Research Article

Phonetica 2019;76:179–200
DOI: 10.1159/000494929

Received: April 3, 2017
Accepted: October 29, 2018
Published online: May 21, 2019

The Goldilocks Zone of Perceptual Learning

Molly Babel[a] Michael McAuliffe[b] Carolyn Norton[a] Brianne Senior[a] Charlotte Vaughn[c]

[a]University of British Columbia, Vancouver, BC, Canada; [b]McGill University, Montreal, QC, Canada; [c]University of Oregon, Eugene, OR, USA

Abstract

Background/Aims: Lexically guided perceptual learning in speech is the updating of linguistic categories based on novel input disambiguated by the structure provided in a recognized lexical item. We test the range of variation that allows for perceptual learning by presenting listeners with items that vary from subtle within-category variation to fully remapped cross-category variation. ***Methods:*** Experiment 1 uses a lexically guided perceptual learning paradigm with words containing noncanonical /s/ realizations from s/ʃ continua that correspond to "typical," "ambiguous," "atypical," and "remapped" steps. Perceptual learning is tested in an s/ʃ categorization task. Experiment 2 addresses listener sensitivity to variation in the exposure items using AX discrimination tasks. ***Results:*** Listeners in experiment 1 showed perceptual learning with the maximally ambiguous tokens. Performance of listeners in experiment 2 suggests that tokens which showed the most perceptual learning were not perceptually salient on their own. ***Conclusion:*** These results demonstrate that perceptual learning is enhanced with maximally ambiguous stimuli. Excessively atypical pronunciations show attenuated perceptual learning, while typical pronunciations show no evidence for perceptual learning. AX discrimination illustrates that the maximally ambiguous stimuli are not perceptually unique. Together, these results suggest that perceptual learning relies on an interplay between confidence in phonetic and lexical predictions and category typicality.

Introduction

Lexically guided perceptual learning in speech uses novel input, often disambiguated by the linguistic scaffolding offered by a recoverable lexical frame, to update linguistic categories. It has been argued that perceptual learning is a means by which

KARGER

© 2019 S. Karger AG, Basel

E-Mail karger@karger.com
www.karger.com/pho

Molly Babel
Department of Linguistics, University of British
Columbia, 2613 West Mall, Totem Field Studios
Vancouver, BC V6T 1Z4 (Canada)
E-Mail molly.babel@ubc.ca

listeners manage the immense amounts of cross-talker variability in spoken language. When listeners are exposed to the novel input in sentences produced by nonnative accents, they rapidly learn and adapt in response to these nonnative accents, generalizing their implicitly updated phonetic and phonological knowledge to novel voices (Baese-Berk, Bradlow, & Wright, 2013; Bradlow & Bent, 2008; Clarke & Garrett, 2004). This adaptation occurs even for utterances that match on sublexical or syntactic levels (Cooper & Bradlow, 2016). Children also show the ability to rapidly adapt to unfamiliar accents (Holt & Bent, 2017), and both children and adults demonstrate better performance – in terms of speed or magnitude of learning – in sentences that are high in semantic predictability (Bradlow & Alexander, 2007; Holt & Bent, 2017).

Scaling downwards from the sentential level, listeners' enhanced ability to perceptually learn in linguistically predictable environments is also supported at the single-word level. Norris, McQueen, and Cutler (2003) introduced a paradigm for lexically guided perceptual learning in speech using individual lexical items. They synthesized a fricative sound that was ambiguous between /s/ or /f/ – denoted here as /?sf/ – and used it to replace word-final /s/ or /f/ fricatives in Dutch words. Listeners were then exposed to the ambiguous fricative in contexts where they would *expect* an /f/ or an /s/. They quantified their listeners' perceptual learning in a post-test categorization task with an /ɛs/–/ɛf/ continuum and demonstrated that listeners who were exposed to the ambiguous fricative in the context of /s/ words increased their perception of what was an acceptable /s/ in the post-test, while listeners exposed to the ambiguous /?sf/ in /f/ words likewise expanded their /f/ category to include these ambiguous sounds. Crucially, listeners exposed to the acoustically identical /?sf/ fricatives in nonwords showed no perceptual learning, as they had no linguistic structure to guide their interpretation of the ambiguous fricative. Note that in addition to lexical status assisting in the interpretation of a sound, the phonological position of the critical sound within the word also appears to be crucial. When ambiguous sounds replace critical fricatives in initial position, listeners do not show perceptual learning (Jesse & McQueen, 2011; McAuliffe & Babel, 2016). Similarly, when ambiguous sounds hamper a listener's ability to identify the item as a word (Clarke-Davidson, Luce, & Sawusch, 2008; Scharenborg & Janse, 2013; Witteman, Weber, & McQueen, 2013), the listener is less likely to perceptually learn. This lexically guided perceptual learning paradigm has been exploited considerably (Kraljic & Samuel, 2005, 2006, 2007; Maye, Aslin, & Tanenhaus, 2008; Reinisch, Weber, & Mitterer, 2013; Weatherholtz, 2015). Researchers have also demonstrated that visual information about place of articulation of a speech sound can disambiguate phonetically anomalous acoustic variation, supporting perceptual learning of speech sounds in the absence of lexical information (Vroomen, van Linden, Keetels, De Gelder, & Bertelson, 2004; Reinisch, Wozny, Mitterer, & Holt, 2014).

With this empirical background, we can consider the mechanisms proposed to account for lexically guided perceptual learning. A necessary component of a model of lexically guided perceptual learning is the association between the phonetic detail of an auditory object and its phonological category, determined in part by lexical context. In cases where we see perceptual learning, this phonetic detail must be used to update the multidimensional phonetic representation of a phonological category. The crux of the theoretical debate is how these connections are made. In TRACE, an interactive model of speech perception, McClelland and colleagues (McClelland & Elman, 1986; McClelland, Mirman, & Holt, 2006) argue that the activation of a lexical item propagates down to a sublexical level, which facilitates the mapping of the auditory input and the phonological category. An alternate model, Merge-B (Norris, McQueen, & Cutler, 2016), is distinguished from TRACE in part by its *lack* of activation from the lexical to sublexical levels. In Merge-B, different lines of evidence are evaluated jointly at a decision stage. The auditory input, ambiguous or not, is considered as evidence for a particular sublexical categorization, and that information is merged with and considered alongside the evidence in support of a particular lexical expectation. Thus, listeners may have a hypothesis about what they expect to hear, and this hypothesis is evaluated in conjunction with the auditory evidence of the perceptual event. This prediction-focused Bayesian framework proposed by Norris and colleagues is conceptually similar to Kleinschmidt and Jaeger's (2015) ideal adapter framework, which is designed with lower-level perceptual categorization in mind, and Clark's (2013) predictive coding model for perception, which is intended to be domain-general.

The simultaneous flexibility (e.g., listeners adapting to new accents) and stability (e.g., exploiting perceptual representations for similar-sounding talkers and accents) is a key feature of human speech perception abilities and a crucial consideration in the design of Bayesian models of perceptual learning (Clark, 2013; Kleinschmidt & Jaeger, 2015; Norris et al., 2016). This balance of flexibility and stability is also a necessity in models of diachronic sound change, where recent theories and computational models build in mechanisms to evaluate and potentially discard ambiguous items (Hay, Pierrehumbert, Walker, & LaSchell, 2015; Sóskuthy, 2015; Wedel, 2012) that may nevertheless be accurately recognized. Recently, Hay et al. (2015) have proposed a mechanism that discards tokens or prevents their encoding in memory when that recognition process involved excessive ambiguity, even in cases with ultimately accurate recognition. Although ambiguity and lower encoding strength might not interfere with the ultimate recognition of an intended item, a discard mechanism predicts that in cases of more extreme category atypicality, there should be less perceptual learning since these items are not used to update phonetic representations.

