Laboratory Phonology
Journal of the Association for Laboratory Phonology

Babel, M., et al. 2021. Asymmetries in perceptual adjustments to non-canonical pronunciations. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 12(1): 19, pp. 1–43. DOI: https://doi.org/10.16995/labphon.6442

Open Library of Humanities

Asymmetries in perceptual adjustments to non-canonical pronunciations

Molly Babel, Department of Linguistics, University of British Columbia, Canada, molly.babel@ubc.ca

Khia A. Johnson, Department of Linguistics, University of British Columbia, Canada

Christina Sen, San Diego State University/University of California, San Diego Joint Doctoral Program in Language and Communicative Disorders, USA

This paper examines two plausible mechanisms supporting sound category adaptation: directional shifts towards the novel pronunciation or a general category relaxation of criteria. Focusing on asymmetries in adaptation to the voicing patterns of English coronal fricatives, we suggest that typology or synchronic experience affect adaptation. A corpus study of coronal fricative substitution patterns confirmed that North American English listeners are more likely to be exposed to devoiced /z/ than voiced /s/. Across two perceptual adaptation experiments, listeners in test conditions heard naturally produced devoiced /z/ or voiced /s/ in critical items within sentences, while control listeners were exposed to identical sentences with canonical pronunciations. Perceptual adaptation was tested via a lexical decision test, with devoiced /z/ or voiced /s/, as well as a novel alveopalatalized pronunciation, to determine whether adaptation was targeted in the direction of the exposed variant or reflected a more general relaxation. Results indicate there was directional and word-specific adaptation for /z/-devoicing with no evidence for generalization. Conversely, there was evidence of /s/-voicing generalizing and eliciting general category relaxation. These results underscore the role of perceptual experiences, and support an evaluation stage in perceptual learning, where listeners assess whether to update a representation.

1. Introduction

Due to the enormous amount of linguistic and social information simultaneously available in the speech stream, speech perception is a challenging task. When considering the variability listeners encounter within and across talkers, the act of understanding spoken language becomes an even more complex enterprise. Both within- and between-talker variation can arise due to differences in physiology (e.g., Weirich & Simpson, 2014; Ladefoged & Broadbent, 1957), articulatory control (e.g., Johnson & Beckman, 1997), emotional state and personality (e.g., Dewaele & Li, 2014), and dialect (e.g., Clopper & Smiljanic, 2015). While this list is not exhaustive, such differences result in variable spectral and temporal information that must be processed on some level by listeners for successful recognition and communication. Despite different sources of variation, phonetic variability is typically highly structured—patterns of variation are not random, and listeners are able to exploit both the surrounding linguistic context and talker-specific social information to categorize, comprehend, and adapt to speech (Holt & Bent, 2017; Bradlow & Bent, 2008).

1.1. Perceptual adjustments to phonemic categories

How do listeners adapt to and accommodate variable pronunciations? Listeners' rapid adaptation abilities have been studied under the umbrellas of perceptual learning or phonemic recalibration—behaviours that allow for listeners' adjustments to phonemic categories. In the lab, adaptive behaviours have been extensively tested using an experimental paradigm termed lexically-guided perceptual learning, which generally functions as follows: Listeners are exposed to novel or phonetically ambiguous pronunciations of a particular sound in lexical contexts that provide the linguistic scaffolding to guide the interpretation of the intended word. Norris, McQueen, and Cutler (2003) introduced the paradigm using Dutch words containing the fricatives /f/ and /s/. If an ambiguous sound—denoted [?sf]—was presented in the context of a word like *witlof* "chicory," listeners expanded their /f/ category to include these pronunciations. Conversely, if the ambiguous [?sf] sound was heard in the context of an /s/ word like *naaldbos* "pine forest," listeners adjusted their /s/ category. The adjusted category boundaries were demonstrated in a categorization task with items on a synthesized [s]-[f] continuum. Listeners who heard the ambiguous sound in /f/ contexts showed learning targeted in a particular direction, expanding their /f/ category to specifically accommodate the ambiguous fricative at the expense of /s/, while those who heard the ambiguous sound in the context of /s/ words expanded their /s/ category at the expense of /f/. Crucially, listeners did not make perceptual adjustments when the phonetically ambiguous sound was heard in the context of nonwords, as there was no lexical content listeners could use to make the connection between [?sf] and /s/ or /f/. These adjustments reflect actual retuning of linguistic categories and not exclusively decision criterion shifts (Clarke-Davidson, Luce, & Sawusch, 2008).

There seem to be some limitations to perceptual learning, and further identifying and articulating (some of) those limitations is a goal of the current study. Adaptation seems to be limited, for example, if the phonetic ambiguity of the crucial phoneme is extreme, as is the case of heavily non-native-accented speech, where listeners also do not show adaptation (Witteman, Weber, & McQueen, 2013). Likewise, if listeners do not accurately recognize the words with ambiguous pronunciations as words, they are less likely to learn the novel pronunciations (Scharenborg & Janse, 2013). Adaptation appears limited in its generalization to voices and across-languages, only applying when the voices or cross-linguistic patterns share sufficient acoustic-phonetic or perceptual similarity (Eisner & McQueen, 2005; Kraljic & Samuel, 2005; Reinisch & Holt, 2014; Mitterer & Reinisch, 2017, Reinisch, Weber, & Mitterer, 2013). Together, adaptation requires lexical or other top-down support, in addition to signal-based or bottom-up similarity.

Taking advantage of the natural variation in stop realization within and across languages, scholars have identified some additional limits to the generalization of what is learned. For example, native English listeners naïve to Dutch-accented English not only show adaptation to the devoiced final-stops in Dutch-accented English, but can generalize voiced stop devoicing to initial position as well, at least when not presented with counterevidence to initial devoicing patterns (Eisner, Melinger, & Weber, 2013). Counterevidence inhibited this learning: Listeners do not generalize to initial position when listeners are exposed to items with initial voiced stops, which do not undergo devoicing in Dutch-accented English. When presented with a native English speaker who *only* exhibited word-final stop devoicing (and not the more global and systematic effects of a non-native accent), listeners showed learning of the pattern, but did not generalize it to word-initial positions, as they had with the Dutch-accented English speaker. These results crucially suggest that the overall accent of a talker informs the generalization strategy. Using Mandarin-accented English, Xie, Theodore, and Myers (2017) demonstrated that adaptation to final /d/-devoicing results in both a recalibration of category boundaries and a reorganization of the category-internal structure. Adaptation was evident in a cross-modal priming task, where listeners showed higher levels of lexical activation for devoiced /d/-final words without negative effect on lexical activation for /t/-final words. These adjustments to Mandarin-accented final-stop devoicing (as observed in the priming data) are different from those results observed by McQueen et al. (2006), who found that listeners completely remapped categories (e.g., negative consequences for doo*f* when training was words like doo*s*).

Adaptation has also been found for vowels (Maye, Aslin, & Tanenhaus, 2008; Witteman et al., 2013; Weatherholtz, 2015; Babel, Senior, & Bishop, 2019). The novel pronunciations used in the work on vowel adaptation implement changes in vowel pronunciation that use or model natural accents, which often fully and categorically replace one vowel quality with that of another vowel category (e.g., *tooth* pronounced as [tʊθ], not the canonical [tuθ] for North American dialects

of English). Some of these vowel studies have tackled the mechanisms of perceptual adaptation by assessing whether exposure to a novel accent adjusts criteria for word identification. For example, Maye et al. (2008) crucially controlled for whether learning was specifically tuned in the targeted direction of the novel pronunciations or whether exposure to a unique accent led to a more general relaxing of criteria for word identification. They accomplished this by testing word endorsement rates of items like [witʃ] when listeners had been exposed to [wɛtʃ] for what is typically pronounced as [wɪtʃ] "witch" in North American English. Their results indicate that listeners' adjustments to the novel pronunciations were targeted in the specific direction of listeners' exposure, as opposed to a more general relaxation of criteria for what constitutes an acceptable realization of a vowel. Conversely, several experiments in Weatherholtz (2015) do provide evidence of relaxing criteria for vowel categories: Listeners exposed to a back vowel raising chain shift learned an exposure-specific pattern, whereas listeners exposed to a back vowel lowering chain shift showed evidence of category relaxation, accepting items that showed a lowering or raising pattern. While the mechanism behind this exposure-specific category relaxation is not understood, it is not necessarily a surprising outcome, as small distortions to a word's consonant or vowel pronunciations do not impede successful recognition (Connine, Blasko, & Titone, 1993; Connine, Titone, Deelman, & Blasko 1997; Andruski, Blumstein, & Burton, 1994; Brouwer, Mitterer, & Huettig, 2012). Successful recognition of a word despite a deviant pronunciation, however, may or may not have an effect on the subsequent encoding of that item (Todd, Pierrehumbert, & Hay, 2019).

The specificity of the category—for example whether what is learned is tied to a phoneme-level, an allophonic-level, or an even more surface-level of representation—appears to be sensitive to the exposure conditions (Eisner et al., 2013; Xie et al., 2017; Mitterer & Reinisch, 2017). What remains unclear about the processes involved is whether a more general category relaxation may be a distinct mechanism that also supports perceptual learning, one that complements a more directionally targeted phonemic recalibration mechanism. Zheng and Samuel (2019) present evidence that phonemic recalibration, accent adaptation, and selective adaptation are all distinct mechanisms. Zheng and Samuel's approach to accent adaptation leverages a non-native accent that has specific shifts in pronunciation that lead to what have been termed 'bad maps' (Sumner, 2011). These 'bad maps' are categorical mismatches in pronunciation patterns, as in the case of Mandarin-accented English /θ/ pronounced as [s]. In such cases, there is deviation from a listener's local accent throughout the signal, though crucially, the acoustics-to-phoneme mapping is more disrupted for particular sound categories. Focusing on such categories, Zheng and Samuel found that listeners exposed to Mandarin-accented English sentences were subsequently more likely to call nonwords words (e.g., calling a nonword like [sæŋkfʊɬ] a word), which indicates general criteria relaxation. Support for a general category relaxation mechanism can also be found in the literature on accent accommodation, both in development (Schmale, Cristia, & Seidl, 2012;

Schmale, Seidl, & Cristia, 2015; White & Aslin, 2011), and with adult listeners (Baese-Berk, Bradlow, & Wright, 2013). Inconclusive support for expansion mechanisms also comes from computational modelling. Hitzcenko and Feldman (2016) implement Kleinschmidt and Jaeger's (2015) ideal adaptor framework in an effort to computationally model the Maye et al. (2008) data and assess whether the mechanism(s) at play in the listeners in Maye and colleagues' study is one of (i) expanding, (ii) shifting, (iii) shifting and expanding, or (iv) remapping. The relevant parameters are the category means and covariance and the model's confidence in those values (e.g., high or low). The four implementations hit-and-miss the behavioural data to different degrees, but overall, the expand, shift, and expand and shift models all provide decent fits with models that include expansion perhaps aligning somewhat better to Maye and colleagues' data.

It is important to broach the question of how degree of deviation might limit or constrain the ease with which listeners adapt—or even whether they adapt at all. There are limitations on the acoustic similarity of what kind of substitution is acceptable while still allowing for perceptual adaptation (e.g., Witteman et al., 2013; Babel, McAuliffe, Norton, Senior, & Vaughn, 2019), as is the case with other behaviours that showcase perceptual flexibility (e.g., the phoneme restoration effect; Samuel, 1981). Though note, it is possible that simply increasing the amount of exposure may provide a boost in learning for highly divergent pronunciations. There also seem to be other constraints on lexically-guided perceptual learning with such constraints potentially stemming from a number of underlying causes. For example, Kraljic, Brennan, and Samuel (2008a) illustrate that listeners do not adapt to more [ʃ]-like /s/ sounds in the context of *str*-clusters. In another example, Kraljic, Samuel, and Brennan (2008b) find that listeners do not shift phonetic categories if ambiguous productions are accompanied by video clips suggesting the speaker had a pen in her mouth during production. They reason that if phonetic variation can be attributed to an entity beyond the speaker (i.e., phonologically conditioned dialect patterns or pen-in-mouth disruptions), listeners will fail to learn the pattern (for a similar account, see Cutler, 2012). Alternatively, if there is not sufficient lexical activation to support the mapping of the novel pronunciation to a phonemic category, boundary adjustments do not occur (Jesse & McQueen, 2011; McAuliffe & Babel, 2016). Thus, the amount or nature of phonetic variation, quantity of exposed items, talker-related attribution (i.e., causal inference), and lack of top-down support may all constrain listeners' perceptual adjustments.

Lastly, we can also ask whether all speech sounds are updated equivalently in perceptual learning paradigms. The answer to this question appears to be no. Zhang and Samuel (2014) report on a lexically-guided perceptual learning study where listeners were exposed to an ambiguous /s/-/f/ fricative in the context of /s/ or /f/ words in highly predictable sentences. In addition to showing perceptual learning in clear speech, conversational speech, and under two cognitive load manipulations, Zhang and Samuel found an asymmetry across the /s/ and /f/ conditions. Compared to a control condition, listeners exposed to the ambiguous fricative in

the context of /s/ words learned the novel pronunciation, generalizing their updated category distribution to the categorization test stimuli. Listeners exposed to the same ambiguous fricative in the context of /f/, however, did not. Zhang and Samuel's explanation for this hinges on the less robust spectral cues for /f/ compared to /s/, suggesting that listeners may simply be less sensitive to phonetic variation in non-sibilant fricatives. Note, however, that several other studies have found evidence for the perceptual retuning of /f/ in the face of /s/ (e.g., Schuhmann, 2014; Bruggeman & Cutler, 2020; Reinisch & Holt, 2014; and for retuning of /f/ compared to a control group, see Chan, Johnson, & Babel, 2020). Drozdova, Van Hout, and Scharenborg (2016) also uncover an asymmetry in the perceptual retuning of /l/ and /ɹ/ in British English with British English and Dutch (L2 English) listeners. Dutch listeners showed more recalibration for English /ɹ/ than the British English listeners, which Drozdova and colleagues link to the wide range of rhotic variation in (L1) Dutch, reasoning that this larger phonetic range allows listeners to tap into pre-existing variation they have been exposed to, and thus adapt with ease.

