−2.01	0.31	−6.56	<0.01
Trial Type = /ʃ/	−0.18	0.35	−0.50	0.62
Attention	−0.35	0.25	−1.42	0.15
Word Position	−0.42	0.25	−1.70	0.09
Trial Number * Trial Type = /s/	0.41	0.12	3.36	<0.01
Trial Number * Trial Type = /ʃ/	0.01	0.13	0.05	0.96
Trial Number * Attention	−0.07	0.15	−0.48	0.63
Trial Number * Word Position	0.07	0.15	0.50	0.62
Attention * Trial Type = /s/	0.67	0.37	1.182	0.07
Attention * Trial Type = /ʃ/	0.38	0.29	1.31	0.19
Word Position * Trial Type = /s/	0.46	0.52	0.88	0.38
Word Position * Trial Type = /ʃ/	0.47	0.29	1.60	0.11
Attention * Word Position	−0.57	0.50	−1.16	0.25
Trial Number * Attention * Trial Type = /s/	0.13	0.24	0.52	0.60
Trial Number * Attention * Trial Type = /ʃ/	0.16	0.26	0.62	0.53
Trial Number * Word Position * Trial Type = /s/	−0.45	0.24	−1.84	0.07
Trial Number * Word Position * Trial Type = /ʃ/	−0.08	0.26	−0.29	0.77
Trial Number * Attention * Word Position	−0.04	0.29	−0.13	0.89
Attention * Word Position * Trial Type = /s/	−0.39	0.74	−0.52	0.60
Attention * Word Position * Trial Type = /ʃ/	−0.01	0.58	−0.01	0.99
Trial Number * Attention * Word Position * Trial Type = /s/	−0.36	0.49	−0.53	0.59
Trial Number * Attention * Word Position * Trial Type = /ʃ/	−0.23	0.52	−0.45	0.66

2. Exposure response times

Note: Random slope for Attention by Word removed due to near-zero variance causing convergence warnings, and random effect structure was specified as uncorrelated.

Random effects:

Groups	Name	Variance	Standard deviation
Word	Intercept	0.17	0.41
Subject	Intercept	0.17	0.41
	Trial Number	0.02	0.13
	Trial Type = /s/	0.07	0.26
	Trial Type = /ʃ/	<0.01	0.04
	Trial Number * Trial Type = /s/	0.02	0.13
	Trial Number * Trial Type = /ʃ/	0.01	0.09

Fixed effects:

Predictor	Estimate	St. Error	*t*-value	*p*-value
Intercept	−0.04	0.07	−0.65	0.51
Trial Number	−0.3	0.02	−1.61	0.11
Trial Type = /s/	0.37	0.09	4.12	< 0.01
Trial Type = /ʃ/	−0.04	0.11	−0.35	0.73
Attention	0.05	0.09	0.55	0.59
Word Position	0.03	0.09	0.35	0.72
Trial Number * Trial Type = /s/	−0.8	0.03	−3.05	<0.01
Trial Number * Trial Type = /ʃ/	0.02	0.02	1.06	0.29
Trial Number * Attention	−0.06	0.03	−1.82	0.07
Trial Number * Word Position	0.06	0.03	1.93	0.06
Attention * Trial Type = /s/	−0.07	0.07	−1.01	0.31
Attention * Trial Type = /ʃ/	−0.07	0.04	−1.53	0.13
Exposure type * Trial Type = /s/	−0.38	0.15	−2.61	<0.01
Exposure type * Trial Type = /ʃ/	0.02	0.04	0.39	0.71
Attention * Word Position	−0.02	0.17	−0.13	0.90
Trial Number * Attention * Trial Type = /s/	0.02	0.05	0.37	0.71
Trial Number * Attention * Trial Type = /ʃ/	0.03	0.05	0.56	0.57
Trial Number * Word Position * Trial Type = /s/	0.06	0.05	1.06	0.29
Trial Number * Word Position * Trial Type = /ʃ/	−0.06	0.05	−1.10	0.27
Trial Number * Attention * Word Position	0.00	0.07	0.02	0.98
Attention * Word Position * Trial Type = /s/	0.06	0.14	0.045	0.65
Attention * Word Position * Trial Type = /ʃ/	−0.01	0.09	−0.14	0.89
Trial Number * Attention * Word Position * Trial Type = /s/	0.18	0.11	1.6	0.10
Trial Number * Attention * Word Position * Trial Type = /ʃ/	−0.09	0.09	−0.96	0.34

3. Categorization models

a. Experimental data only

Random effects: Given the few levels of Item ($n = 4$), the number of parameters to be modeled had to be low. As such, only key random slopes were kept in the model, and the interactions of Step with Attention and Word Position were removed.