Yet, there is also the issue of subtler shifts in pronunciation. It is likely that more typical – yet still not prototypical – productions will easily be perceived as the intended category and therefore they may more readily induce perceptual learning than a perfectly ambiguous category. This line of thinking underscores Bradlow and Bent's (2008) speculation that their listeners' greater degree of adaptation to foreign-accented speakers with higher intelligibility was due to facilitated lexical feedback for those higher intelligibility speakers. However, these typical-but-not-prototypical instances may be in some sense too typical and thus not merit adaptation for an ideal adapter (Kleinschmidt & Jaeger, 2015; Norris et al., 2016); Kleinschmidt & Jaeger refer to this as "recognizing the familiar." While listeners may be sensitive to subtle acoustic phonetic differences in the realization of phonetic and phonological categories between talkers and indeed track and use this information in talker recognition (e.g., Allen & Miller, 2004; Theodore, Myers, & Lomibao, 2015), in some instances, small changes may fall within the familiar range and not warrant an update (see also Cutler, 2012). We know that listeners' sensitivity to phonetic differences is highest in psychoacoustically focused tasks where listeners' attention is directed towards perception-oriented listening (Cutler, Mehler, Norris, & Segui, 1987). When listeners are engaged in a task that directs their attention to comprehension of lexical items (e.g., a lexical decision task) where the sound of interest is not in word-initial position, they may show decreased sensitivity to subtler shifts in production, akin to reduced sensitivity to within-category variation generally (e.g., Liberman et al., 1957).

Most lexically guided perceptual learning paradigms use maximally ambiguous stimuli in selecting the targets of perceptual learning, where "maximal ambiguity" is determined by either acoustic or perceptual means (e.g., Norris et al., 2003; Kraljic, Brennan, & Samuel, 2008a; Kraljic, Samuel, & Brennan, 2008b; Reinisch et al., 2013). Regardless of the method of assessment, the goal is to determine the extent of learning that occurs when listeners are presented with stimuli that are perfectly ambiguous between two categories (i.e., stimuli where half of the time a sound is categorized as a member of one category and half the time as a member of a different category). Robust perceptual learning has been found with such methods. Our goal here is to assess how the typicality of the pronunciation variant affects perceptual learning. There is already empirical evidence that phonetic variation that deviates considerably from a canonical pronunciation (i.e., with heavily accented items in a nonnative accent, Witteman et al., 2013), that interferes with lexical recognition (i.e., when listeners have lower thresholds in a lexical decision task, Scharenborg & Janse, 2013), or that is less intelligible because of signal degradation (i.e., one-channel speech, Sohoglu & Davis, 2016) does not lead to perceptual learning. Further, Ganong's (1980) finding that listeners' boundaries shift more in accordance with lexical bias at their own perceptual category boundary than at continuum end points suggests different amounts of retuning across degrees of ambiguity. Several recent computational models of diachronic sound change (Hay et al., 2015; Sóskuthy, 2015; Wedel, 2012) suggest the need for filtering excessively ambiguous tokens from phonetic memory, even in cases where the item may have been comprehended as intended. Synchronic models, like the ideal adapter framework (Kleinschmidt & Jaeger, 2015), predict that subtler phonetic variants might not warrant an update to the system. Real-life exposure to pronunciation variants is likely to range from highly typical to highly atypical variants (cf. Sumner, 2011), both from incipient sound changes and from speakers from diverse dialect backgrounds. Given this real-life variation, understanding the bounds of perceptual learning in terms of phonetic ambiguity is ultimately important for our understanding of how perceptual learning may function in real-life speech situations. In experiment 1, we test listeners' perceptual learning of /s/ words across four points along a continuum from /s/ to /ʃ/: "typical" /s/ (70% /s/ word identification), "ambiguous" /s/ (50% /s/ word identification), "atypical" /s/ (30% /s/ word identification), and "remapped" /s/ (0% /s/ word identification). We predict that listeners will show the most learning at the ambiguous point, the point at which maximum ambiguity occurs. This point should facilitate adaptation because it is neither too familiar nor too different from the typical signal to interfere with recognition, making it the "Goldilocks" zone of perceptual learning.

Experiment 1: Perceptual Learning

Methods

Participants

One hundred twenty listeners participated in this study, 25 in each of the four conditions in experiment 1 and 20 in a pretest to determine items used in experiment

1. Additionally, data from the 25 control participants in McAuliffe (2015) are used in the analyses below. All participants were self-reported native English speakers with no known speech, language, or hearing impairments. We recruited participants from the University of British Columbia community and compensated them with either 10 CAD or course credit for their time.

Stimuli

Stimuli were part of a larger set from McAuliffe (2015). A college-aged (early twenties), male, phonetically aware native speaker of Vancouver English produced 100 nonwords, 60 filler words, 20 critical items, 20 control words, and 8 test words. Nonwords were bi- or trisyllabic items that abided by English phonotactics (e.g., *lindel*). Filler stimuli were English words that did not contain any sibilant fricatives. The critical items consisted of English words with /s/ word-medially (e.g., *carousel* / keɹəsɛɫ/), while the control items were English words with /ʃ/ either word-initially (n = 10, e.g., *shoulder* /ʃoʊɫdɹ̩/) or word-medially (n = 10, e.g., *cushion* /kʊʃn̩/). The eight test items consisted of four monosyllabic /s/-/ʃ/ minimal pairs with the contrasting phoneme in word-initial position (*sack-shack, sigh-shy, sin-shin, sock-shock*). Apart from the medial /s/ in the critical items, the initial or medial /ʃ/ in the control words, and the initial /s/ or /ʃ/ in the test words, none of the stimuli contained any sibilants.

The speaker produced the critical items both normally and with the /s/ replaced with /ʃ/ (i.e., *carousel* was produced as both /kɛɹə**s**ɛɫ/ and /kɛɹə**ʃ**ɛɫ/). These productions were then time-aligned and morphed together using TANDEM-STRAIGHT in Matlab (Kawahara et al., 2008) to create twenty 11-step continua (Fig. 1) ranging from the most /s/-like pronunciation (step 11) to the most /ʃ/-like pronunciation (step 1). The minimal word pairs for the test continua were synthesized into 11-step continua in a similar manner. All nonwords, fillers, and control items were resynthesized in TANDEM-STRAIGHT to control for stimulus quality.

The 20 critical item continua and 8 minimal pair continua were pretested using a categorization task in E-Prime 2.0 (Psychology Software Tools, 2012). Twenty native English speakers were presented with stimuli from the various steps and asked to categorize each item as either a word or nonword (for the critical item continua) or as an "s" word or a "sh" word (for the minimal pair continua) using a response box. Each step from each of the continua was presented once, randomized for each participant, and the button labels were counterbalanced across participants, so half the participants used the left button for word and the other half the right button. Furthermore, within each of those halves, half used the left button for "s" word response and half the right button.

For each continuum, we calculated the proportion of "word" responses – or "s" responses for the minimal pair continua – at each step. The 6 steps nearest to the 50% crossover points on the minimal pair continua were used as test items in all experiments. For the critical item continua, the steps corresponding to 70% /s/ word identification (typical /s/), 50% /s/ word identification (ambiguous /s/), 30% /s/ word identification (atypical /s/), and 0% /s/ word identification (remapped /s/; these items were identified as nonwords containing /ʃ/ 100% of the time) were selected for use in the appropriate conditions in the exposure phase. The step numbers for each critical item for each condition are reported in Table 1 in the Appendix. These values are also shown visually in Figure 1 where step 11 and step 1 indicate the /s/ and /ʃ/ end points