This overview of the theoretical and empirical literature on adaptation leads to the objectives for the current study, which are multifaceted. We seek to (i) directly test whether asymmetries in sound patterns affect perceptual adjustments, and (ii) assess whether the adaptation mechanism is targeted in the direction of the exposed pronunciation or reflects a general relaxing of criteria or both. We investigate a potential asymmetry in perceptual learning by testing the learnability of changes in the voiced and voiceless English coronal fricatives: /z/ and /s/. Specifically, we assess whether listeners perform differently when exposed to naturally-produced devoiced /z/ (perceived as [s]) and voiced /s/ (perceived as [z]). This question has roots in questions about learnability in phonology (e.g., the presence of channel biases in phonology; e.g., Moreton & Pater, 2012, Martin & Peperkamp, 2020), and in whether there are boundary conditions on retuning for novel pronunciations in lexically-guided perceptual learning.

We use a lexical decision task as a test of perceptual learning, and quantify differences in word endorsement rates across experimental and control groups for items with a voicing change (i.e., /z/-devoicing or /s/-voicing). This paradigm allows us to quantify both learning specificity and the generalization of the learned (or not) voicing change in novel words not presented during the exposure phase. As opposed to the two-alternative forced choice categorization task often used in phoneme recalibration studies, our use of a lexical decision paradigm to quantify learning has the distinct advantage of allowing us to test whether phonemic adjustments are directional, showing increased word endorsement only in the direction of the exposed change, or whether the adjustments involve general relaxation of category boundaries. To assess whether exposure to devoiced /z/ or voiced /s/ results in directional adaptation or general relaxation, we tested word endorsement rates of both [s] and [ʒ] pronunciations in canonical /z/ words. Endorsement of [ʒ] pronunciations in addition to [s] pronunciations would be evidence in support of a general category relaxation mechanism. Likewise, [ʃ] pronunciations in canonical /s/ words are used to

test category relaxation in response to exposure to /s/-voicing. Both the voicing change shifts and the place of articulation shifts are pronunciation changes of one feature edit distance from the canonical pronunciation, and as a result are close to the canonical productions. Following our corpus study in Section 2, we provide much more finely honed predictions of listener behaviour.

1.2. Coronal fricatives in English

English coronal fricatives offer a nice test case because they are clearly asymmetrical in their cross-linguistic distribution and within-English voicing patterns. English coronal fricatives also have attested alveopalatalized variants in casual speech, though these are rare (at least in the Buckeye Corpus [Pitt et al., 2007]; see corpus study in Section 2). Aerodynamic constraints make voiced fricatives less common cross-linguistically compared to voiceless fricatives (Ohala, 1983; Moran & McCloy, 2019). Devoicing patterns are logically assumed to result from the difficulty in maintaining the pressure gradient between the oral and subglottal cavities required to maintain voiced turbulent airstreams. The difficulty of maintaining voicing during fricative production explains why voiceless fricatives are typically longer in duration than their voiced counterparts. These aerodynamic constraints result in devoicing patterns being described as more phonetically natural, particularly for obstruents in utterance-final position (Ladefoged & Johnson, 2011). Also, phonological rules of English fricative final-devoicing are more common than the reverse—within and across varieties of English—though there is the small set of voiceless fricative final words which can voice when coupled with the plural morpheme (e.g., *houses* can be pronounced as [haʊsəz] or [haʊzəz]; MacKenzie, 2018). In a study of such words, MacKenzie (2018) finds that /s/-voicing rates are dropping precipitously in apparent time, further reiterating the asymmetry between /s/-voicing and /z/-devoicing. Overall, both cross-linguistic and English-specific patterns favour the learnability of /z/-devoicing and the reduced learnability of /s/-voicing, offering a clear test for whether asymmetries in phonemic adjustments are biased by experience.

While there is at least one dialect of English that engages in word-initial fricative voicing (West Country English; Wells, 1982), voicing of phonologically voiceless fricatives is rare both within and across English dialects. Conversely, word-final fricative devoicing is widely documented in varieties of American and British English (e.g., Docherty, 1992; Veatch, 1989; Stevens, Blumstein, Glicksman, Burton, & Kurowski, 1992). Similarly, many languages have phonological patterns where obstruent voicing contrasts are described as neutralizing in word-final positions. Phonetic studies illustrate that these neutralizations are often incomplete, with subtle acoustic-phonetic cues differentiating the underlying obstruent categories (e.g., Warner, Good, Jongman, & Sereno, 2006; Warner, Jongman, Sereno, & Kemps, 2004; and references therein). Kleber, John, and Harrington (2010) demonstrate that naïve listeners are perceptually sensitive to incomplete neutralizations, and use subtle acoustic distinctions in recognition (although far from categorically). In an artificial language learning study, Myers and Padgett (2014) demonstrated

that English-speaking listeners more readily learned an utterance-final fricative devoicing pattern compared to a voicing pattern, though participants learned and generalized both patterns to word-final utterance-medial environments. To maintain our focus on phoneme-specific adjustments, we focus on adaptation to fricative voicing changes in word-medial position where the pattern is not tied to morphological alternations, but rather to pronunciation variation attributable to a speaker or dialect.

Smith (1997) examined the propensity to devoice /z/ in English in depth for a small set of speakers (n = 4) in utterances where the target fricatives occurred in word-initial, word-medial, word-final, and utterance-final position. As a testament to the multidimensional nature of voicing contrasts, Smith measured vocal fold vibration with an electroglottograph (EGG), duration of the fricatives and the preceding vowels, and oral airflow. Smith found that her four talkers varied considerably in terms of the degree of /z/-devoicing—one speaker typically produced vocal fold vibration throughout /z/, one speaker produced most of her /z/ tokens with no vocal fold vibration, and two speakers typically vibrated their vocal folds for some portion of the fricative. Three of the four speakers produced a clear duration difference between /z/ and /s/, regardless of whether the /z/ was realized as voiced, partially devoiced, or fully devoiced. Speakers produced longer vowels before phonologically voiced /z/, although three of the four speakers did not produce longer vowels before phonologically voiced /z/ when produced with full voicing. Generally, across all types of /z/ realizations, speakers produced /z/ with lower mean and maximum airflow. Together, these results suggest that the difference between /s/ and /z/ extends well beyond the categorical distinction of whether or not vocal folds are vibrating. Smith summarizes her results as providing evidence that American English speakers can "devoice /z/ in almost any environment" (Smith, 1997: 498).

Thus, not only is there typological evidence that voiced fricatives are more likely to devoice than the reverse, but at least in lab speech, English /z/ is more likely to devoice than /s/ is to voice. There is no literature, to our knowledge, documenting the probability or magnitude of /s/-voicing. This lack of evidence, however, does not mean /s/-voicing does not exist. Thus, to assess /z/-devoicing and /s/-voicing on a categorical level and to quantify listeners' previous experience with sibilant fricatives and voicing patterns, we consider spontaneous speech, which exhibits large amounts of reduction and pronunciation variation (Johnson, 2004; Pluymaekers, Ernestus, & Baayen, 2005).

2. Experiment 1: Corpus study of /s/ and /z/ realizations in the Buckeye Corpus

We estimate listeners' previous experience with voiced /s/ and devoiced /z/ by comparing expected citation forms to pronounced forms of voiced and voiceless fricatives in the Buckeye Corpus (Pitt et al., 2007). The Buckeye Corpus is a collection of recordings of spontaneous

conversational speech from 40 white native speakers of English from Columbus, Ohio, USA, and has detailed phonetic transcriptions for all recordings. We use this corpus because of its size, the casual nature of the spontaneous speech, and the crucial fact that it is phonetically transcribed. Our approach uses categorical coding of fricative categories in the Buckeye Corpus: We examine categorical changes in the transcription character symbol selected to represent the surface form for what is /s/ or /z/ in the canonical citation forms. This was done using the transcription data in the phonetic and citation tiers of the Buckeye Corpus.

2.1. Methods

2.1.2. Materials

As the goal was to assess the overall probability of categorical voicing changes for all /s/ and /z/, regardless of position in the word (or participation in morphophonological processes), all items with /s/ or /z/ in any position in the expected citation form transcriptions were identified in the Buckeye Corpus. This resulted in 38,934 instances of citation /s/ and 23,076 instances of citation /z/.

2.1.3. Procedure

The surface behaviour of the coronal fricatives was assessed from the phonetic transcripts. An automated procedure separated items into matches or items that required human decisions. Items where the citation form (e.g., 'dictionary' pronunciation) and the phonetically-transcribed pronounced surface form showed perfect alignment (e.g., there were no changes between citation and pronounced forms) were automatically tagged as matches. There were 17,448 perfect matches for /s/ items (44.8%) and 8,831 perfect matches for /z/ items (38.2%). Fricatives were also automatically paired and identified as substitutions in citation and surface forms if the two strings matched in terms of both the number of characters and the identity of all characters *except* the citation form /s/ or /z/. For example, a word like *there's* with the dictionary pronunciation of <dh eh r z> that was transcribed in the Buckeye as <dh eh r zh> (filename s0101b.word, timestamp 44.59) was automatically coded as a [z] to [ʒ] substitution. However, *there's* that was transcribed as <eh r z> (filename: s2402a.word, timestamp: 328.105) was sent to the human decisions list.[1] Just less than 1% of citation /s/ items were automatically identified as substitutions, but 10.3% of /z/ citations were; and all but 1% of those were cases where /z/ devoiced to [s].

The remaining items were manually matched by annotators in a custom-written command-line program using the following process: The citation transcription was presented with the item's pronounced transcription. The fricative of interest was visually flagged by a symbol (>)

[1] The Buckeye corpus uses Arpabet transcriptions and those are used in our reporting from this corpus.

in the citation pronunciation and the characters in the pronounced word were numbered. Using the transcriptions alone (without audio), annotators identified the number in the citation form string associated with the fricative of interest or marked the fricative as deleted from the string.

2.2. Results

The counts for labels that were automatically and manually applied are shown for /s/ and /z/ words in **Tables 1** and **2**. There are many instances of applied categories that have small counts and are likely erroneous labels due to mistakes in the human decision process or overly simple assumptions for our automatic labeling (e.g., /s/ as [ɑ]). What is clear from these numbers is that /s/ is much less likely to surface as [z] than /z/ is to surface as [s]. Over 96% of /s/ words have an [s] as their transcribed category, and only 1% of /s/ words were transcribed as [z]. On the other hand, 77% of /z/ words are labeled with [z], and a total of 18% of words were transcribed with an [s] in place of the citation /z/. These data from spontaneous speech confirm prior laboratory findings: /z/ is much more likely to surface as [s] than /s/ is to surface as [z].

The voicing-matched alveopalatal English fricatives are equally likely to surface: [ʃ] and [ʒ] both occur 1.7% of the time for /s/ and /z/, respectively. This equivalence is useful in our adjudication between directional adaptation and general category relaxation, as both substitutions are attested, but equivalently unlikely.

2.3. Discussion and predictions

The results of this corpus study confirm that, at least for this dialect of North American English, listeners are more likely to be exposed to /z/-devoicing than /s/-voicing in spontaneous speech.[2] The probability of listeners hearing substitutions of [ʒ] for /z/ and [ʃ] for /s/ is low but roughly equivalent. These data allow us to clarify our predictions for how asymmetries in sound patterns will impact adaptation and acceptability of category variation.

We predicted that English listeners would be willing to identify words with /z/-devoicing as words *a priori*, but that exposure to devoiced /z/ in lexical contexts would cause listeners to be even more likely to identify these items as words. We expected that listeners exposed to devoiced /z/ would update their expectations about the talker in the experiment, and generalize to novel devoiced /z/ words, as this talker would be labelled a 'devoicer.' Given that the devoiced /z/ taps into previous experiences with devoiced /z/, we expected that listeners would not generally relax category boundaries for /z/. That is, after being exposed to devoiced /z/, we did not

[2] While we acknowledge that the population sample of the Buckeye Corpus (Midland American English in Columbus, Ohio) is not identical to that of the listener population sample (Canadian English in Vancouver, British Columbia), the fricative voicing patterns are robust to the point that we assume the basic pattern holds for speakers of Canadian English as well given that, to our knowledge, no previous literature suggests major differences in fricative voicing patterns in North American English.