Groups	Name	Variance	Standard deviation
Item	Intercept	0.21	0.46
	Step	0.08	0.28
	Attention	0.01	0.11
	Exposure type	<0.01	0.03
	Attention * Word Position	0.03	0.16
Subject	Intercept	1.11	1.05
	Step	0.42	0.64

Fixed effects:

Predictor	Estimate	Standard Error	*z*-value	*p*-value
Intercept	1.03	0.25	4.05	<0.01
Step	−2.14	0.16	−13.63	<0.01
Word Position	−0.13	0.23	−0.57	0.57
Attention	0.29	0.23	1.27	0.21
Step * Word Position	−0.01	0.15	−0.07	0.94
Step * Attention	−0.13	0.15	−0.88	0.38
Word Position * Attention	−0.99	0.46	−2.17	0.03
Step * Word Position * Attention	0.36	0.29	1.22	0.22

b. Combined with control

Random effects: Note: The random slope for Condition and Step * Condition by Item was removed due to near-zero variance resulting in convergence warnings.

Groups	Name	Variance	Standard deviation
Item	Intercept	0.20	0.44
	Step	0.09	0.30
Subject	Intercept	1.06	1.03
	Step	0.39	0.62

Fixed effects:

Predictor	Estimate	Standard Error	*z*-value	*p*-value
Intercept	0.44	0.31	1.42	0.15
Step	−2.58	0.21	−12.49	<0.01
Condition = No Attention/Word Initial	0.42	0.30	1.40	0.16
Condition = Attention/Initial	0.63	0.31	2.05	0.04
Condition = No Attention/Final	1.05	0.31	3.37	<0.01
Condition = Attention/Final	0.25	0.30	0.84	0.40
Step * Condition = No Attention/Initial	0.46	0.20	2.31	0.02
Step * Condition = Attention/Initial	0.42	0.20	2.06	0.04
Step * Condition = No Attention/Final	0.29	0.21	1.41	0.16
Step * Condition = Attention/Final	0.61	0.20	3.08	<0.01

[1]Deviance contrast coding is used for all two-level independent variables, so the intercept of the model represents the grand mean (Word Position: Word-initial = 0.5, Word-medial = −0.5; Attention: No attention = 0.5, Attention = −0.5). Main effects for factors are calculated with other factors held at their average value, rather than at an arbitrary reference level. Trial Type (Filler, /s/, and /ʃ/) was coded using treatment (dummy) coding with Filler as the reference level. Numeric independent variables were centered prior to inclusion in models.

[2]Continuum was used as a random effect, serving as the Item grouping factor. However, there was only four minimal pair continua used in the categorization, so the random effect status may not be warranted. The estimates for the continua effects are likely not very reliable, but differences between continua are not the principle question being investigated. Use of a by-Continuum random effect structure with maximal random slopes allowed for estimation of the fixed effects that are not driven by one particular minimal pair continuum.

[3]While the step of the continua is sampled from six discrete levels, it is entered as a numeric variable in the models to reduce the complexity of models. Graphs of the categorization results show continua step as categorical factor to aid interpretation.

Babel, M. (**2010**). "Dialect divergence and convergence in New Zealand English," Lang. Soc. **39**(4), 437–456.

Brysbaert, M., and New, B. (**2009**). "Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English," Behav. Res. Methods **41**(4), 977–990.

Clark, A. (**2013**). "Whatever next? Predictive brains, situated agents, and the future of cognitive science," Behav. Brain Sci. **36**(3), 181–204.

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

Connine, C. M., Blasko, D. G., and Titone, D. (**1993**). "Do the beginnings of spoken words have a special status in auditory word recognition?," J. Mem. Lang. **32**(2), 193–210.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (**1987**). "Phoneme identification and the lexicon," Cognit. Psychol. **19**(2), 141–177.

Eisner, F., and McQueen, J. M. (**2005**). "The specificity of perceptual learning in speech processing," Percept. Psychophys. **67**(2), 224–238.

Gibson, E. J. (**1963**). "Perceptual learning," Ann. Rev. Psychol. **14**(1), 29–56.

Gilbert, C., Sigman, M., and Crist, R. (**2001**). "The neural basis of perceptual learning," Neuron **31**, 681–697.

Goldstone, R. L. (**1998**). "Perceptual learning," Ann. Rev. Psychol. **49**(1), 585–612.

Idemaru, K., and Holt, L. L. (**2014**). "Specificity of dimension-based statistical learning in word recognition," J. Exp. Psychol.: Human Percept. Perform. **40**(3), 1009–1021.

Jesse, A., and McQueen, J. M. (**2011**). "Positional effects in the lexical retuning of speech perception," Psychonomic Bull. Rev. **18**, 943–950.

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., and Banno, H. (**2008**). "Tandem-straight: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3933–3936.

Kleber, F., Harrington, J., and Reubold, U. (**2012**). "The relationship between the perception and production of coarticulation during a sound change in progress," Lang. Speech **55**, 383–405.

Kleinschmidt, D. F., and Jaeger, T. F. (**2015**). "Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel," Psychol. Rev. **122**(2), 148–203.

Kraljic, T., Brennan, S. E., and Samuel, A. G. (**2008a**). "Accommodating variation: Dialects, idiolects, and speech processing," Cognition **107**(1), 54–81.