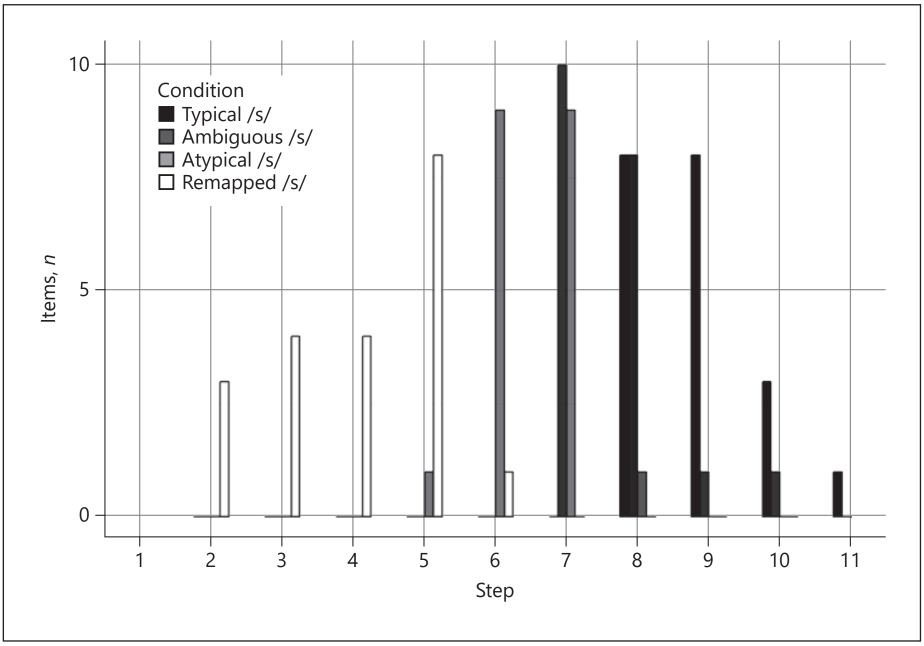

Fig. 1. Distribution of critical items at each step for the four perceptual learning conditions. Step 1 and step 11 indicate the /ʃ/ and /s/ end points of the continua, respectively, and higher steps are more similar to the /s/ end point.

of the continua, respectively. The distributions of the steps associated with each critical item in each of the four perceptual learning conditions – typical, ambiguous, atypical, and remapped (these terms are used in the remainder of the text) – are presented, illustrating that for each condition, a distribution of steps associated with shifts in word identification rates was used.

Procedure

Following the methods in McAuliffe (2015), each condition of the perceptual learning experiment consisted of two parts: (1) an exposure phase, in which participants were introduced to the critical items with ambiguous fricatives, and (2) a test phase, in which participants categorized minimal pair stimuli as "s" or "sh." Participants were given oral instructions about both tasks at the beginning of the experiment, and written instructions were also provided on screen prior to the start of each task. Following completion of the two tasks, participants filled out a language background questionnaire.

Exposure: Lexical Decision Task. Exposure took the form of a lexical decision task. Participants were seated in sound-attenuated cubicles and wore AKG-K240 headphones. On each trial, listeners were first presented with a blank screen for 500 ms. Then a response screen with two options (e.g., *1 = word, 5 = nonword*) was displayed for 500 ms, followed by an auditory stimulus presented over the headphones. Participants had 3,000 ms from the onset of the auditory stimulus to decide if the item was an English word or a nonword and to register their response via a button box. Button-label assignments were counterbalanced across participants (i.e., half the participants used the left button to respond with a word response and the other half the right button). If participants did not respond within the 3,000 ms time frame, a message appeared on screen informing them that no response had been detected, and the experiment progressed to the next trial. Listeners completed a total of 200 trials and were given self-paced breaks every 50 trials. Trial order was pseudorandomized to ensure the first 6 trials contained filler items and that critical or control items did not appear consecutively, with each listener in a condition receiving a different trial order. All listeners heard the same filler words, filler nonwords, and control /ʃ/ words in the exposure phase, but the critical /s/ items differed according to condition assignment; listeners were presented with either typical, ambiguous, atypical, or remapped critical /s/ items.

Test: Categorization Task. The test phase was structured similarly to the exposure phase. Listeners completed a two-alternative, forced-choice identification task. On each trial, they heard one of the 6 steps from one of the minimal word continua and were presented with two options on the screen, the "s" word (e.g., *sack*) or the corresponding "sh" word (e.g., *shack*). They were given 3,000 ms after stimulus onset to indicate their response by pressing the appropriate button on the response box. Participants completed 168 trials (4 continua × 6 steps × 7 repetitions of each step) and received self-paced breaks every 40 trials.

A control group of 25 participants completed just the test phase (from McAuliffe, 2015). Their results served as a baseline categorization function.

Experiment 1 Results

Lexical Endorsement in Exposure Phase

Responses faster than 200 ms or slower than 2,500 ms from the onset of the trial were excluded from analysis. Accuracy on the lexical decision task was high overall, 94% for filler words and 88% for nonwords. Trials with nonword stimuli were excluded for further analyses of accuracy. A logistic mixed effects model was used to analyze word recognition accuracy as the dependent variable, fitted in R (R Core Team, 2017) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015).[1] Fixed effects for exposure item type (filler, /s/, /ʃ/), and category typicality (typical, ambiguous, atypical, remapped) were used to assess the differences in word endorsement rates of the critical /s/ words compared to other item types across the different exposure conditions. Random effects for subjects and words, as well as random slopes for each were added where possible (by-speaker random slope for exposure item type and by-word random slope for category typicality). Both exposure item type and category typicality were contrast coded with filler and typical as the reference levels, respectively.

Within the typical category exposure, /s/ words were endorsed at a lower rate than filler items only marginally [B = –0.79; SE = 0.45; z = –1.76, p = 0.08]; however, there was significantly less endorsement of /s/ words compared to filler items for the ambiguous category [B = –1.24; SE = 0.60; z = –2.08; p = 0.04], atypical [B = –3.13; SE = 0.55; z = –5.66; p < 0.001], and remapped [B = –3.89; SE = 0.55; z = –7.04; p <

[1] Formula used: accuracy – exposure item type × category typicality + (1 + exposure item type | subject) + (1 + category typicality | word).

0.001]. A post hoc analysis of endorsement rates for /s/ words across conditions was done using the lsmeans package (Lenth, 2016), with p values adjusted using the Tukey method. Endorsement rates for /s/ words were not significantly different between typical and ambiguous [B = 0.37; SE = 0.48; z = 0.77; $p = 0.89$], but rates were significantly different between both ambiguous and atypical [B = 2.13; SE = 0.46; z = 4.66; $p < 0.001$], and atypical and remapped [B = 1.39; SE = 0.43; z = 3.21; $p = 0.007$]. Endorsement rates for filler words and control /ʃ/ words did not differ significantly across exposures [all $p > 0.10$]. These results are shown in Figure 2.

Categorization in Test Phase

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 interpreted as indicative of perceptual learning, since a higher likelihood of /s/ response is associated with an expanded /s/ category. As above, responses faster than 200 ms or slower than 2,500 ms from the onset of the trial were excluded from analysis. Listeners' categorizations were analyzed in a logistic mixed effects model using step and exposure category typicality (control, typical, ambiguous, atypical, remapped) as predictors.[2] There were random intercepts for subject (with step as a random slope) and item (with step, typicality, and their interaction as random slopes). The control condition was used as the reference level. Step was centered, so effects around zero represent categorization behavior similar to the pretest participants (and the control participants), and effects above zero represent participants categorizing more of the continua as the /s/ word in the minimal pair rather than the /ʃ/ word.

The model had a nonsignificant intercept [B = 0.16; SE = 0.28; z = 0.58; $p = 0.56$] for the control condition. The ambiguous condition was significantly different from the control condition [B = 1.06; SE = 0.28; z = 3.67; $p < 0.01$], but the typical condition showed no effect [B = 0.15; SE = 0.30; z = 0.50; $p = 0.62$]. There were trending effects for both atypical [B = 0.48; SE = 0.30; z = 1.59; $p = 0.11$] and remapped [B = 0.48; SE = 0.29; z = 1.64; $p = 0.10$]. All interactions between step and category typicality were not significant [all $p > 0.10$].

Discussion

All effects trended towards perceptual learning (more of the continuum categorized as /s/), but the effect sizes in conditions other than ambiguous may be smaller than could be reliably found with the power in this study. The sample size used in this study (25 participants per condition) was based on previous work (McAuliffe & Babel, 2016; Reinisch et al., 2013) for reliably finding moderate effect sizes.