Citation forms with /s/				
IPA	**Arpabet**	**Automatically tagged**	**Manually tagged**	**Total no. of observations**
a	ah	1	3	4
ɑ	aw	0	1	1
ʧ	ch	12	32	44
d	d	0	1	1
ð	dh	2	2	4
ɛ	eh	1	6	7
n̩	en	0	2	2
f	f	0	1	1
h	hh	3	1	4
ɪ	ih	0	4	4
i	iy	0	1	1
ʤ	jh	0	1	1
k	k	1	1	2
m	m	0	2	2
n	n	0	3	3
p	p	0	1	1
ɹ	r	0	3	3
s	s	17448	20060	37508
ʃ	sh	180	469	649
t	t	6	20	26
θ	th	7	11	18
tˀ	tq	0	1	1
ɐ	uh	0	1	1
deletion	xx	0	167	167
z	z	143	320	463
ʒ	zh	0	16	16
TOTALS		17804	21130	38934

Table 1: Counts of substitutions in IPA and Arpabet for underlying /s/ in the Buckeye Corpus from automatic and manual tagging.

expect listeners to change their threshold for /z/, such that [ʒ] pronunciations would become acceptable.

Conversely, we expected listeners to exhibit a different response for /s/, as fully voiced /s/ is incredibly rare. While we anticipated that exposure to voiced /s/ pronunciations would

Citation forms with /z/				
IPA	**Arpabet**	**Automatically tagged**	**Manually tagged**	**Total no. of observations**
æ	ae	0	1	1
a	ah	1	10	11
a͡ɪ	ay	0	3	3
ʧ	ch	2	1	3
d	d	0	11	11
ð	dh	3	0	3
ɾ	dx	2	8	10
ɛ	eh	0	2	2
n̩	en	0	4	4
ɹ̩	er	0	4	4
f	f	0	1	1
h	hh	1	0	1
ɪ	ih	2	12	14
i	iy	0	8	8
ʤ	jh	2	2	4
k	k	0	3	3
l	l	0	5	5
n	n	0	21	21
o͡ʊ	ow	0	2	2
ɹ	r	0	5	5
s	s	2152	2118	4270
ʃ	sh	56	86	142
t	t	0	3	3
θ	th	2	3	5
ʊ	uh	1	2	3
v	v	2	3	5
deletion	xx	0	284	284
z	z	8831	9042	17873
ʒ	zh	156	227	383
TOTALS		11213	11871	23084

Table 2: Counts of substitutions in IPA and Arpabet for underlying /z/ in the Buckeye Corpus from automatic and manual tagging.

increase the identification of these items as words, we expected this pattern to be qualitatively different from the word endorsement rates for the more frequent pattern of /z/-devoicing, as this reflects listeners' experiences with the more probable /z/-devoicing than /s/-voicing; the baseline word endorsement rates for /z/-devoicing are predicted to be much higher than those for the /s/-voicing. Because /s/-voicing is such a rare change in listeners' experiences, the talker may essentially be labelled as atypical and the adaptation mechanism may be different, reflecting the listener's lower degree of certainty that the produced [z] indeed maps onto the /s/ category. Under this general relaxation response scenario, listeners may show increased acceptability of [ʃ] pronunciations as well as to novel words with a [z] pronunciation. Given the rarity of [ʒ] for /z/ and [ʃ] for /s/ in spontaneous speech, we anticipated these items as having lower word endorsement rates than items with the voicing change for listeners in the control groups, who were given no reason to adjust their phoneme boundaries or criteria in the exposure phase. The hypotheses described above are summarized in **Table 3**.

3. Material selection for experiments 2 and 3

Experiments 2 and 3 target different critical sounds—/z/ and /s/—but share a procedure for creating the stimuli. Section 3 outlines this process and the process of winnowing to a final stimuli list. Relevant acoustic characteristics of the stimuli are also described in this section. Sections 4 and 5 report on the experiments that used these stimuli.

3.1. Word and sentence selection

Target lexical items with non-initial /s/ or /z/ were identified. These items contained only a single sibilant fricative. To reduce ambiguity in the lexical frame, these items were confirmed

	Experimental condition of Experiment 2: /z/-devoicing	**Experimental condition of Experiment 3: /s/-voicing**
Hypothesized mechanism	Directional adaptation	General relaxation
Talker behaviour is …	Expected	Unexpected
Baseline word endorsement behaviour	Very high rates for devoiced /z/	Higher than control listeners, but lower than /z/-devoicing
Listener behaviour for novel words (tests generalization)	Yes, listeners form representation of talker as 'devoicer'	Yes, as voicing change falls within general relaxation
Listener behaviour for place change (tests general relaxation)	No, exposure to devoiced /z/ reinforces prior knowledge	Yes, as place change also falls within general relaxation

Table 3: Summary of mechanisms, rationale, and predictions for experimental groups across Experiments 2 and 3.

to not be able to form words if the target fricative was replaced by another fricative (see the list of stimuli in the Appendix). Further, all targets had two to four syllables, and were embedded within moderately predictable sentences containing no sibilant fricatives other than the critical /s/ or /z/. Filler sentences (n = 100) were composed with no sibilant fricatives.

3.2. Audio recordings

An adult female monolingual English speaker produced the sentence and single word materials. These auditory stimuli were recorded using a head-mounted microphone with a SoundDevices USB PreAMP in a sound-attenuated cubicle at a 44.1kHz sampling rate with 16 bit depth. Recorded materials were trimmed of extraneous silence and RMS-amplitude normalized to 70 dB, ensuring no clipped samples. Three versions of each critical /s/ and /z/ sentence were recorded, corresponding to the canonical pronunciation (control), (de)voicing, and alveopalatalized item types. To address the potential for unintended mispronunciations and inconsistencies, the speaker produced several instances of each critical sentence and word. The final recordings were selected based on perceived clarity and consistency by the third author.

3.2.1. Pronunciation accuracy: Transcription

Critical items from the sentences were excised and presented to two trained linguists blind to the purpose of the experiment along with all of the filler single word items. Each linguist independently phonetically transcribed each item, which were presented in a unique random order without word-level labels. This was done to confirm that the critical words with the (de) voicing and alveopalatal pronunciations were categorically perceived as intended. Any items where the transcribers disagreed or where the transcription did not match the intended voicing or place were eliminated from the pool of potential materials.

3.2.2. Pronunciation accuracy: Acoustics

The items selected based on transcription accuracy were also confirmed to be appropriate via acoustic analysis. The onsets and offsets of aperiodic energy associated with frication were used to identify the fricative. Using the identified interval, three measurements were made: (i) the *percentage unvoiced* of the fricative was estimated using the Praat Voice Report function (Boersma & Weenink, 2020), (ii) the raw *fricative duration*, and (iii) the duration of the fricative divided by the duration of the word it occurred in—henceforth, *ratio duration*. These three measurements were used to assess whether there are reliable differences in the manifestation of voicing when the underlying fricative differs, and when the item was produced in a sentence (for exposure) versus in isolation (for test). For example, is the [s] in *assembly* as [ə.sɛm.bli] equivalent to the [s] when *appetizer* is produced as [æ.pɛ.taɪ͡.sɻ̩]? **Table 4** provides the summary statistics for

Underlying fricative in word	Produced Fricative	Stimuli type	Percent Unvoiced	Fricative Duration (ms)	Ratio Duration
s	s	sentence	89.79 (3.42)	115.38 (16.93)	0.19 (0.04)
s	s	words*	88.48 (5.02)	120.86 (15.27)	0.18 (0.03)
z	s	sentence	90.33 (7.06)	116.78 (8.57)	0.20 (0.04)
z	s	words	89.51 (5.04)	123.25 (11.68)	0.19 (0.04)
s	ʃ	words	89.70 (4.38)	133.85 (15.67)	0.20 (0.03)
z	z	sentence	75.54 (13.53)	80.13 (7.36)	0.14 (0.03)
z	z	words*	42.63 (33.61)	83.39 (11.245)	0.13 (0.03)
s	z	sentence	62.76 (22.35)	82.91 (12.24)	0.13 (0.03)
s	z	words	34.63 (32.30)	89.24 (14.35)	0.13 (0.03)
z	ʒ	words	30.11 (32.26)	87.55 (17.24)	0.13 (0.02)

Table 4: Summary statistics (means, with standard deviation in parentheses) for acoustic measures related to voicing. The stimuli types accompanied by an asterisk were not used as stimuli for the experiment, but were recorded to allow comparison of shifted pronunciations to naturally produced tokens.

all stimuli types used in the experiment as well as the same words pronounced canonically. Canonical productions of single words were recorded separately explicitly for these acoustic comparisons.

As these data show, items produced as [s] in word and sentence contexts are more likely to be unvoiced, have longer raw fricative durations, and longer ratio durations than items that were produced with [z], regardless of whether the canonical form contained /s/ or /z/. To quantify whether exposure sentence stimuli were well-matched, and to corroborate the transcription described in the previous section, the following comparisons were made.

To assess the acoustic equivalence of target words in the exposure stimuli, underlying target /s/ words produced with [s] from the control condition of Experiment 3 (/s/-voicing, $n = 36$) were compared with devoiced /z/ words from the experimental condition of Experiment 2 (/z/-devoicing, $n = 36$). Similarly, underlying target /z/ words produced as [z] in the control condition of the /z/-devoicing experiment ($n = 36$) were compared with voiced /s/ words ($n = 36$) from the experimental condition of Experiment 3 (/s/-voicing). A series of ANOVAs were run separately for [s] and [z] items (but differing in underlying representation), using percentage unvoiced, fricative duration, and ratio duration as dependent measures and the underlying fricative as the single independent variable. There were no significant effects, suggesting that [s] and [z] were produced consistently in sentence contexts regardless of the voicing of the fricative in the canonical pronunciation of the word.

To assess the acoustic equivalence of the single word test items, underlying /z/ items produced as [s] (n = 36) and underlying /s/ items produced as [z] (n = 36) were compared to the /s/ and /z/ items produced in their canonical form, respectively; again, these canonical single word items are not used as stimuli, but were recorded to make these comparisons. The fricatives from single word environments were assessed separately for [s] and [z] pronunciations through a series of ANOVAs with percentage unvoiced, fricative duration, and ratio duration as dependent measures with the underlying fricative as an independent variable. There was no significant effect of underlying fricative in any of the three ANOVAs. Together, these results suggest that the pronunciation of the /z/ as [s] and /s/ as [z] in sentence and word contexts well-matched canonical pronunciations of /z/ and /s/ by the same speaker.

The /s/ as [ʃ] and /z/ as [ʒ] test items were not compared to /ʃ/ and /ʒ/ as produced by this speaker in natural contexts because appropriate comparison items were not recorded. Nonetheless, [ʃ] and [ʒ] are well-matched to [s] and [z], respectively, in terms of acoustic presentation of voicing (i.e., measures of percent unvoiced, fricative duration, and ratio duration), which indicates these productions were similarly well-matched.

4. Experiment 2: Learning /z/-devoicing

In Experiment 2 listeners were presented with devoiced /z/ using a sentence exposure task. Perceptual adaptation was assessed by listeners' endorsements of items with devoiced /z/ as words in a lexical decision test following the exposure phase. Generalization was tested through the presentation of novel devoiced /z/ words, which comprised items not presented in the exposure phase. To determine whether adaptation was targeted in the direction of /z/-devoicing or reflected a more general relaxation of /z/ criteria, items containing [ʒ] in place of [s] were included in the second half of the test block—of these, half were heard during the exposure phase with [s], and half were novel /z/ words. The performance of listeners presented with the devoiced /z/ items in exposure was compared to those in a control group who heard the same items with canonical /z/ pronunciations during exposure. All participants completed the same lexical decision test.

4.1. Methods

We report on materials and procedures first, followed by participants, as this order allows us to contextualize participant outlier removal in a more transparent manner.

4.1.1. Materials

The final stimuli for the exposure phase for Experiment 2 consisted of 56 semantically coherent filler sentences, randomly sampled from the 100 possible filler sentences, and two versions of the

14 semantically predictable critical sentences, in which pronunciation patterns and intonation of the sentence was consistent, but the voicing of the /z/ in the target word in the sentence varied according to condition. The control version comprised sentence-final critical words produced in their canonical form (e.g., *busy* [bɪzi]). In the experimental /z/-devoicing condition, the sibilant in the sentence-final /z/ word was produced as [s] (e.g., *busy* [bɪsi]). The test stimuli for the test phase consisted of the 14 critical exposure words (produced as single words) from training, 14 novel /z/ words (randomly sampled per participant from a possible list of 22 items), 42 nonwords (phonotactically-legal maximal pseudowords[3] randomly sampled per participant from a pool of 110 nonwords), and 70 filler words (randomly sampled per participant from a pool of 103 filler words), some of which happened to be presented in the exposure sentences.[4] The large number of filler words served to bias listeners to respond to the devoiced /z/ items as nonwords. The yoking of particular items to the exposure phase—and thus particular heard and novel items at test—was due to limitations of which stimuli were viable in terms of inter-transcriber agreement. A list of the final experimental materials used in this experiment is provided in the Appendix. Lexical frequency (log frequency per million) of items in the final word list was estimated with the SUBTLEX-us corpus (Brysbaert & New, 2009). The frequency of /z/ items used in exposure (M = 2.23, *SD* = 0.66) was not different than those used as novel items (M = 2.05, *SD* = 0.67; *Welch's* $t(28.36) = 0.81$, $p = 0.42$).

4.1.2. Procedure

Participants completed the task up to four at a time in sound-attenuated cubicles outfitted with AKG headphones, a desktop PC, and a PST serial response box. All auditory stimuli were presented at a comfortable listening level (approximately 65dB SPL) over the headphones. The experiment was controlled by E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA).

As noted previously, the study comprised two parts: an exposure phrase with auditorily presented sentences and an auditory lexical decision test phase. Half of the participants were assigned to a control condition where /z/ words were produced with the canonical voiced [z] pronunciation, and half were assigned to the experimental /z/-devoicing condition where all /z/ exposure items were devoiced and pronounced as [s]. The lexical decision test phase was identical for the two groups of listeners.