Kraljic, T., and Samuel, A. G. (**2005**). "Perceptual learning for speech: Is there a return to normal?," Cognitive Psychol. **51**(2), 141–178.

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. (**2008b**). "First impressions and last resorts: How listeners adjust to speaker variability," Psychol. Sci. **19**(4), 332–338.

Kuhl, P. K. (**1979**). "Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories," J. Acoust. Soc. Am. **66**(6), 1668–1679.

Labov, W. (**1972**). *Sociolinguistic Patterns* (No. 4) (University of Pennsylvania Press, Philadelphia, PA).

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

McClelland, J. L., and Elman, J. L. (**1986**). "The TRACE model of speech perception," Cognitive Psychol. **18**(1), 1–86.

McClelland, J. L., Mirman, D., and Holt, L. L. (**2006**). "Are there interactive processes in speech perception?," Trends Cognit. Sci. **10**(8), 363–369.

McLennan, C. T., and Luce, P. A. (**2005**). "Examining the time course of indexical specificity effects in spoken word recognition," J. Exp. Psychol.: Learn., Memory, Cognit. **31**(2), 306–321.

Mirman, D., McClelland, J. L., Holt, L. L., and Magnuson, J. S. (**2008**). "Effects of attention on the strength of lexical influences on speech perception: Behavioral experiments and computational mechanisms," Cognit. Sci. **32**(2), 398–417.

Mitterer, H., Scharenborg, O., and McQueen, J. M. (**2013**). "Phonological abstraction without phonemes in speech perception," Cognition **129**(2), 356–361.

Norris, D., and Cutler, A. (**1988**). "The relative accessibility of phonemes and syllables," Percept. Psychophys. **43**(6), 541–550.

Norris, D., and McQueen, J. M. (**2008**). "Shortlist B: A Bayesian model of continuous speech recognition," Psychol. Rev. **115**(2), 357–395.

Norris, D., McQueen, J. M., and Cutler, A. (**2000**). "Merging information in speech recognition: Feedback is never necessary," Behav. Brain Sci. **23**(3), 299–325.

Norris, D., McQueen, J. M., and Cutler, A. (**2003**). "Perceptual learning in speech," Cognit. Psychol. **47**(2), 204–238.

Pitt, M., and Szostak, C. (**2012**). "A lexically biased attentional set compensates for variable speech quality caused by pronunciation variation," Lang. Cognit. Process. **27**, 37–41.

Pitt, M. A., and Samuel, A. G. (**1990**). "Attentional allocation during speech perception: How fine is the focus?," J. Mem. Lang. **29**(5), 611–632.

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.

Reinisch, E., Weber, A., and Mitterer, H. (**2013**). "Listeners retune phoneme categories across languages," J. Exp. Psychol.: Human Percept. Perform. **39**(1), 75–86.

Reinisch, E., Wozny, D. R., Mitterer, H., and Holt, L. L. (**2014**). "Phonetic category recalibration: What are the categories?," J. Phonetics **45**, 91–105.

Samuel, A. G. (**1981**). "Phonemic restoration: Insights from a new methodology," J. Exp. Psychol.: General **110**(4), 474–494.

Scharenborg, O., and Janse, E. (**2013**). "Comparing lexically guided perceptual learning in younger and older listeners," Attn., Percept., Psychophys. **75**(3), 525–536.

Scharenborg, O., Weber, A., and Janse, E. (**2015**). "The role of attentional abilities in lexically guided perceptual learning by older listeners," Attn., Percept., Psychophys. **77**(2), 493–507.

Shankweiler, D., Strange, W., and Verbrugge, R. R. (**1977**). "Speech and the problem of perceptual constancy," in *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, edited by R. Shaw and J. Bransford (Lawrence Erlbaum Assoc., Hillsdale, NJ), pp. 315–345.

Sumner, M., and Kataoka, R. (**2013**). "Effects of phonetically-cued talker variation on semantic encoding," J. Acoust. Soc. Am. **134**(6), EL485–EL491.

Sumner, M., and Samuel, A. G. (**2009**). "The effect of experience on the perception and representation of dialect variants," J. Mem. Lang. **60**(4), 487–501.

Theodore, R. M., Blumstein, S. E., and Luthra, S. (**2015**). "Attention modulates specificity effects in spoken word recognition: Challenges to the time-course hypothesis," Attn., Percept., Psychophys. **77**(5), 1674–1684.

Vroomen, J., van Linden, S., de Gelder, B., and Bertelson, P. (**2007**). "Visual recalibration and selective adaptation in auditory-visual speech perception: Contrasting build-up courses," Neuropsychologia **45**(3), 572–577.

Wedel, A. (**2012**). "Lexical contrast maintenance and the organization of sublexical contrast systems," Lang. Cognit. **4**(4), 319–355.

Witteman, M. J., Weber, A., and McQueen, J. M. (**2013**). "Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation," Attn., Percept., Psychophys. **75**(3), 537–556.

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.