As shown in Figure 3a, the categorization function for the typical condition is the least different from the control condition, indicating a lack of perceptual learning for this group. Listeners presented with the atypical tokens and remapped tokens show moderate amounts of perceptual learning, but not significantly more than those in the control condition. Listeners in the ambiguous condition, however, show robust perceptual learning. Figure 3b plots the effect sizes of each typicality condition taken from the logistic mixed-effects model. These are the coefficients and standard errors

[2] Formula used: accuracy – step × category typicality+ (1 + step | subject) + (1 + step × category typicality | item).

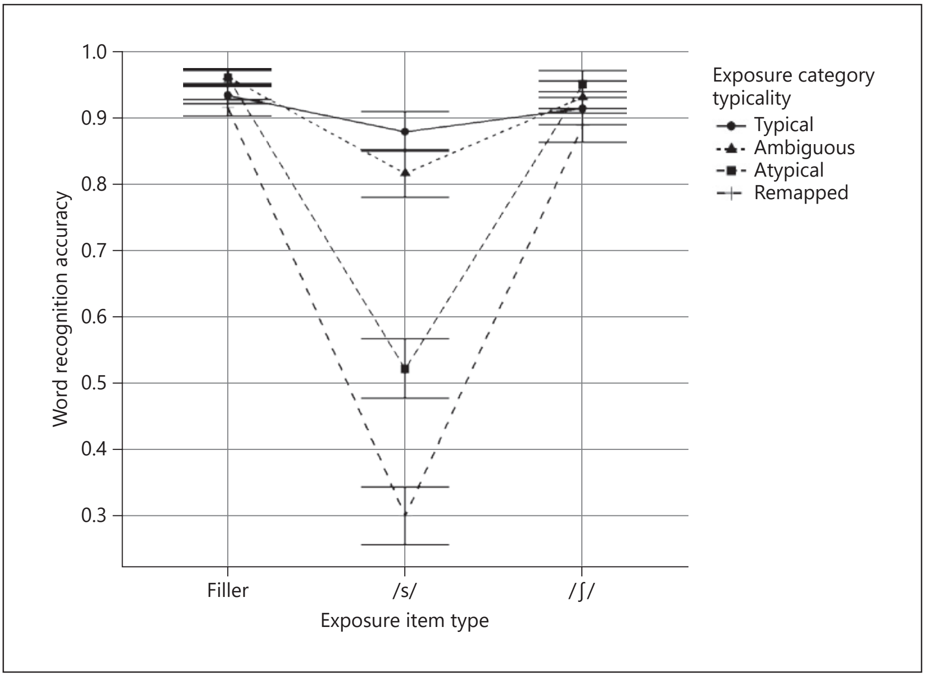

Fig. 2. Within-subject mean accuracy (*y* axis) for filler, /s/ words, and control /ʃ/ words (*x* axis) in the lexical decision exposure phase for the four experimental groups. Error bars represent 95% confidence intervals.

of the statistical models and are another visualization of the increased magnitude of perceptual learning for listeners exposed to items that together presented the maximally ambiguous /s/ distribution.

Listeners appear to learn the most from stimuli that are maximally ambiguous, where maximum ambiguity is determined by pretests in which listeners categorized word-nonword continua as words and nonwords. Does this mean that the maximally ambiguous stimuli are perceptually unique? In other words, are listeners maximally sensitive to this phonetic variant? To assess this, we conducted speeded AX discrimination tasks (i.e., where participants decide whether two items are same or different) with the whole-word stimulus listeners heard in the lexical decision exposure phase of the perceptual learning tasks (experiment 2A) and with the fricative-vowel sequences of those words excised as free-standing CV sequences (experiment 2B).

Experiments 2A and 2B: Categorical Perception

Experiment 2A: Categorical Perception with Words

Methods

Participants. Twenty listeners participated in this task. As in experiment 1, all participants were self-reported native English speakers recruited from the University

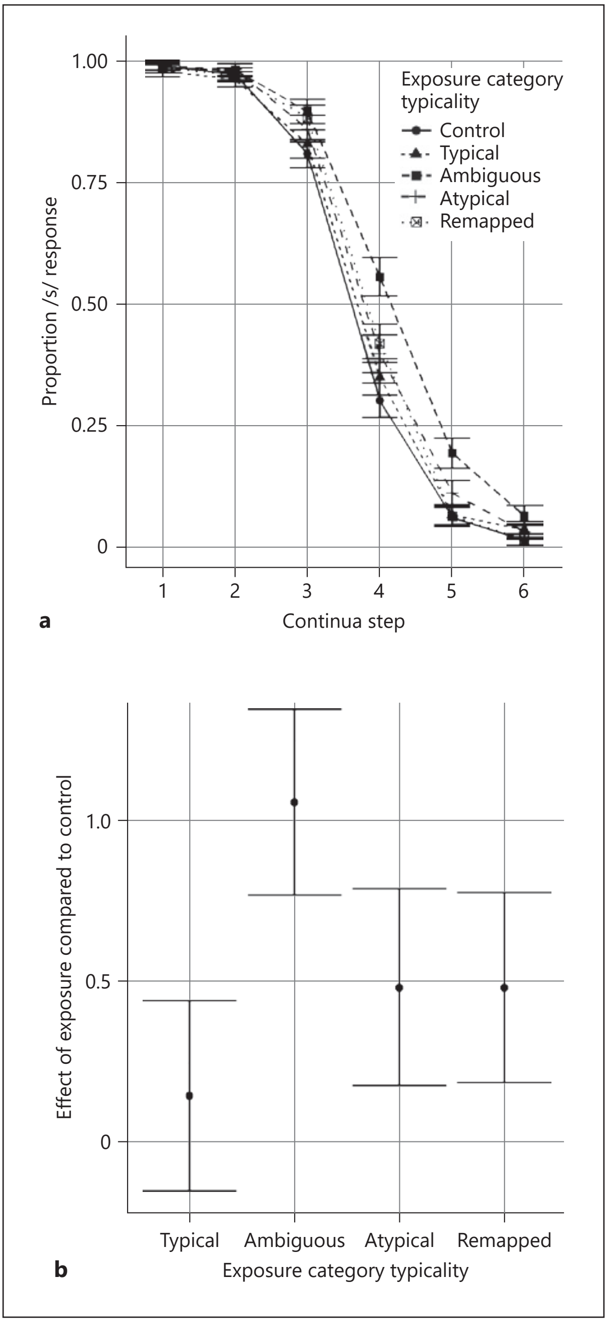

Fig. 3. a Proportion /s/ word responses (y axis) across the 6-step minimal pair continua (x axis) for the four experimental conditions and the control condition from McAuliffe (2015). Error bars represent 95% confidence intervals. **b** Effect sizes (y axis) across experimental conditions (x axis). Effect size is plotted as the coefficient and standard error from the mixed effects models. Experimental condition is ordered from most typical (typical) to least typical (remapped) /s/.

DOI: 10.1159/000494929

of British Columbia community and reported no known speech, language, or hearing impairments. Participants were compensated with partial course credit for their participation. None of the listeners in this experiment had participated in experiment 1.

Stimuli. Stimuli for the whole-word categorization task consisted of 40 critical words containing a word-initial (e.g., *celery*) or word-medial /s/ (e.g., *cursor*, from McAuliffe & Babel, 2016) that were used in the exposure phase of experiment 1. Stimuli maintained the category typicality steps corresponding to typical, ambiguous, atypical, and remapped categories used in experiment 1, along with a new set of "canonical" items (100% /s/ word endorsement from the pretest).

Procedure. Participants were seated up to four at a time at individual PC workstations outfitted with AKG-K240 headphones and a PST serial response box. Upon completion of the perception task, participants completed a language background questionnaire.