[3] Maximal here refers to the fact that the pseudowords differed from any real word with respect to multiple phones.

[4] There was no intention of balancing or controlling for the presence or absence of filler words in the filler sentences. The overlap varied by participant, as each participant's list was randomly sampled from the viable pool of items, and the overlap simply existed as a function of our lack of creativity composing semantically coherent sentences that lacked the critical sounds of interest.

In the exposure phase, 70 sentences (14 critical, 56 filler) were presented in a pseudo-random order such that no two critical trials were adjacent. The number of filler sentences presented before the first critical sentence varied randomly from six to eight across experimental conditions. Participants were informed to listen carefully to the sentences and falsely instructed that there would be comprehension questions following the sentence list, but otherwise not required to do anything while listening to the presented stimuli. Each sentence was separated by a 2000 ms pause. Participants were allowed a self-timed break after the exposure phase.

The lexical decision test phase began after the participants' self-administered break. Participants were presented with a single item over headphones and were asked to classify the item as a 'word' or 'not a word' using the button box provided. The buttons "1" or "5" for 'word' and 'not a word' were counterbalanced across participants. These response options were visually presented—numerically and orthographically—on a computer monitor and participants were given up to 1500 ms to respond. All items (n = 140) were pseudo-randomized across participants as described below. There were 42 nonwords and 70 filler words, none of which contained sibilant fricatives. In the first half of the test block, listeners were presented with 14 critical items where the underlying /z/ was pronounced as a [s]. Half of these test items had occurred in the exposure sentences and half were novel words. These items were fully randomized within the block. The second half of the test block contained 14 critical items that tested for general relaxation of the /z/ category, in which the fricative was pronounced as [ʒ]. Half of these items were lexical items that had occurred in the exposure sentences (with the exposure pronunciation as [s] or [z] depending on the condition) and half were novel in the context of the experiment. After completing the exposure and test phases, listeners completed a language background questionnaire. Participants who inquired about the (lack of) comprehension questions for the sentences were informed that those instructions were included to ensure they attended to the sentences.

4.1.3. Participants

A total of 135 adults from the Metro Vancouver community participated in this study. Participants' data were removed prior to the analysis if they did not report English as one their native languages (n = 55), reported a speech/hearing impairment (n = 1), or were below 90% accuracy on filler items (n = 5; following the exclusionary criteria of Sumner & Samuel, 2009). Seventy-four participants were retained for the analysis (Control: n = 32, Experimental: n = 42). Participants varied in gender (47 female, 23 male, 5 did not report), and of those who reported their age (n = 72), the majority were undergraduate student-aged (M = 20.18, Median = 20, *SD* = 2.26). Participants self-identified with various racial and ethnic backgrounds (White = 22, Chinese = 14, South Asian = 10, Filipino = 6, Korean = 5, Other/Mixed = 17), and typically had knowledge of multiple languages (M = 3.22 , *SD* = 1.73), including English. Participants were compensated with partial course credit or $10 CAD.

4.2. Analysis and results

Participant responses were removed if the reaction time was below 200 ms, or more than three standard deviations above the grand mean—this resulted in the removal of 0.2% of the data.

The experimental data were analyzed using Bayesian multilevel logistic regression models implemented with the *brms* R package (Bürkner, 2017). The *brms* package provides a simple interface to the popular and widely-used Stan probabilistic programming language (Stan Development Team, 2021), and adopts the familiar formula specification of models. Bayesian analysis methods are desirable from both conceptual and practical perspectives, as outlined by Vasishth and colleagues in a 2018 tutorial paper on the topic (Vasishth, Nicenboim, Beckman, Li, & Kong, 2018). While similar to frequentist mixed effects models in accounting for variability in the population ('fixed effects') and between groups within the population ('random effects'), Bayesian models allow for graded statements about the strength of evidence for all parameters. A practical benefit is that Bayesian models can be fit with maximal random effects and not run into convergence problems.

Inference in Bayesian models is based on the *posterior* distributions of parameters, which represent a range of possible parameter values accompanied by probabilities, rather than the point estimates provided in frequentist models. The posterior distribution is a combination of prior expectations (discussed below) and the likelihood of observing the data under the model. The posterior distribution gives information about the size of the effect and the confidence with which the effect size can be interpreted. As the models described in this paper use logistic regression, the posterior distributions represent the possible parameter values in the log odds space. For example, a posterior distribution with most of its probability mass in the [2.3, 4.6] range indicates that the parameter leads to a 10-fold to 100-fold increase in odds for the response variable—that is, the parameter indicates a large effect size. Bayesian models also allow us to express a degree of confidence in an effect's direction; this is shown in the reporting of the proportion of the posterior distribution with the same direction as the posterior mean. Prose descriptions of confidence are necessarily subjective, and we use the following conventions: Probabilities equal to 1 represent strong or consistent evidence, probabilities between 0.9 and 1 represent moderate evidence, probabilities between 0.8 and 0.9 are weak, and those below 0.8 offer little to no evidence.

In all models, the dependent variable was Word Endorsement, where 1 corresponds to participants' response of 'word,' and 0 to 'not a word.' Each of the models had population-level (fixed) effects for Item Type, Condition, and their interaction. Both were weighted effect coded categorical variables, in which levels are compared against the weighted mean. For example, the effect for a specific level of Item Type would indicate that the level differs from the weighted mean across all levels. Weighted effect coding accounts for unbalanced data (i.e., more filler items than critical items) and facilitates the interpretation of both main effects and interactions

(Nieuwenhuis et al., 2017). When levels are balanced, weighted effect coding is identical to effect coding.[5] Levels and weights are reported for each model below. Additionally, in all models, there were by-Participant random slopes for Item Type. This is summarized in the *brms* formula used for each model: *Word Endorsement* ~ *Item Type* × *Condition* + *(1* + *Item Type* | *Participant)* + *(1* | *Word)*.

Models also shared general specifications. Priors for the intercept and all population-level effects were *regularizing,* weakly informative Student's *t* distributions ($\nu = 3$, $\mu = 0$, $\sigma = 2.5$), following the recommendations from Gelman, Simpson, and Betancourt (2017) and Vasishth et al. (2018). This prior largely serves to keep parameter estimates within a sensible range for the log-odds space—that is, approximately 93% of this *t* distribution falls between log odds of −6.9 and 6.9, which corresponds to probabilities of 0.001 and 0.999, respectively. While extreme parameter values are possible with this prior, the prior is regularizing because it requires more evidence for extreme parameter values to be reflected in the posterior. The default weakly informative *brms* priors for the group-level and correlations of group-level parameters were used—Student's t ($\nu = 3$, $\mu = 0$, $\sigma = 2.5$), half Student's t ($\nu = 3$, $\mu = 0$, $\sigma = 2.5$), and LKJ correlation ($\eta = 1$) distributions, respectively. All models were run with four chains. Each chain had 5,000 iterations (including 2,500 warm-up iterations). This resulted in a total of 10,000 post-warmup samples in each model. Across all models, there were no divergent transitions, and $\hat{R}$ values were very close to 1 (and all < 1.05), which indicates that chains were well-mixed and that all models converged. Likewise, the effective sample size for all population-level effects was sufficiently large. For discussion regarding model specifications, see Vasishth at al. (2018), and McElreath (2020: p. 287). In the following sections, results are reported as posterior mean estimates of model parameters (β), along with 95% highest density credible intervals (henceforth, CrI)—the narrowest interval that contains 95% of the posterior distribution. As noted above, the posterior distribution of a parameter indicates a range of possible parameter values and their associated probabilities. While the interpretation of the posterior mean is analogous to the point estimates provided in frequentist mixed-effects models, the posterior distribution as a whole offers a richer picture. As the credible intervals for some parameters span zero, the probability of the effect's direction is also reported (i.e., if β is negative, the probability of $\beta < 0$).

4.2.1. Model: Learning /z/-devoicing

This model assesses whether or not participants learned /z/-devoicing, and excludes critical test items that were not previously heard during exposure. In the model, Item Type was weighted effect coded with four levels (*Item Type Filler*: Filler Word = 1 , [s] = 0, [ʒ] = 0, Nonword =

[5] For additional information on weighted effect coding, see https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faqwhat-is-effect-coding/.

−1.69; *Item Type Heard [s]*: Filler = 0, [s] = 1, [ʒ] = 0, Nonword = −0.17; *Item Type Heard [ʒ]*: Filler = 0, [s] = 0, [ʒ] = 1, Nonword = −0.16). Condition was weighted effect coded with two levels (Control = 1, Experimental = −0.76). Participants' mean word endorsement rates are summarized by Item Type and Condition in **Figure 1**.

The results of the Bayesian multilevel regression model of word endorsement for the Control and Experimental listener groups' adaptation to /z/-devoicing are described here and depicted in **Figure 2**.[6] To limit the complexity of the interactions, the first model for Experiment 2 focused on word endorsement behaviour for filler words, filler nonwords, and words presented in the exposure phase sentences, which were presented at test with either [s] or [ʒ] pronunciations. Generalization to novel items is addressed later.

The model intercept indicates that there is a slight and consistent bias to endorse items as words ($\beta = 2.09$, CrI = [1.69, 2.53], $Pr(\beta > 0) = 1$). There is evidence that participants in the Control condition have a slightly higher word endorsement rate overall ($\beta = 0.30$, CrI = [0.02, 0.59], $Pr(\beta > 0) = 0.98$). Filler word items were very likely to be endorsed as words ($\beta = 3.96$, CrI = [3.59, 4.37], $Pr(\beta > 0) = 1$), and the Condition × Item Type Filler interaction indicates that Control participants were slightly more likely to endorse filler words as words ($\beta = 0.24$, CrI = [0.04, 0.48], $Pr(\beta > 0) = 0.99$), providing evidence that Experimental condition listeners

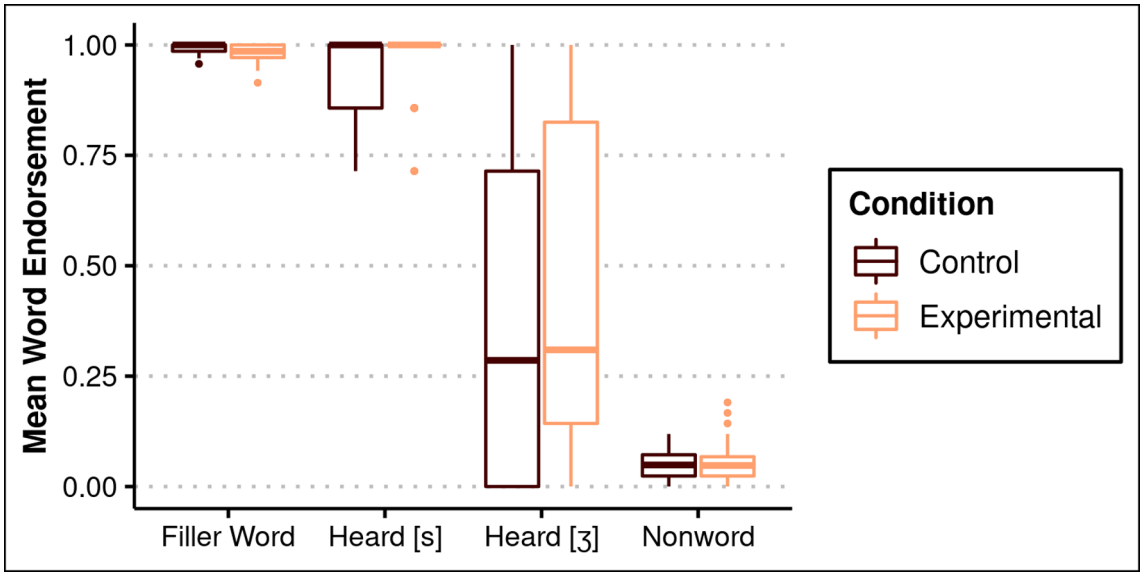

Figure 1: Mean word endorsement rates for filler words, nonwords, and all previously heard critical items across the two conditions used in the analysis of learning /z/-devoicing.

[6] Full model summaries are provided in the Appendix.

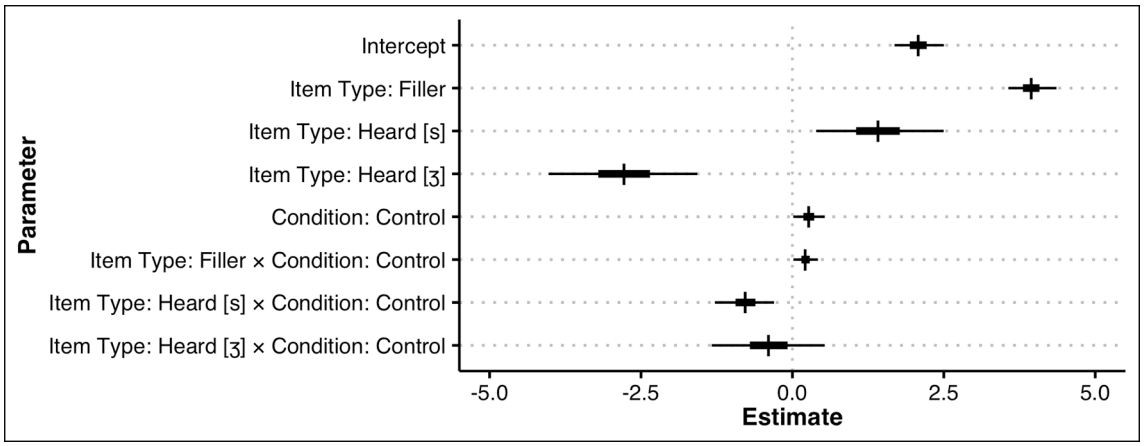

Figure 2: Population-level parameters for the learning /z/-devoicing model. Thin lines represent 95% CrI and thick lines represent 50% CrI. The posterior mean estimate for each parameter is indicated by the vertical tick mark.

may have globally adjusted their criteria for word endorsement, becoming more conservative and calling fewer filler items words—we offer possible explanations for this in the interim discussion. Overall, the population-level parameter for Item Type [s] provides evidence that previously heard [s] items are more likely to be endorsed as words ($\beta = 1.39$, CrI $= [0.36, 2.40]$, $Pr(\beta > 0) = 0.99$). The interaction of Condition and Item Type [s] provides strong evidence that Control listeners were somewhat less likely to identify [s] pronunciations as words ($\beta = -0.76$, CrI $= [-1.27, -0.28]$, $Pr(\beta < 0) = 1$). This is clear evidence in support of an adjustment to [s] pronunciations in /z/ words for the Experimental participants. While critical items presented as [ʒ] were less likely to be endorsed as words for both experimental and control listener groups ($\beta = -2.68$, CrI $= [-3.90, -1.55]$, $Pr(\beta < 0) = 1$), the evidence that conditions differed in how they responded to [ʒ] items was weak ($\beta = -0.44$, CrI $= [-1.34, 0.48]$, $Pr(\beta < 0) = 0.84$). Together with the low estimate for Item Type [ʒ], this indicates that, overall, listeners in both conditions were very unlikely to endorse these items as words, though note the wide range of variability in **Figure 1**.

4.2.2. Model: Generalizing /z/-devoicing

A second model addressed the question of whether participants generalized their learning of /z/-devoicing to novel words, that is, words not heard during the exposure phase. For this model, Filler Words and Nonwords were excluded, and all critical items included (both Heard and Novel). All aspects of the model structure were identical to that of the previous section, with the following exceptions. Item Type was weighted effect coded with four levels (*Item Type Novel [s]*: Heard [s] $= -0.97$, Novel [s] $= 1$, Heard [ʒ] $= 0$, Novel [ʒ] $= 0$; *Item Type Heard [ʒ]*: Heard [s] $= -0.94$, Novel [s] $= 0$, Heard [ʒ] $= 1$, Novel [ʒ] $= 0$; *Item Type Novel[ʒ]*: Heard [s] $=$

-0.94, Novel [s] $= 0$, Heard [ʒ] $= 0$, Novel [ʒ] $= 1$). Condition was weighted effect coded with two levels (Control $= 1$, Experimental $= -0.75$). Participants' mean word endorsement rates are summarized by Item Type and Condition in **Figure 3**.

The model intercept indicates that there is an overall bias towards endorsing items as words ($\beta = 1.04$, CrI $= [0.47, 1.63]$, $Pr(\beta > 0) = 1$). There is strong evidence that novel words pronounced with [s] are endorsed as words ($\beta = 1.16$, CrI $= [0.59, 1.76]$, $Pr(\beta > 0) = 1$), but little to no evidence that this interacts with Condition ($\beta = 0.13$, CrI $= [-0.28, 0.56]$, $Pr(\beta < 0) = 0.73$). That is, listeners in the experimental condition did not generalize their learning of /z/-devoicing to novel /z/ words pronounced with [s]. Overall, listeners were less likely to endorse Heard ($\beta = -1.74$, CrI $= [-2.35, -1.14]$, $Pr(\beta < 0) = 1$) and Novel ($\beta = -2.23$, CrI $= [-2.82, -1.68]$, $Pr(\beta < 0) = 1$) items where /z/ was pronounced as [ʒ] as words. The evidence that this interacted with Condition is weak for Novel [ʒ] items ($\beta = 0.20$, CrI $= [-0.18, 0.59]$, $Pr(\beta < 0) = 0.85$) and non-existent for Heard [ʒ] items ($\beta = 0.06$, CrI $= [-0.42, 0.56]$, $Pr(\beta > 0) = 0.61$). These results are depicted in **Figure 4**.

4.3. Discussion

Regardless of condition assignment, listeners were very likely to identify /z/ words with devoiced [s] pronunciations as words. Listeners in the experimental condition who were exposed to devoiced /z/ in training were more likely to endorse heard /z/ words with devoicing as words,

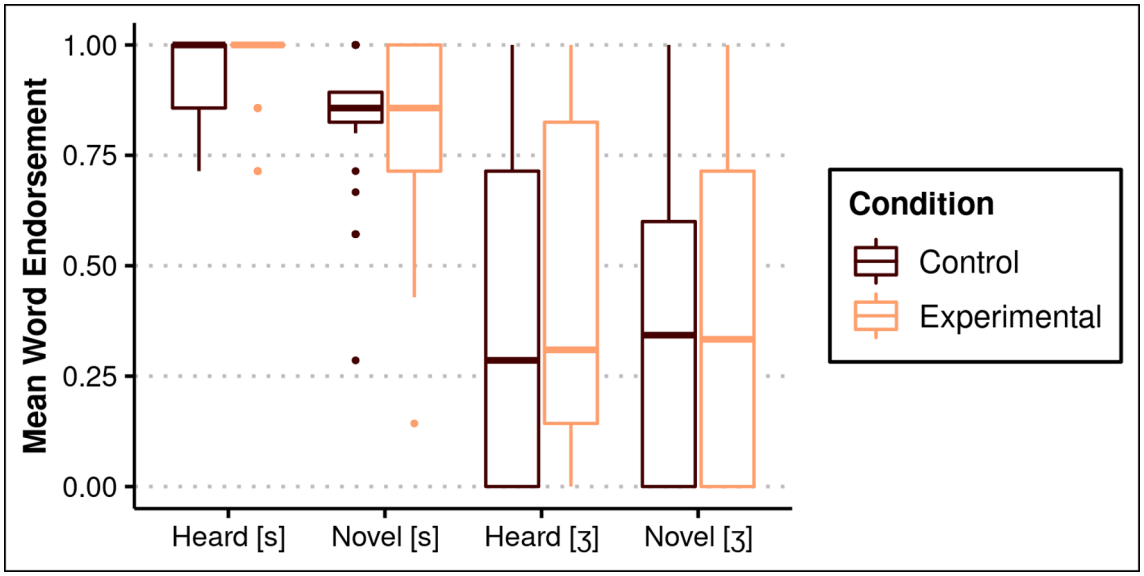

Figure 3: Mean word endorsement rates for previously Heard and Novel critical Item Types (both [s] and [ʒ] pronunciations) across the two conditions used in the analysis of generalizing /z/-devoicing.

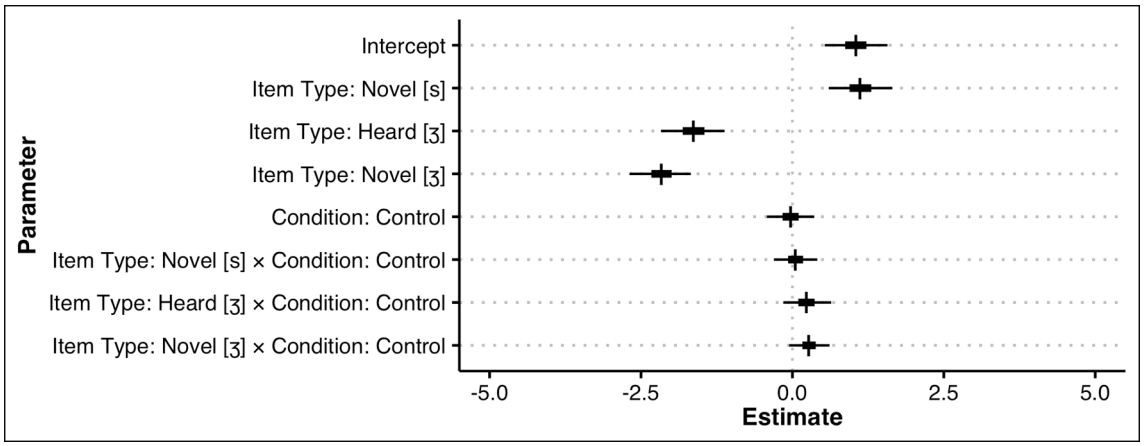

Figure 4: Population-level parameters for the generalizing /z/-devoicing model. Thin lines represent 95% CrI and thick lines represent 50% CrI. The posterior mean estimate for each parameter is indicated by the vertical tick mark.

indicating that they had adjusted their thresholds for acceptable or identifiable realizations of /z/ in a directional manner for the items they were exposed to. However, there was no evidence for generalization for listeners in the experimental group: While listeners in the experimental condition were more likely to identify previously heard devoiced /z/ words as words, there was no evidence that novel /z/ words pronounced with [s] were more likely to be identified as words. This outcome highlights the potentially word-specific nature of perceptual adaptation.

In the generalization model, there was weak evidence that control listeners were *more* likely to endorse novel /z/ words with [ʒ] pronunciations as words than experimental listeners, suggesting not only that experimental condition listeners used a directional adaptation mechanism to adjust their /z/ category, but also that control listeners exposed to multiple novel pronunciations at test may have relaxed their criteria for /z/ in novel words across the course of the test phase, as they were, for the first time, exposed to this talker producing /z/ as [s] and then [ʒ]. This suggests that while listeners did not make any general category relaxation for /z/ in the context of lexical items they heard as [z] in the exposure phase, there is weak evidence that /z/ category boundaries were slightly relaxed to include [ʒ]-like pronunciations for novel lexical items. When listeners had not previously been presented with counter-evidence for specific lexical items, control listeners show a tendency towards category relaxation in the test phase. This pair of outcomes indicates that the same underlying category can be subjected to different mechanisms depending on the nature of the variation in the input. Simply, the nature of the input triggers the adaptation mechanism.

Note that there was moderate evidence that listeners exposed to devoiced-/z/ pronunciations in the exposure phase were less accurate on the categorization of filler words. This is a small (well below a doubling of odds) but consistent effect, and may be due to an apparent difference in the

word-nonword balance across conditions. Experimental participants have learned to perceive target items (which are part of the nonword count in the stimuli distribution) as words; they identify fewer filler words as words. An alternative possibility is that by virtue of learning that a talker makes certain pronunciation changes, listeners may anticipate additional pronunciation changes which render licit words as nonwords. Both of these interpretations are speculative and warrant future consideration, and stem from a small effect.

In summary, as expected, both listener groups were likely to call devoiced /z/ items words, as they are all exposed to such pronunciations in their natural input. Listeners who were exclusively presented with devoiced /z/ in the exposure phase were even more likely to identify these items as words. Their adaptations were limited to the specific items they were exposed to and in the direction of variation to which they were exposed. That is, /z/-devoicing was not extended to novel /z/ items that were not presented in the exposure sentences and alveopalatalized [ʒ] pronunciations of /z/ were not more likely to be identified as words. This lack of generalization was unexpected, and we discuss this further in the general discussion.

5. Experiment 3: Learning /s/-voicing

Experiment 2 established that listeners adjust their lexical decision thresholds to a talker's /z/-devoicing pattern for items they were exposed to, albeit in a somewhat narrow manner. In Experiment 3 we tested whether listeners learned an /s/-voicing pattern, which is both typologically marked and very likely outside of listeners' regular perceptual experience in English, as demonstrated by Experiment 1. The materials and procedures for Experiment 3 are equivalent to those for Experiment 2, with the exception that critical items were underlyingly /s/ words, and listeners were exposed to /s/-voicing instead of /z/-devoicing. To assess whether any adjustments were due to a directional adaptation or a general relaxation mechanism, items with [ʃ] were included in the second half of the test block.

5.1. Methods

5.1.1. Materials

Like in Experiment 2, the stimuli for the exposure phase in Experiment 3 comprised 56 semantically coherent filler sentences, randomly sampled from a pool of 100 filler sentences, and two versions of the 14 semantically predictable critical sentences. The control conditions had sentence-final critical words produced in their canonical form (e.g., *cassette* [kəˈsɛt]). In the experimental /s/-voicing condition, the sibilant in the critical /s/ word was produced as [z] (e.g., *cassette* [kəˈzɛt]). The test stimuli for the lexical decision task consisted of the 14 critical /s/ words heard during exposure (half with [ʃ] in place of [z]), 14 novel /s/ words (half with [ʃ] in place of [z]; randomly sampled by participant from a list of 22 possible novel words), 42 nonwords (phonotactically-legal maximal pseudowords randomly sampled by participant from

a pool of 110 nonwords), and 70 filler words (randomly sampled by participant from a pool of 103 filler words). Again, the large number of filler words serves to bias listeners to respond to the voiced /s/ items as nonwords. The full list of items is presented in the Appendix.

Like for the /z/ items, lexical frequency (log frequency per million) of the final stimuli list were estimated using the SUBTLEX-us corpus (Brysbaert & New, 2009). The /s/ items used in exposure (M = 1.93, *SD* = 0.61) were matched in frequency with the novel items (M = 2.03, *SD* = 0.59; Welch's $t(27.25) = -0.48$, $p = 0.63$) that were used in the lexical decision test. A list of the critical sentences and lexical items from the lexical decision task are shown in the Appendix. The lexical frequency of the items used in Experiment 2 (/z/-devoicing) and Experiment 3 (/s/-voicing) did not differ significantly from one another (Welch's $t(69.27) = -0.83$, $p = 0.41$).