Participants received instructions orally and visually on the desktop prior to the experiment, instructing them to complete an AX categorization task of whole-word pairs. Stimuli contained a critical word-medial or word-initial /?sʃ/ fricative. Listeners were presented with pairs of adjacent different stimuli (e.g., canonical/typical, typical/ambiguous, ambiguous/atypical, atypical/remapped; "different trials") or pairs of identical stimuli ("same trials") with a 500-ms interstimulus interval. Participants were given up to 5,000 ms to indicate whether the auditory stimulus pairs were the "same" or "different" by pressing the appropriate button on the response box (e.g., 1 = same, 5 = different). Button-label assignments were counterbalanced across participants. Participants received visual feedback (accuracy and response time) following each trial. After every 50 trials, participants were provided with a self-administered break. Participants completed a total of 544 trials (272 same; 272 different). Not all words had the same number of comparisons because of the shape of the word endorsement response functions for some items. As reported in Table 1 in the Appendix, the continuum steps associated with adjacent categories were identical for some items (i.e., the response function was flat across a range of steps).

Experiment 2A Results

Although listeners responded to stimuli containing /s/ in both initial (e.g., *silver*, /sɪlvɹ̩/) and medial (e.g., *croissant*, /kɹəsɑnt/) positions, we restricted this analysis to responses to /s/-medial stimuli in order to be maximally comparable to the exposure stimuli in experiment 1. Response time outliers, defined as responses more than 2 standard deviations away from the group mean, were removed, eliminating 4.8% of the data. The remaining trials have a mean response time of 1,064 ms (SD = 287). Accuracy on the remaining trials was used to calculate the signal detection theoretic values d' and c for the analysis of the categorical perception data. d' is a measure of sensitivity to the signal versus noise based on the inverse of the cumulative distribution function of the hit rate (H) and false alarm rate (FA), with higher values of d' indicating greater sensitivity. It was calculated for roving same-different discrimination tasks following Macmillan and Creelman (2004) using the psyphy package in R (Knoblauch, 2014). c, a measure of response bias, was calculated as $-0.5 \times (z(H) + z(FA))$, where z refers to the inverse of the cumulative distribution function; based on how we calculated these values, a positive value indicates a bias to respond "same."

For the sensitivity analysis, d' values from the word and CV experiments were analyzed in a repeated-measures ANOVA with pair (e.g., canonical/typical, typical/

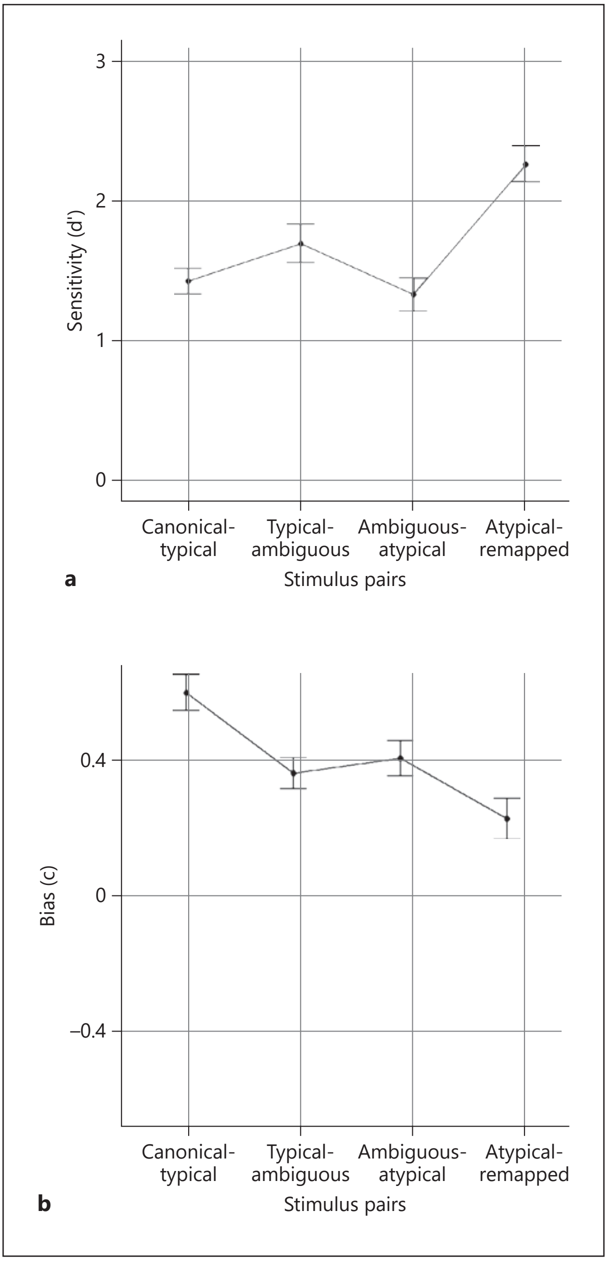

Fig. 4. Experiment 2A (words). **a** d', a measure of sensitivity, on the *y* axis for each stimulus pair type (*x* axis). Higher values indicate higher perceptual sensitivity. **b** c, a measure of response bias, on the *y* axis for stimulus pairs (*x* axis). Error bars show standard errors.

ambiguous, etc.) repeated across listeners. The d' analysis showed a main effect of pair [F(3, 99) = 21.24, $p < 0.001$]. The Tukey honest significant difference test was used to compare sensitivity between stimulus pairs. Listeners discriminated the atypical/remapped pair more accurately than canonical/typical ($p < 0.001$), typical/ambiguous

($p = 0.006$), and ambiguous/atypical ($p < 0.001$) pairs. None of the other comparisons within each stimulus type were significantly different in the Tukey tests. These results, shown in Figure 4a, indicate that listeners were most accurate at discriminating between pairs on the /ʃ/ side of the continuum.

To examine listeners' response bias across the continuum pairs, we computed and analyzed c with pair as a within-listener variable. The analysis for c revealed a main effect of pair [$F(3, 99) = 18.23$, $p < 0.001$]. As shown in Figure 4b, listeners showed a more numerically gradual change in bias across the pairs. Post hoc Tukey tests indicate that listeners were more biased to respond "same" on the far word side of the continuum. The greatest bias towards a "same" response occurred with canonical/typical pairs, as compared to typical/ambiguous ($p = 0.01$), ambiguous/atypical ($p = 0.05$), and atypical/remapped ($p < 0.001$) pairs. All other comparisons for these stimuli were not significant.

Experiment 2B: Categorical Perception with CVs

The results from experiment 2A provide no support for the maximally ambiguous step being perceptually unique to listeners. However, it is possible that listeners' lexical and linguistic biases clouded their ability to perceive small differences between different items (Ganong, 1980; Pisoni & Tash, 1974). To address this we ran a nearly identical experiment, presenting listeners with CV sequences composed of the /?s/ fricatives and the immediately following vocalic unit to determine whether the maximally ambiguous step stood out as perceptually unique in the absence of lexical bias.[3]

Methods

Participants. Seventeen novel listeners who had not previously participated in either of the previous experiments completed this task. All participants were self-reported native English speakers with no known speech, language, or hearing impairments, and they were recruited from the same population as in experiments 1 and 2A. Participants were compensated with partial course credit for their participation.

Stimuli. The stimuli consisted of 40 CV sequences excised from the 40 words used in experiment 2A. CV sequences were excised using the following boundaries: the onset of the fricative was identified as the onset of aperiodic energy, and the offset of the following vocoid was identified as the cessation of periodic energy. Of the 544 trials presented to participants, 288 trials were of CV sequences excised from fricative medial words.

Procedure. The procedure was identical to that of experiment 2A.

Experiment 2B Results

Analyses were conducted on the CV sequences excised from the fricative medial stimuli. Responses which made more than 2 SDs from the overall mean response time

[3] A reviewer points out that many of these excised CV sequences themselves create words (e.g., /si/ and /ʃi/ excised from the canonical and remapped end points of galaxy create the words see and she). Including fricative-rhotic sequences, which create words like sure and sir, 9 of the 20 CV sequences mapped to real CV words. Being excised out of word-medial positions from multisyllabic words, these items retain any coarticulation from the surrounding context and are considerably shorter in duration (mean: 275 ms, SD: 54 ms) than the real word equivalents of these words (for example, test items sigh and shy were 635 and 622 ms, respectively). This fact and the monotony of the experiment make it less likely that listeners processed these items as fully word-like (Cutler et al., 1987).