5.1.2. Procedure

The procedure for Experiment 3 was identical to that of Experiment 2.

5.1.3. Participants

There were 135 adult participants from the Metro Vancouver community in Experiment 3. Using the same criteria as in Section 4.1.3., participants were excluded if they were not self-reported native speakers of English (n = 34), had a speech or hearing impairment or did not answer the question (n = 7), or scored below 90% on filler words (n = 7). There were 87 participants retained in the analysis (Control: n = 46, Experimental: n = 41). Participants varied in gender (64 female, 16 male, 1 fluid, 1 non-binary, 4 did not report), and of the participants who reported their age (n = 84), the majority were undergraduate student-aged (M = 21.08, Median = 20, *SD* = 3.90). Participants self-identified with various racial and ethnic backgrounds (White = 31, Chinese = 20, South Asian = 8, Filipino = 3, Southeast Asian = 3, Other/Mixed = 22), and typically had knowledge of multiple languages (M = 3.86, *SD* = 1.32), including English. Participants were compensated with partial course credit or $10 CAD.

5.2. Analysis and results

The analysis for Experiment 3 was nearly identical to that of Experiment 2. Less than 0.2% of the data was removed due to reaction time filtering. The models for Experiment 3 have the same formula, weak priors, and specifications as described in Section 4.2. For reference, the formula was: $\textit{Word Endorsement} \sim \textit{Item Type} \times \textit{Condition} + (1 + \textit{Item Type} \mid \textit{Participant}) + (1 \mid \textit{Word})$.

5.2.1. Model: Learning /s/-voicing

This model assesses whether or not participants learned /s/-voicing, and excludes critical test items that were not heard during exposure. In the model, Item Type was weighted effect coded

with four levels (*Item Type Filler*: Filler Word = 1 , [z] = 0, [ʃ] = 0, Nonword = −1.69; *Item Type Heard [z]*: Filler = 0, [z] = 1, [ʃ] = 0, Nonword = −0.16; *Item Type Heard [ʃ]*: Filler = 0, [z] = 0, [ʃ] = 1, Nonword = −0.16). Condition was weighted effect coded with two levels (Control = 1, Experimental = −1.12). Mean word endorsement rates are summarized by Item Type and Condition in **Figure 5**.

In the model of learning /s/-voicing, the intercept indicates that there is an overall bias towards endorsing items as words ($\beta = 1.79$, CrI = [1.41 , 2.20], $Pr(\beta > 0) = 1$), though there was little to no evidence of a meaningful difference across conditions ($\beta = 0.09$, CrI = [−0.12, 0.31], $Pr(\beta > 0) = 0.80$). There is strong evidence that filler words were consistently endorsed as words ($\beta = 4.13$, CrI = [3.77, 4.52], $Pr(\beta > 0) = 1$), and that this interacted with Condition. Control participants were slightly more likely to endorse filler words as words ($\beta = 0.32$, CrI = [0.16, 0.49], $Pr(\beta > 0) = 1$), suggesting that listeners in the Experimental group may have globally adjusted their criteria for word endorsement, becoming more conservative with respect to what constitutes a word, as in Experiment 2. There is weak evidence that Heard [z] items were less likely to be endorsed as words ($\beta = -0.64$, CrI = [−1.76, 0.47], $Pr(\beta < 0) = 0.87$), and furthermore, the Condition × Item Type [z] parameter indicates that Control participants were less likely to endorse Heard [z] items as words ($\beta = -0.57$, CrI = [−1.27, 0.10], $Pr(\beta < 0) = 0.95$). This provides some evidence that listeners in the Experimental group adjusted their /s/ criteria to accommodate [z] pronunciations. There is strong evidence that [ʃ] items were less likely to be endorsed as words overall ($\beta = -3.38$, CrI = [−4.76, −2.06], $Pr(\beta < 0) = 1$). Additionally, Control participants were less likely to identify [ʃ] items as words ($\beta = -0.51$,

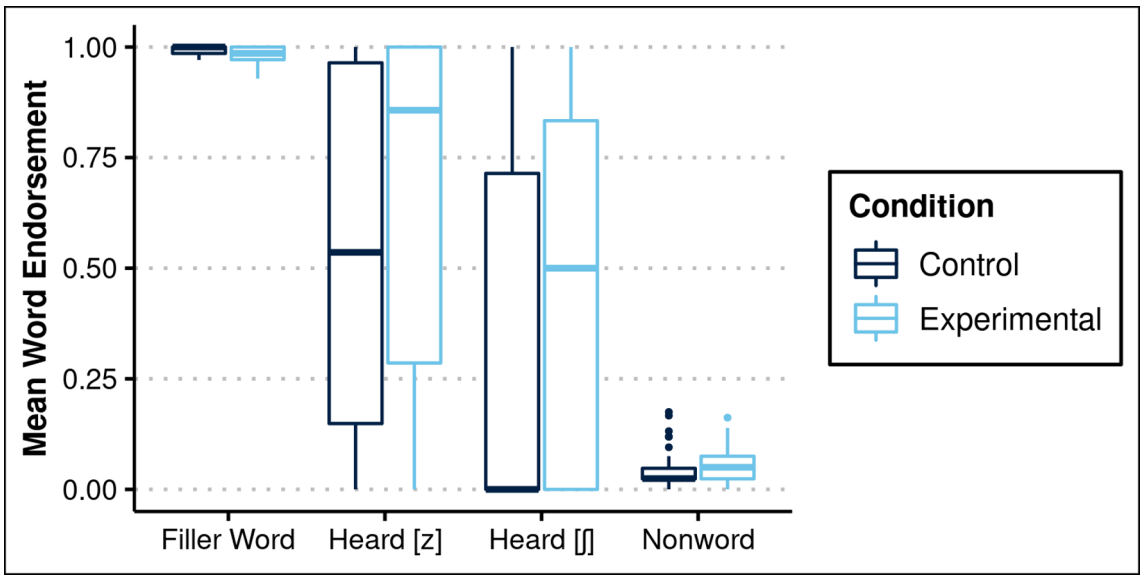

Figure 5: Mean word endorsement rates for filler words, nonwords, and all previously heard critical items across the two conditions used in the analysis of learning /s/-voicing.

CrI = [−1.43, 0.40], Pr(β < 0) = 0.87), though given the probability of the effect's direction, this should be interpreted as weak evidence that Experimental participants generally relaxed their word endorsement thresholds. The population-level parameters for the model of learning /s/-voicing are depicted in **Figure 6**.

5.2.2. Model: Generalizing /s/-voicing

This model addresses the question of whether participants generalized their learning of /s/-voicing to novel words—those not heard during exposure. As in Section 4.2.2, Filler Words and Nonwords were excluded, while all Heard and Novel critical items were retained. The same model structure was used. Item Type was weighted effect coded with four levels (*Item Type Novel [z]*: Heard [z] = −0.99, Novel [z] = 1, Heard [ʃ] = 0, Novel [ʃ] = 0; *Item Type Heard [ʃ]*: Heard [z] = −0.99, Novel [z] = 0, Heard [ʃ] =1 , Novel [ʃ] = 0; *Item Type Novel [ʃ]*: Heard [z] = −1.01, Novel [z] = 0, Heard [ʃ] = 0, Novel [ʃ] = 1). Condition was weighted effect coded with two levels (Control = 1, Experimental = −1.14). Participants' mean word endorsement rates are summarized by Item Type and Condition in **Figure 7**.

In the model of generalizing /s/-voicing, there is a slight overall bias to call items nonwords (β = −0.87 , CrI = [−1.75, −0.04], Pr(β < 0) = 0.98), unlike each of the previous three models. Overall, there was some moderate evidence that novel /s/ items pronounced with [z] are more likely to be endorsed as words (β = 0.32, CrI = [−0.19, 0.84], Pr(β > 0) = 0.90), and strong evidence that [ʃ] pronunciations for /s/ items are less likely to be endorsed as words, whether Heard (β = −0.95, CrI = [−1.55, −0.41], Pr(β < 0) = 1) or Novel (β = −1.16 , CrI = [−1.80, −0.56], Pr(β < 0) = 1). There is moderate evidence that Control participants

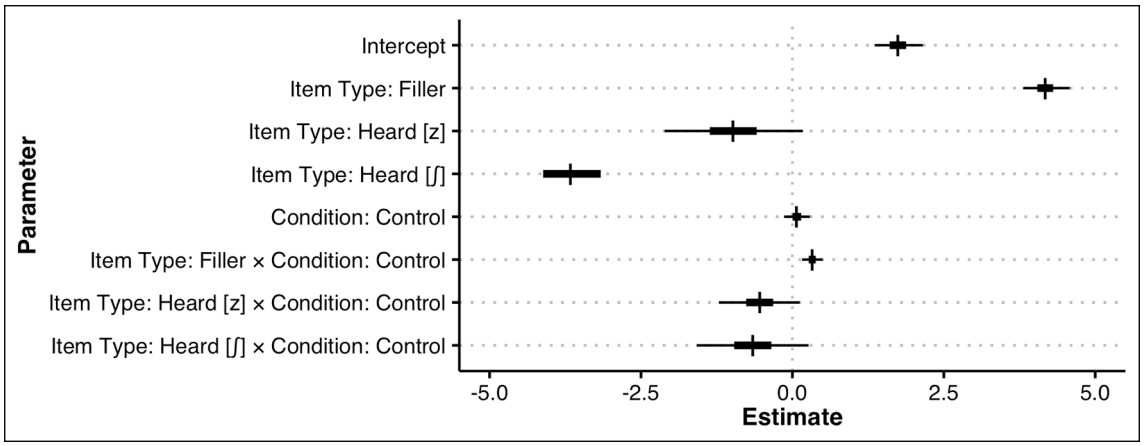

Figure 6: Population-level parameters for the learning /s/-voicing model. Thin lines represent 95% CrI and thick lines represent 50% CrI. The posterior mean estimate for each parameter is indicated by the vertical tick mark.

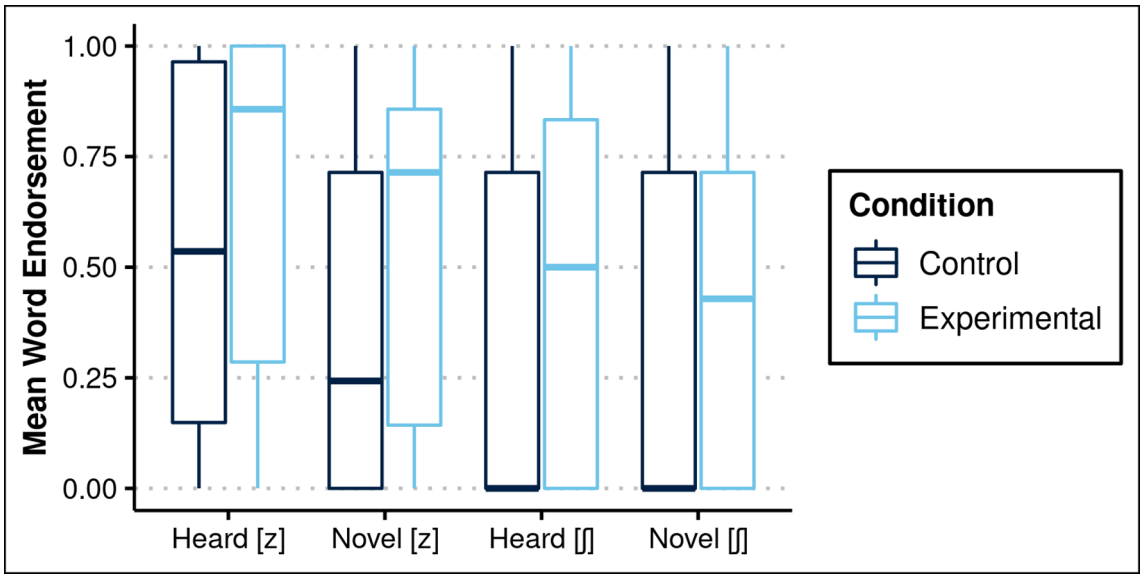

Figure 7: Mean word endorsement rates for previously Heard and Novel Item Types (both [z] and [ʃ] pronunciations) across the two conditions used in the analysis of generalizing /s/-voicing.

are less likely to endorse items as words across the board ($\beta = -0.57$, CrI $= [-1.30, 0.17]$, $Pr(\beta < 0) = 0.94$), suggesting that Experimental participants may have generally relaxed their criteria for /s/ which manifests here as greater acceptance of [ʃ] pronunciations. There was no evidence that Condition interacted with Item Type, as each interaction term substantially overlaps with zero, possibly because the pattern is well-captured by the main effect of Condition. The population-level parameters described here are depicted in **Figure 8**.

5.4. Discussion

Listeners were assigned to an experimental condition where /s/ words were produced with [z]—a voiced fricative at the same place of articulation—or to a control condition where a typical [s] was heard in the same sentence set. At test, all listeners were presented with /s/ items from exposure containing /s/-voicing, novel /s/ words with voicing, heard /s/ words pronounced with [ʃ], and novel /s/ words with [ʃ]. Bayesian multilevel logistic regression models present very weak evidence that listeners exposed to voiced /s/ adapted their /s/ category to specifically accommodate these pronunciations. That is, there is little to no evidence of a directional adjustment mechanism in play in response to exposure to /s/-voicing. However, there is weak-to-moderate evidence that, at test, listeners exposed to the voiced /s/ in exposure were more likely to call any non-canonical pronunciation of an /s/ word a word than listeners in the control condition, suggesting that any change in /s/ category structure was the result of a more general relaxation of /s/ criteria, as opposed to any directional adjustments towards voiced /s/.