DOI: 10.1159/000494929

were removed, constituting 5.2% of the data. The remaining trials had a mean response time of 866 ms (SD = 317). Then d' and c were calculated on these trials. For the sensitivity analysis, d' values were analyzed in a repeated measures ANOVA with pair repeated across listeners. The analysis showed a main effect of pair [$F(3, 78) = 35.25$, $p < 0.001$]. The Tukey HSD test was used to compare sensitivity between stimulus pairs. Like in experiment 2A, listeners discriminated the atypical/remapped pairs more accurately than: canonical/typical ($p = 0.001$), typical/ambiguous ($p < 0.001$), and ambiguous/atypical ($p < 0.001$) pairs. None of the other comparisons within each stimulus type were significantly different from each other. Like with the whole-word stimuli, listeners were most accurate on the /ʃ/ side of the continuum, as shown in Figure 5a.

Listeners' response biases were analyzed with c as the dependent measure and stimulus pair as a within-listener independent variable. This analysis returned a main effect of pair [$F(3, 78) = 15.97$, $p < 0.001$]. As can be seen in Figure 5b, listeners were least biased towards a same response with the atypical/remapped pair compared to ambiguous/atypical ($p = 0.01$), typical/ambiguous ($p = 0.004$), and canonical/typical ($p = 0.01$). None of the other comparisons were significant.

To assess whether sensitivity or bias were related to response time, we compared a listener's mean response time for each pair to their d' and c values. There was a tendency for slower response times to be associated with greater perceptual sensitivity [$r(106) = 0.19$, $p = 0.05$], and there was no relationship between bias and response time [$r(106) = -0.03$, $p = 0.72$].

Discussion of Experiments 2A and 2B

To assess whether the maximally ambiguous items from experiment 1 were perceptually distinct we conducted two AX discrimination tasks with the whole words (experiment 2A) and excised fricative vowel sequences (experiment 2B) from items with canonical, typical, ambiguous, atypical, or remapped endorsement rates. Perceptual sensitivity was quantified using d', and response bias was evaluated using c.

Listeners' perceptual sensitivity with these stimuli peaked with the atypical/remapped pair, indicating that it was not until the fricative was fully remapped to /ʃ/ that listeners showed a clear perceptual boundary. While sensitivity was unaffected by stimulus type, listener response bias was affected. An interaction between stimulus type and pair indicated that when presented with stimuli as whole words, listeners were maximally biased to respond "same" for the canonical/typical pairs, indicating a "same" bias on the side of the continuum that is more word-like. Listeners who did the discrimination task with CVs excised from the word items showed the opposite pattern; when presented with excised CV sequences, listeners' bias to respond "same" remained stable across the first three continuum pairs and decreased with the atypical/remapped pair on the /ʃ/ side of the continua. Together, these results indicate that the maximally ambiguous fricatives are not perceptually unique in either a sensitivity or bias analysis. While not a corroboration of these results, note that, as reported in Table 1 in the Appendix, the average step selected for the ambiguous items is closer to the average step for the typical and atypical items than the typical and atypical items are to the canonical and remapped items, respectively. This indicates that the end points of the continua were more perceptually unique in the categorization pretest as well.

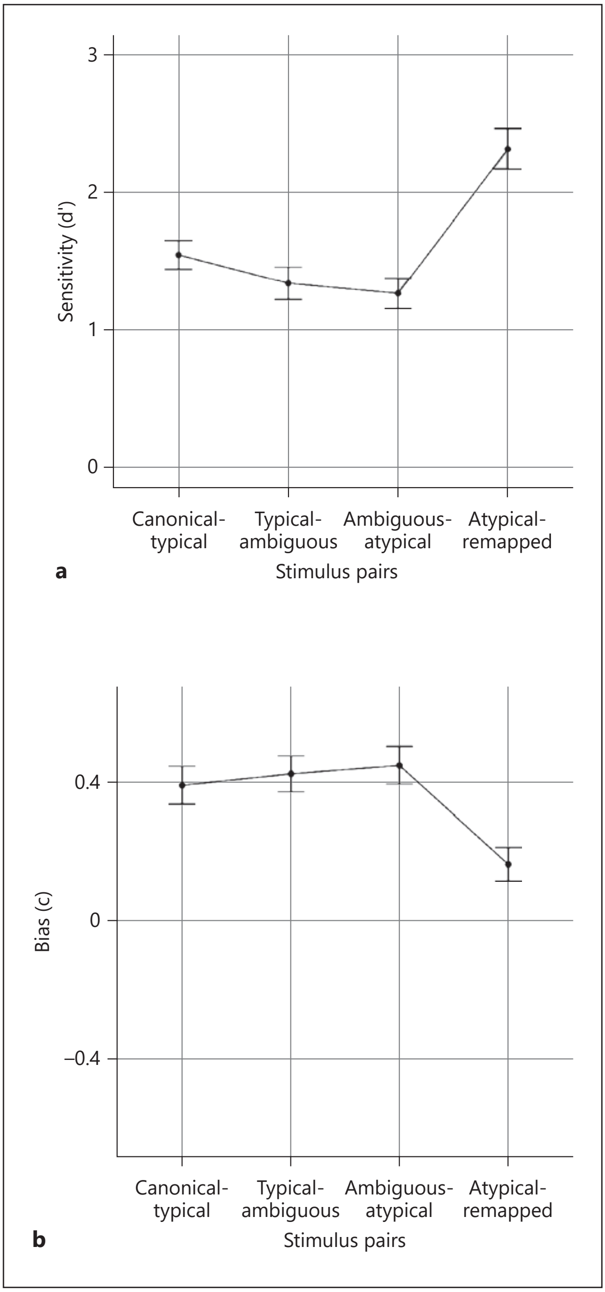

Fig. 5. Experiment 2B (CVs). **a** d', a measure of sensitivity, on the *y* axis for each stimulus pair type (*x* axis). Higher values indicate higher perceptual sensitivity. **b** c, a measure of response bias, on the *y* axis for stimulus pairs (*x* axis). Error bars show standard errors.

Discussion

Lexically guided perceptual learning in speech hinges on at least two fundamental steps. Crucially, (i) a listener must be sensitive, on some level, to the phonetic detail of a novel pronunciation in order to update a phonetic category with the experienced acoustic-phonetic information, and (ii) the novel pronunciation has to be apprehended in an appropriate context (e.g., a recognizable lexical item).

To guide listeners' interpretation of the signal, listeners were presented with noncanonical /s/ realizations in lexical items that *should* contain /s/ and do not have a direct /ʃ/ competitor. This lack of competition should make the interpretation of the items as a word containing /s/ more likely (relevant to point ii above). Listeners were exposed to /s/ words that varied in the typicality of the /s/ based on a previous group's word endorsement rates for /s/ word (e.g., *carousel,* /kɛɹəsɛɫ/) to nonword (e.g., *caroushel,* /kɛɹəʃɛɫ/) continua. Listeners in the typical condition heard /s/ words with /s/ variants that had been correctly identified as the /s/ word only 70% of the time. Those in the ambiguous condition were presented with /s/ words that had been identified correctly 50% of the time. Listeners in the atypical condition heard an /s/ sound that was highly unusual and had been identified as a word containing /s/ in only 30% of the time in the pretest. The remapped condition presented a group of listeners with the /s/ words with pronunciation variants that were fully /ʃ/-like (closest to 0% identified as a word).

Listeners in the lexically guided perceptual learning task heard these /s/ items at levels of typicality that corresponded to their condition assignment alongside filler word items, control /ʃ/ words with a typically produced /ʃ/ (e.g., *shovel*), and nonword items. In the lexical decision task, listeners exposed to the typical (though still not canonical) /s/ items did not categorize these items as words at significantly lower rates than the filler word items. Listeners who heard ambiguous, atypical, or remapped /s/ items did identify these items as words at lower rates than the filler words. As the /s/ items became less canonical, the likelihood of listeners accepting these items as words decreased. The literature on lexically guided perceptual learning clearly shows that a lexical frame is necessary to provide the context for updating a phonetic category (Norris et al., 2003), and some work has shown that listeners need to actively identify the items as words in order to show perceptual learning (Scharenborg, Weber, & Janse, 2015).

Given this, one might expect that listeners who heard the typical /s/ words would show the most perceptual learning in the test phase, as they show the highest rates of word endorsement in the exposure phase. To the contrary, these listeners show the lowest rates of learning. We must acknowledge, however, that the magnitude of learning may be scaled to the amount of what is to be learned. Listeners who heard the items with the typical /s/ pronunciations may have learned to accept slightly more /ʃ/ as /s/, but with a smaller effect size, and thus we would have needed a larger sample to detect it. Additionally, it may be that listeners in this condition have internally restructured their /s/ category in a way not measurable by our categorization test, but which may have been evident in a more sensitive task, such as a category goodness task (e.g., Xie, Theodore, & Myers, 2017).

The argument that when the target is more atypical, it requires greater learning cannot explain all of our results, however, or else the remapped items would have demonstrated the highest degree of learning. Thus, our results suggest that the difference between what is to be learned and what has been perceived cannot be too great; the deviance of an item from its canonical form needs to be in the "Goldilocks" zone for learning to occur. Listeners rejected a perceptual readjustment of /s/ to include more /ʃ/ when the /s/ in the stimuli was the most /ʃ/-like to begin with. These results are in line with prior findings that the speech perception system balances flexibility and stability; listeners in the atypical and remapped conditions showed conservative, stable behavior and were less willing to retune their category boundary in response to stimuli which were clearly outside of the norm (cf. Kraljic et al., 2008b). These findings are reminiscent of Sohoglu and Davis (2016), who found that listeners learned best from partially intelligible (6-channel speech) than from unintelligible (1-channel) or highly intelligible (24-channel) speech. It should be noted that learning of complete remappings has been found in several studies (Sumner, 2011; Weatherholtz, 2015) for different contrasts (bilabial stops; vowels) in different paradigms (exposure to the deviant pronunciations via accent rating; narrative exposure), indicating that future work should tease apart variables responsible for differences across studies. On the whole, our results contribute to our understanding of the limits of perceptual learning in relation to various degrees of phonetic variation, and this is important for our ultimate understanding of perceptual learning in real-life speech situations, which are highly variable and may involve many noncanonical pronunciations. Perceptual learning is often heralded as a primary way that listeners adapt to unfamiliarly accented speakers. However, accented speech presents listeners with a range of speech sounds from more to less typical, and our findings indicate that perceptual learning is most evident when listeners encounter perfectly ambiguous sounds. This suggests that perceptual learning may not be the only or best mechanism that listeners use for adapting to accented speakers.

In order to ensure that listeners were sensitive to the phonetic variation in the critical items, we used AX discrimination tasks with the words used in the perceptual learning tasks and sequences of the /s/ items and the right-adjacent sonorant. Listeners were most sensitive to the phonetic variation on the /ʃ/ side of the continuum; the most ambiguous items, which listeners learned the most from, were not perceptually unique. This indicates that listeners are not more phonetically sensitive to the ambiguous step and suggests that listeners' ability to learn from maximally ambiguous items may stem from a post-perceptual assessment of exemplar quality. In the process of weighing the phonetic and contextual evidence for a linguistic decision, listeners may be assessing whether an item merits inclusion in updating a category or whether that instance should be discarded. The perceptual learning discard pile may be, like the discard pile proposed for sound change (Hay et al., 2015; Sóskuthy, 2015; Wedel, 2012), independent from the actual recognition of the item. In this experiment, listeners presented with the typical tokens endorsed /s/ items at higher rates than other token types in the exposure phase, and they showed the smallest evidence of perceptual learning. Such findings are consistent with previous work suggesting that listeners may rely less on top-down information when their certainty about bottom-up information in the signal is highest (e.g., Ganong, 1980; Kuperberg & Jaeger, 2016; Pitt & Samuel, 1993; Vaughn & Kendall, 2018; Warren, 1970). Assuming a Merge-B style mechanism where streams of information are considered and evaluated at a decision stage (Norris et al., 2000, 2016), when bottom-up information is most uncertain (in this case, in the ambiguous, and to a lesser extent the atypical, condition), listeners weight top-down information more heavily (in this case, their decision to classify an ambiguous item as a word). Thus, they may be more likely to demonstrate perceptual retuning based on that top-down information. Similarly, listeners with a high degree of bottom-up certainty about the lexicality of a token (in this case, in the typical and remapped conditions) may be less inclined to recalibrate their category boundary so it is in line with that top-down information. These results suggest that listeners juggle and consider multiple streams of information in the process of retuning phonetic categories; neither bottom-up nor top-down information is the single determining factor in the retuning of phonetic categories. Moreover, post-perceptual evaluation appears as necessary for the category adjustments elicited in the lab through perceptual learning paradigms, as they are in the category adjustments made over longer time scales in community-level sound changes.

Conclusion

These data illustrate that perceptual learning is indeed enhanced with maximally ambiguous stimuli. Excessively atypical pronunciations show attenuated perceptual learning, while more typical (yet still not canonical) pronunciations show no evidence for perceptual learning. AX discrimination tasks illustrate that the maximally ambiguous stimuli are not perceptually unique. Together, these results suggest that perceptual learning relies on an interplay between confidence in phonetic and lexical predictions and category typicality of the acoustic signal. A conservative post-perceptual goodness-of-fit assessment appears crucial to the process. While there is evidence to suggest that perceptual learning is a low-level perceptual process (Clarke-Davidson et al., 2008), these results suggest that post-perceptual processes likely also factor into the perceptual learning mechanism at large (Norris et al., 2000).

Acknowledgments

Thanks to Graham Haber, Jobie Hui, Michelle Chan, and Zoe Lawler for their assistance in creating stimuli and collecting data for these experiments. This project was partially supported by an award from SSHRC (#F16-04616).

Statement of Ethics

The study protocol was approved by the Behavioural Ethics Review Board at the University of British Columbia. Participants provided oral informed consent prior to completing the research tasks.

Disclosure Statement

The authors declare no financial interest or nonfinancial interest in the subject matter of this research.

Appendix

Table 1

Language Background Questionnaire

Please answer the questions below to the best of your ability. Please ask the experimenter if you have any questions or concerns.

1. Are you a native speaker of English? In this case, “native” refers to your first language. Yes/No

2. If English is not your native language, is it your dominant language? Yes/No

3. If English is not your native language, what is/are your native language(s)?

4. Regardless of whether English is your native or dominant language, what variety of English do you speak? Please specify a dialect (e.g., Newfoundland English, Southern US English, etc.) if you would like.
___American English
___Australian English
___British English
___Canadian English
___Indian English
___Irish English
___Hong Kong English
___Jamaican English
___New Zealand English
___Scottish English
___Singaporean English
___South African English
___Other. Please, specify:

5. What gender do you identify as? _______________

6. What is your racial or ethnic heritage? Check all that apply
___First Nations
___Asian
___Pacific Islander
___Black
___White
___Hispanic
___South Asian
___Other. Please, specify: _________

7. What is your age? _____

8. Are you right-handed or left-handed? ____

9. What cities or towns have you lived in? Beginning with the place where you were born, please list each town or city (and country, if appropriate) you have lived in for 6 months or more.

10. What is your proficiency in English?
(1) not at all, (2) poorly, (3) fairly well, (4) fluently.
Reading ___
Writing ____
Speaking ___
Listening ___

11. At what age did you start learning English? _____

12. Do you have knowledge of any languages other than English? This can include both languages you speak natively and ones you have learned in educational settings. Yes/No

13. What other languages do you have knowledge of? Please include both languages you speak natively and ones you have learned in educational settings. When did you start learning this language? How would you rate your proficiency in reading, writing, speaking, and understanding it? (1) Not at all, (2) poorly, (3) fairly well, (4) fluently

14. Which language(s) do you most commonly speak:
At home?
At work?
At school?
With friends?
With parents?
With grandparents?