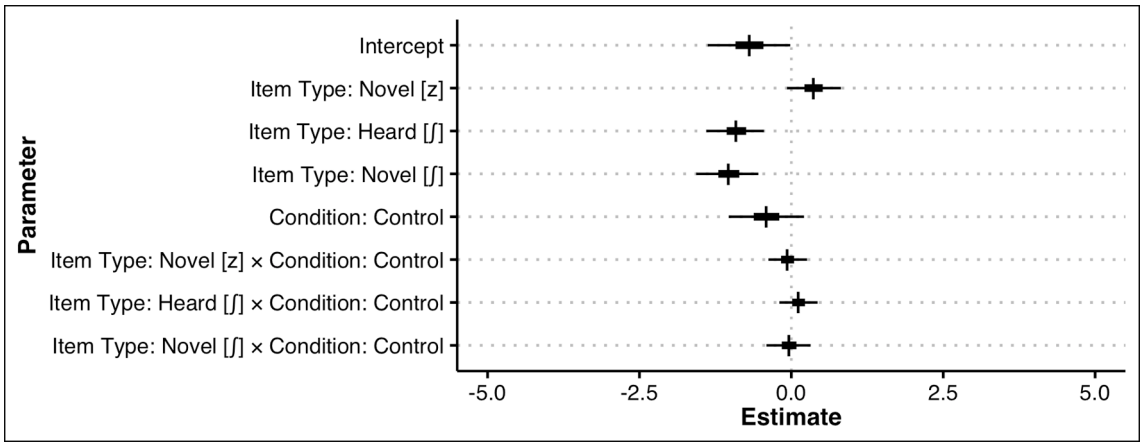

Figure 8: Population-level parameters for the generalizing /s/-voicing model. Thin lines represent 95% CrI and thick lines represent 50% CrI. The posterior mean estimate for each parameter is indicated by the vertical tick mark.

6. General discussion

Our goal was to examine asymmetries in adaptation to non-canonical pronunciations, focusing on the voicing patterns of coronal fricatives, given the typological and English-specific tendencies for these fricatives to devoice as opposed to voice. In Experiment 1, we first confirmed that North American English speakers in the Buckeye Corpus produce substantially more (categorical) /z/-devoicing than /s/-voicing in spontaneous speech. Such a confirmation reinforces an expected asymmetry that is reflected in the behavioural data of Experiments 2 and 3: We expected listeners to adjust their word endorsement behaviours in different ways to /z/-devoicing and /s/-voicing. The devoicing of /z/ is not only typologically frequent and phonetically natural, but also comparatively frequent in spontaneous speech in North American English. Categorical voicing of /s/, on the other hand, is typologically rare, phonetically unnatural, and very rare in spontaneous North American English. In Experiments 2 and 3, we presented listeners with naturally-produced words containing coronal fricatives that had been produced with either devoicing (/z/ → [s]) or voicing (/s/ → [z])—compared to their canonical citation forms—in sentences. Groups of listeners in control conditions were presented with the same sentences and words with their typical fricative realizations (/z/→ [z]; /s/ → [s]). We tested whether listeners in the Experimental conditions adjusted their fricative categories, identifying more items with a change in coronal fricative voicing as words in a lexical decision test. To assess whether adjustments are targeted in the direction of the exposed variant or whether such adjustments are the results of general category relaxation, the test block also presented listeners with an additional change in place of articulation. In the latter part of the test phase, listeners were presented with items where an expected /z/ was replaced by [ʒ] and /s/ by [ʃ].

The Bayesian statistical analysis provides nuance in evaluating the evidence for directional adaptation and general relaxation mechanisms in perceptual learning. While listeners in both the control and experimental conditions of Experiment 2 (/z/-devoicing) showed strong evidence of identifying devoiced /z/ items as words, those in the experimental condition were more likely to identify these items as words, demonstrating directional adaptation. These same listeners, however, did not generalize their adjustments to novel devoiced /z/ items, underscoring just how lexically-specific the adjustments were, though both listener groups did identify these items as words at high rates, which is expected given it is a pronunciation pattern listeners have substantial experience with. Listeners in the /z/-devoicing control condition (Experiment 2) were more likely than those in the experimental condition to accept [ʒ] pronunciations as words, suggesting that exposure to novel pronunciations at test, after having been presented with canonical /z/ pronunciations in exposure, may have triggered a general category relaxation mechanism at that point. It is worth reiterating that the evidence for this behaviour in the control condition was weak, and did not extend to previously heard /z/ words pronounced as [ʒ] at test.

The behaviours in Experiment 3 (/s/-voicing) were quite different. While the evidence was weak and weak-to-moderate, respectively in the learning and generalization models, experimental condition listeners who were exposed to voiced /s/ were more likely to identify [z] *and* [ʃ] pronunciations as words compared to control listeners. In the generalization model, this was an across-the-board increase in word endorsement rates, applying to all items, heard or novel. This suggests that exposure to an atypical pronunciation during exposure may trigger a generalizable and more general category relaxation strategy, compared to the directional and non-generalized adjustments observed for participants exposed to /z/-devoicing. These complementary results suggest that targeted adaptation and category relaxation are not necessarily mutually exclusive mechanisms, but rather are triggered in response to the stimuli. Crucially, experimenters must acknowledge that test stimuli are indeed more exposure to the talker, and thus may exert influence on listeners' responses. **Table 5** summarizes our results with respect to our hypotheses, which were all based on predictions about the behaviour of the experimental group. Note that in mapping these summaries to our analyses, all of the Bayesian models assess the evidence for general relaxation, while only the second models for each experiment assess the evidence for generalization to novel words. While the results of these experiments offer insight into adaptation processes for listeners in the control conditions as well, such information is not summarized in **Table 5**.

The goal of these experiments was to take advantage of English phonology and fricative typology as a way to establish that there are predictable asymmetries in lexically-guided perceptual learning in speech, compared to asymmetries which have simply been fortuitously stumbled upon (e.g., Zhang & Samuel, 2014). We predicted that adjustments to /z/-devoicing would be different from those for /s/-voicing for two very different reasons.

	Experimental condition of Experiment 2: /z/-devoicing	**Experimental condition of Experiment 3: /s/-voicing**
Hypothesized mechanism	Directional adaptation	General relaxation
Talker behaviour is …	Expected	Unexpected
Baseline word endorsement behaviour	Very high rates for devoiced /z/ **Result: True, and also higher than control listeners**	Higher than control listeners, but lower than /z/-devoicing **Result: True**
Listener behaviour for novel words (tests generalization)	*Yes, listeners form representation of talker as 'devoicer'* **Result: No evidence of generalization**	Yes, as voicing change falls within general relaxation **Result: Yes, evidence of generalization**
Listener behaviour for place change (tests general relaxation)	No, exposure to devoiced /z/ reinforces prior knowledge **Result: Very weak evidence of generalization**	Yes, as place change also falls within general relaxation **Result: Yes, evidence of generalization**
Takeaway	**Evidence for directional adaptation, in a more narrow word-specific form than anticipated.**	**Evidence for general relaxation mechanism, though effects are highly variable and weaker than expected.**

Table 5: Summary of mechanisms, rationale, and predictions and results for *experimental* groups from Experiments 2 and 3. This table reproduces the hypotheses in Table 3, and adds the results for the respective cells in bold. The unexpected result is italicized.

The first reason supporting the learnability of devoiced /z/ stems from the voiced fricative's typological rarity. Voiced fricatives are typologically less common than voiceless fricatives across the languages of the world. Producing a voiced fricative involves an aerodynamic configuration that maintains the subglottal and supraglottal pressure differential for voicing whilst generating high enough pressure behind the constriction to create turbulent airflow. Voiced fricatives often devoice or become approximants in languages (i.e., debuccalization). Thus, the argument can be made that voiceless fricatives are more natural than voiced fricatives, and that listeners may be more likely to learn this more natural pattern. Indeed, in the learning of phonological patterns, there is some evidence that learners appear to be biased towards learning what is phonetically natural (though disentangling naturalness is a challenge; Martin & Peperkamp, 2020). Wilson (2006) develops "substantively biased phonology," where phonological learners have knowledge of their phonetic apparatus, and this biases them towards phonological patterns that are phonetically natural. The influence of this bias is not so strong as to prevent unnatural sound patterns from emerging through sound changes, but rather accounts for listeners' preference for phonetically natural patterns. Note that these phonological biases are studied in artificial language learning paradigms that lack lexical scaffolding to facilitate the interpretation of

noncanonical input. Comparisons between what is learnable in an artificial language learning task and what is learnable in an adaptation task are thus imperfect, but still valuable.

The observation of English listeners being more experienced with /z/-devoicing than /s/-voicing is our second reason for expecting /z/-devoicing to be learned more readily than /s/-voicing. It is well established that /z/ devoices in both lab speech (Docherty, 1992; Veatch, 1989; Stevens et al., 1992; Smith, 1997) and spontaneous speech across varieties of English. Experiment 1 echoes this prior work, in that we find evidence of /z/-devoicing and a lack of evidence for frequent /s/-voicing in conversational speech with data from the Buckeye Corpus. Why might these synchronic patterns of variation affect the nature of perceptual adaptation? Let us consider the proposed mechanisms that support perceptual learning in the first place.

Perceptual learning is a retuning of sublexical representations, which means that the nature of those representations is of crucial importance for these theories. The two leading models—TRACE (McClelland & Elman, 1986; McClelland, Mirman, & Holt, 2006) and Merge-B (Norris et al., 2000; Norris et al., 2016)—disagree about *where* that retuning takes place. The current study does not arbitrate between these models, but instead offers insight into other aspects of the adaptation process, and is thus applicable to both models. The multidimensional phonetic distribution of sublexical /z/ likely includes devoicing, given its frequency in our corpus analysis. On the other hand, the sublexical representation of /s/ likely does not include any instances of voicing based on listeners' prior experiences with their language. Simply, the sublexical representations reflect listeners' prior experiences with speech sounds, and mechanisms supporting lexically-guided perceptual adaptation appear to be a means by which these are updated. So, the fact that listeners regularly experience devoiced /z/ in their natural ambient input and not voiced /s/ can account for the fact that the /z/ control group of Experiment 2 had a high rate of devoiced /z/ endorsement, while the /s/ control group of Experiment 3's endorsement of voiced /s/ items was lower. But, why does experience hearing devoiced /z/ in the exposure phrase boost word endorsement of the exposed items at test for the /z/ group, but the group exposed to voiced /s/ shows no evidence of directional word-specific learning? Why does the /s/-voicing group exhibit evidence of global /s/ category relaxation instead?

Recent work by Tzeng, Theodore, and Nygaard (2021) clearly demonstrates—using /s/-/ʃ/ recalibration—that phonemic adjustments represent a cumulative experience update. Given this, we expect that /s/-voicing would be learnable. Kraljic, Samuel, and Brennan (2008b) argue that learning can be blocked when listeners can attribute the deviant pronunciation to an incidental speech trait, as opposed to a characteristic one. Kraljic and Samuel (2011) revise this reasoning, and instead appeal to episodic models of speech perception: Listeners learn that incidental ambiguous pronunciations are perceptual experiences that update episodes associated with particular experiences, such as talking with a pen in one's mouth. They argue that different representations are updated according to different experiences. In our case, with the lack of

robust directional learning of voiced /s/, it does not seem likely that listeners would have treated these as unique or distinct episodes (as in Kraljic & Samuel, 2011), given that they heard multiple instances of voiced /s/ and no counter-evidence to the pattern in the exposure phase. Additionally, while the artificial language learning work discussed above indicates that more natural patterns are more easily learned, listeners are still fully capable of learning unnatural patterns. While it is possible that listeners in the /s/-voicing condition (Experiment 3) did not show targeted directional learning because the pronunciation pattern was deemed an incidental speech trait, following the logic of Kraljic et al. (2008b), we offer an alternative explanation that exists within the sound change literature.

A mechanism that may account for the current results exists more readily in theories about sound change, which could be viewed as *the long game* in accommodating new sound patterns. Several recent models intended to account for diachronic sound change all highlight a need for mechanisms that filter or discard ambiguous items (Wedel, 2006; Sóskuthy, 2015; Hay, Pierrehumbert, Walker, & LaShell, 2015; Todd et al., 2019). Sóskuthy (2015), for example, introduces an ambiguity filter to his computational model of perception-production feedback as a means of accounting for the simultaneous stability and flexibility of sound systems. An item experienced by an agent may be discarded if it is phonetically ambiguous. In Sóskuthy's model, like many other episodically-flavoured models (e.g., Pierrehumbert, 2002, 2003; Johnson, 1997), an update to a category shifts the category to accommodate the newly experienced item into its distribution. Highly frequent words have denser distributions and are thus less affected by new experiences, while low frequency words will be more easily swayed by novel pronunciations. Sóskuthy describes this in terms of phonetic ambiguity. Hay et al. (2015), on the other hand, propose a mechanism with a similar function that operates more on the word level. Hay and colleagues examine the front vowel push chain in New Zealand English using corpus data that span 130 years. They propose a mechanism similar to the feel of Sóskuthy's ambiguity filter: If items are "difficult to understand, they will be less likely to be encoded in memory" (Hay et al., 2015: 88). In their work, the crucial example relates to high frequency and low frequency words at a category boundary. Low frequency words of the encroached upon category will then be more readily comprehended and encoded when they are more acoustic-phonetically distant from the original category. Along similar lines, a computational model of sound change by Todd, Pierrehumbert, and Hay (2019) implements an exemplar-based model that includes evaluation stages for discriminability and typicality prior to a storage stage, which contributes to the modelling of frequency effects in sound change.