15. Do you have any speech or hearing disorders? If "yes", please specify:

16. Where were your caretakers born and raised?

17. What are your caretakers' first languages?

18. What is the highest educational degree you have earned (or are in the process of earning)?

19. What did you think the experiment was about? (Optional)

References

Allen, J. S., & Miller, J. L. (2004). Listener sensitivity to individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America, 115*(6), 3171–3183.

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America, 133*(3), EL174–EL180. doi: 10.1121/1.4789864

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. doi:10.18637/jss.v067.i01

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America, 121*(4), 2339–2349. doi:10.1121/1.2642103

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*(2), 707–729. doi:10.1016/j.cognition.2007.04.005

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181–204. doi:10.1017/S0140525X12000477

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America, 116*(6), 3647–3658. doi:10.1121/1.1815131

Clarke-Davidson, C. M., Luce, P. A., & Sawusch, J. R. (2008). Does perceptual learning in speech reflect changes in phonetic category representation or decision bias? *Attention, Perception & Psychophysics, 70*(4), 604–618. doi:10.3758/PP.70.4.604

Cooper, A., & Bradlow, A. R. (2016). Linguistically guided adaptation to foreign-accented speech. *The Journal of the Acoustical Society of America, 140*(5), EL378–EL384. doi:10.1121/1.4966585

Cutler, A. (2012). *Native Listening: Language Experience and the Recognition of Spoken Words*. Cambridge: MIT Press.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology, 19*(2), 141–177. doi:10.1016/0010-0285(87)90010-7

Ganong, W. F., III. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology. Human Perception and Performance, 6*(1), 110–125. doi:10.1037/0096-1523.6.1.110

Hay, J. B., Pierrehumbert, J. B., Walker, A. J., & LaShell, P. (2015). Tracking word frequency effects through 130 years of sound change. *Cognition, 139*, 83–91. doi:10.1016/j.cognition.2015.02.012

Holt, R. F., & Bent, T. (2017). Children's use of semantic context in perception of foreign-accented speech. *Journal of Speech, Language, and Hearing Research: JSLHR, 60*(1), 223–230. doi:10.1044/2016_JSLHR-H-16-0014

Jesse, A., & McQueen, J. M. (2011). Positional effects in the lexical retuning of speech perception. *Psychonomic Bulletin & Review, 18*(5), 943–950. doi:10.3758/s13423-011-0129-2

Kawahara, H, Morise, M, Takahashi, T, Nisimura, R, Irino, T, & Banno, H (2008, March). TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. *ICASSP. 2008 IEEE International Conference on Acoustics, Speech and Signal Processing*; 2008 March 31 to April 4; Las Vegas (NE), pp. 3933-3936.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, 122*(2), 148–203. doi:10.1037/a0038695

Knoblauch, K. (2014). psyphy: Functions for analyzing psychophysical data in R. R package version 0.1-9. Retrieved from https://CRAN.R-project.org/package=psyphy

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology, 51*(2), 141–178. doi:10.1016/j.cogpsych.2005.05.001

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review, 13*(2), 262–268. doi:10.3758/BF03193841

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language, 56*(1), 1–15. doi.org/10.1016/j.jml.2006.07.010

Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008a). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition, 107*(1), 54–81. doi:10.1016/j.cognition.2007.07.013

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008b). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science, 19*(4), 332–338. doi:10.1111/j.1467-9280.2008.02090.x

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*(1), 32–59. doi:10.1080/23273798.2015.1102299

Lenth, R. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software, 69*(1), 1–33. doi:10.18637/jss.v069.i01

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358–368. doi:10.1037/h0044417

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. New York: Psychology Press.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science, 32*(3), 543–562. doi:10.1080/03640210802035357

McAuliffe, M. (2015). *Attention and salience in lexically guided perceptual learning* (University of British Columbia doctoral dissertation). cIRcle: UBC's Digital Repository: Electronic Theses and Dissertations (ETDs) 2008+. Retrieved from: http://hdl.handle.net/2429/54152

McAuliffe, M., & Babel, M. (2016). Stimulus-directed attention attenuates lexically-guided perceptual learning. *The Journal of the Acoustical Society of America, 140*(3), 1727–1738. doi:10.1121/1.4962529

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1–86. doi:10.1016/0010-0285(86)90015-0

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences, 10*(8), 363–369. do:10.1016/j.tics.2006.06.007

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*(3), 299–325. doi:10.1017/S0140525X00003241

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

Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience, 31*(1), 4–18. doi:10.1080/23273798.2015.1081703

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Attention, Perception & Psychophysics, 15*(2), 285–290. doi:10.3758/BF03213946

Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology. Human Perception and Performance, 19*(4), 699–725. doi:10.1037/0096-1523.19.4.699

Psychology Software Tools Inc. (2012). *E-Prime 2.0* [computer software]. Retrieved from http://www.pstnet.com

R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Reinisch, E., Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology. Human Perception and Performance, 39*(1), 75–86. doi:10.1037/a0027979

Reinisch, E., Wozny, D. R., Mitterer, H., & Holt, L. L. (2014). Phonetic category recalibration: What are the categories? *Journal of Phonetics, 45*, 91–105. doi:10.1016/j.wocn.2014.04.002

Scharenborg, O., & Janse, E. (2013). Comparing lexically guided perceptual learning in younger and older listeners. *Attention, Perception & Psychophysics, 75*(3), 525–536. doi:10.3758/s13414-013-0422-4

Scharenborg, O., Weber, A., & Janse, E. (2015). The role of attentional abilities in lexically guided perceptual learning by older listeners. *Attention, Perception & Psychophysics, 77*(2), 493–507. doi:10.3758/s13414-014-0792-2

Sohoglu, E., & Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *Proceedings of the National Academy of Sciences of the United States of America, 113*(12), E1747–E1756. doi:10.1073/pnas.1523266113

Sóskuthy, M. (2015). Understanding change through stability: A computational study of sound change actuation. *Lingua, 163*, 40–60. doi:10.1016/j.lingua.2015.05.010

Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition, 119*(1), 131–136. doi.org/10.1016/j.cognition.2010.10.018

Theodore, R. M., Myers, E. B., & Lomibao, J. A. (2015). Talker-specific influences on phonetic category structure. *The Journal of the Acoustical Society of America, 138*(2), 1068–1078. doi:10.1121/1.4927489

Vaughn, C., & Kendall, T. (2018). Listener sensitivity to probabilistic conditioning of sociolinguistic variables: The case of (ING). *Journal of Memory and Language, 103*, 58–73. doi.org/10.1016/j.jml.2018.07.006

Vroomen, J., van Linden, S., Keetels, M., De Gelder, B., & Bertelson, P. (2004). Selective adaptation and recalibration of auditory speech by lipread information: Dissipation. *Speech Communication, 44*(1), 55–61. doi.org/10.1016/j.specom.2004.03.009

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*(3917), 392–393. doi.org/10.1126/science.167.3917.392

Weatherholtz, K. (2015). *Perceptual learning of systemic cross-category vowel variation* [Unpublished doctoral dissertation]. Columbus, OH: The Ohio State University.

Wedel, A. (2012). Lexical contrast maintenance and the organization of sublexical contrast systems. *Language and Cognition, 4*(4), 319–355. doi:10.1515/langcog-2012-0018

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception & Psychophysics, 75*(3), 537–556. doi:10.3758/s13414-012-0404-y

Xie, X., Theodore, R. M., & Myers, E. B. (2017). More than a boundary shift: Perceptual adaptation to foreign-accented speech reshapes the internal structure of phonetic categories. *Journal of Experimental Psychology. Human Perception and Performance, 43*(1), 206–217. doi:10.1037/xhp0000285

DOI: 10.1159/000494929