A post-perceptual evaluative stage similar to those proposed by sound change theorists may account for the directional adaptation to devoiced /z/, which did not generalize, compared to the very different category relaxation behaviour, which included generalization to novel items, for voiced /s/. Such a mechanism is congruent with our results, and the prospect of integrating

theoretical mechanisms for category flexibility and stability in both synchronic and diachronic scenarios is appealing. When listening to devoiced /z/ items in sentences, listeners likely had no trouble recognizing the intended word and it was likely not considered an extreme outlier, as listeners regularly experience devoiced /z/. Thus, these devoiced /z/ items were filed away and used to update the /z/ distribution for particular lexical items *only*. This suggests that listeners may have been explicitly learning an alternate pronunciation of particular lexical items (similar to the dual mapping described in Samuel & Larraza, 2015), which in the context of our results did not influence the entire /z/ distribution. While we did not test lexical endorsement for canonical pronunciations, there is no reason to believe that listeners would reject such items (e.g., listeners would likely accept both [dʌsən] and [dʌzən] for dozen, regardless of condition). Listeners presented with sentences that include items with voiced /s/ likely comprehended the utterances as intended. However, given the rarity of experience with voiced /s/, these items may not have passed an ambiguity filter threshold, despite recognition in context. As a result, these items may not have directly updated the distribution, but instead affected the threshold of what may constitute an /s/. These listeners seemed to be more willing to accept any novel pronunciation as /s/ compared to control listeners, as evidenced by the interaction between condition by item types in the /s/-learning model, and the main effect of condition in the /s/-generalization model.

This kind of post-perceptual evaluation is becoming viewed as necessary to account for diachronic sound change—and offers an elegant account of our results—but how does it fit with the models and mechanisms discussed above for lexically-guided perceptual learning? A stage that involves weighting of evidence and goodness-of-fit evaluation seems to align well within a Bayesian model like Merge B (that has a more explicit decision stage built into the model framework), though an evaluation stage is not incompatible with an interactive model of spoken language processing (e.g., TRACE). Building such a stage into our models allows us to account for asymmetries in adaptation that ostensibly relate to the magnitude of experienced category deviation. While we remain agnostic between the two opposing proposals, the degree to which lexical activation feeds back or merges with a phonemic category to allow a sufficient signal-to-phoneme mapping update may be determined by the certainty or confidence in the lexical assessment and deviant sound's category goodness (Babel, McAuliffe, et al., 2019). Ultimately, the current results demonstrate that subphonemic changes with which listeners have linguistic experience facilitate targeted adjustments in the direction of that experience, while novel changes seem to spur a more general relaxation of criteria for the category.

The total lack of generalization to novel devoiced-/z/ words warrants further discussion. We suggest that this is due to listeners learning a context-specific mapping associated with a particular lexical item. Previous studies highlighting a lack of generalization argue that the learned acoustic-auditory features are highly context dependent (Reinisch, Wozny, Mitterer, & Holt, 2014; Reinisch & Mitterer, 2016). The context dependency is surprising, given that

perceptual recalibration is typically assessed with the assumption of generalized learning, as the categorization continua typically used to quantify learning employ novel lexical items or non-words to localize a category boundary shift. Generalization is a hallmark of robust learning and is found in a range of instances in perceptual retuning: for example, within a voice across word position (Eisner et al., 2013) and oral stop places of articulation (Kraljic & Samuel, 2006), in addition to across sufficiently auditorily similar voices (Kraljic & Samuel, 2007; Reinisch & Holt, 2014). The targeted and specific behaviour for devoiced-/z/ merits further investigation, and while the present result is not *entirely* without precedent, it highlights that generalization should not necessarily be the default assumption.

7. Conclusion

We demonstrate that there are asymmetries in what listeners learn and how they adjust in lexically-guided perceptual learning tasks. Listeners adapted to word-specific patterns in /z/-devoicing, illustrating a directional adaptation mechanism. Those exposed to voiced /s/ showed some evidence of engaging in more generic /s/ category relaxation, subtly and variably increasing their probability of endorsing any novel /s/ pronunciation. These results suggest that perceptual learning of speech is moderated by listener knowledge or experience. The different response elicited by these two patterns may be because /z/-devoicing is phonetically natural in the languages of the world and/or because it taps into pre-existing pronunciation variants and is, therefore, readily recognized and accepted as a pronunciation that merits the updating of a word's category-associated distribution. We are, of course, unable to say which of these phonetic factors is ultimately the motivating factor for the lack of learnability of voiced /s/ and the learnability of devoiced /z/, but we clearly establish an important asymmetry in how listeners update phonetic representations. We posit that an evaluation stage, which has been invoked in the sound change literature, accounts for these results. Altogether, these results suggest that adaptation to novel pronunciation may leverage distinct mechanisms based on the nature of the phonetic variability presented.

Additional File

The additional file for this article can be found as follows:

- **Appendix**. Critical sentences used in Experiment 2 and Experiment 3 and statistical model summaries. DOI: https://doi.org/10.16995/labphon.6442.s1

Acknowledgements

Special thanks to Carolyn Norton and Zoe Lawler who played pivotal roles in earlier iterations of this project. Thanks to Martin Oberg for programming assistance, Masako Kato, Sophie Bishop, Stephanie Chung, and Cassandra Savage for their contribution to the coding of the corpus data, and Tristan Bhimaraj for being the speaker. This work has benefited from discussion with many members of the Speech in Context Lab, past and present, especially Brianne Senior, Kathleen Hall, and Michael McAuliffe. Thanks to Charlotte Vaughn and audiences at Acoustics Week in Canada, September 21–23, 2016, for comments on earlier versions of this work. All errors are our own. This work has been supported by an award from Canada's Natural Sciences and Engineering Research Council (MB).

Reproducibility

As permitted by the participant consent forms, the data and materials from Experiments 2 and 3 are available by request made to the researchers.

Competing Interests

The authors have no competing interests to declare.

References

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, *52*(3), 163–187. DOI: https://doi.org/10.1016/0010-0277(94)90042-6

Babel, M., McAuliffe, M., Norton, C., Senior, B., & Vaughn, C. (2019). The Goldilocks zone of perceptual learning. *Phonetica*, *76*(2–3), 179–200. DOI: https://doi.org/10.1159/000494929

Babel, M., Senior, B., & Bishop, S. (2019). Do social preferences matter in lexical retuning? *Laboratory Phonology, 10*(1), 4. DOI: https://doi.org/10.5334/labphon.133

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: https://doi.org/10.1121/1.4789864

Boersma, Paul & Weenink, David (2020). Praat: Doing phonetics by computer [Computer program]. Version 6.1.27, Retrieved from http://www.praat.org/

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

Brouwer, S., Mitterer, H., & Huettig, F. (2012). Speech reductions change the dynamics of competition during spoken word recognition. *Language and Cognitive Processes*, *27*(4), 539–571. DOI: https://doi.org/10.1080/01690965.2011.555268

Bruggeman, L., & Cutler, A. (2020). No L1 privilege in talker adaptation. *Bilingualism: Language and Cognition, 23*, 681–693. DOI: https://doi.org/10.1017/S1366728919000646

Brysbaert, M., & 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. *Behavior Research Methods, 41*(4), 977–990. DOI: https://doi.org/10.3758/BRM.41.4.977

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software, 80*(1), 1–28. DOI: https://doi.org/10.18637/jss.v080.i01

Chan, L., Johnson, K., & Babel, M. (2020). Lexically-guided perceptual learning in early Cantonese-English bilinguals. *The Journal of the Acoustical Society of America, 147*(3), EL277–EL282. DOI: https://doi.org/10.1121/10.0000942

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? *Perception & Psychophysics, 70*(4), 604–618. DOI: https://doi.org/10.3758/PP.70.4.604

Clopper, C. G., & Smiljanic, R. (2015). Regional variation in temporal organization in American English. *Journal of Phonetics*, *49*, 1–15. DOI: https://doi.org/10.1016/j.wocn.2014.10.002

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, *32*(2), 193–210. DOI: https://doi.org/10.1006/jmla.1993.1011

Connine, C. M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, *37*(4), 463–480. DOI: https://doi.org/10.1006/jmla.1997.2535

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: The MIT Press. DOI: https://doi.org/10.7551/mitpress/9012.001.0001

Dewaele, J. M., & Li, W. (2014). Intra-and inter-individual variation in self-reported code-switching patterns of adult multilinguals. *International Journal of Multilingualism*, *11*(2), 225–246. DOI: https://doi.org/10.1080/14790718.2013.878347

Docherty, G. J. (1992). *The timing of voicing in British English obstruents*. Netherlands Phonetics Archives. Berlin, Germany: Foris Publications. DOI: https://doi.org/10.1515/9783110872637

Drozdova, P., Van Hout, R., & Scharenborg, O. (2016). Lexically-guided perceptual learning in non-native listening. *Bilingualism: Language and Cognition*, *19*(5), 914–920. DOI: https://doi.org/10.1017/S136672891600002X

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, *67*(2), 224–238. DOI: https://doi.org/10.3758/BF03206487

Eisner, F., Melinger, A., & Weber, A. (2013). Constraints on the transfer of perceptual learning in accented speech. *Frontiers in Psychology*, *4*, 148, 1–9. DOI: https://doi.org/10.3389/fpsyg.2013.00148

Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, *19*(10), 555–xxx. DOI: https://doi.org/10.3390/e19100555

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: https://doi.org/10.1016/j.cognition.2015.02.012

Hitczenko, K., & Feldman, N. H. (2016). Modeling adaptation to a novel accent. *Proceedings of the Annual Conference of the Cognitive Science Society*, 1–6.

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*, *60*(1), 223–230. DOI: https://doi.org/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*, 943–950. DOI: https://doi.org/10.3758/s13423-011-0129-2

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson, & J.W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 145–165). San Diego, CA: Academic Press.

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama, & K. Maekawa (Eds.), *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium* (pp. 29–54). Tokyo, Japan: The National International Institute for Japanese Language.

Johnson, K., & Beckman, M. (1997). Production and perception of individual speaking styles. *Ohio State University Working Papers in Linguistics*, *50*, 115–125. DOI: https://doi.org/1811/81781

Kleber, F., John, T., & Harrington, J. (2010). The implications for speech perception of incomplete neutralization of final devoicing in German. *Journal of Phonetics*, *38*(2), 185–196. DOI: https://doi.org/10.1016/j.wocn.2009.10.001

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: https://doi.org/10.1037/a0038695

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

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology, 51*(2), 141–178. DOI: https://doi.org/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: https://doi.org/10.3758/BF03193841

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

Kraljic, T., & Samuel, A. G. (2011). Perceptual learning evidence for contextually-specific representations. *Cognition, 121*, 459–465. DOI: https://doi.org/10.1016/j.cognition.2011.08.015

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: https://doi.org/10.1111/j.1467-9280.2008.02090.x

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, *29*(1), 98–104. DOI: https://doi.org/10.1121/1.1908694

Ladefoged, P., & Johnson, K. (2011). *A course in phonetics* (6th ed.). Boston, MA: Wadsworth/Cengage Learning.

MacKenzie, L. (2018). Variable stem-final fricative voicing in American English plurals: Different pa[ð~ θ]s of change. *Language Variation and Change*, *30*(2), 147–174. DOI: https://doi.org/10.1017/S0954394518000091

Martin, A., & Peperkamp, S. (2020). Phonetically natural rules benefit from a learning bias: A re-examination of vowel harmony and disharmony. *Phonology*, *37*(1), 65–90. DOI: https://doi.org/10.1017/S0952675720000044

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: https://doi.org/10.1080/03640210802035357

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: https://doi.org/10.1121/1.4962529

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1–86. DOI: https://doi.org/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. DOI: https://doi.org/10.1016/j.tics.2006.06.007

McElreath, R. (2020). Statistical Rethinking: A Bayesian Course with Examples in R and STAN (2nd ed.). CRC Press. DOI: https://doi.org/10.1201/9780429029608

McQueen, J., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*, 1113–1126. DOI: https://doi.org/10.1207/s15516709cog0000_79

Mitterer, H., & Reinisch, E. (2017). Surface forms trump underlying representations in functional generalisations in speech perception: The case of German devoiced stops. *Language, Cognition and Neuroscience*, *32*(9), 1133–1147. DOI: https://doi.org/10.1080/23273798.2017.1286361

Moran S., & McCloy, D. (Eds.) (2019). *PHOIBLE 2.0.* Jena, Germany: Max Planck Institute for the Science of Human History.

Moreton, E., & Pater, J. (2012). Structure and substance in artificial-phonology learning, part II: Substance. *Language and Linguistics Compass*, *6*(11), 702–718. DOI: https://doi.org/10.1002/lnc3.366

Myers, S., & Padgett, J. (2014). Domain generalisation in artificial language learning. *Phonology, 31*(3), 399–433. DOI: https://doi.org/10.1017/S0952675714000207

Nieuwenhuis, R., te Grotenhuis, M., & Pelzer, B. (2017). Weighted Effect Coding for Observational Data with wec. *The R Journal, 9*(1), 477–485. DOI: https://doi.org/10.32614/RJ-2017-017

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: https://doi.org/10.1017/S0140525X00003241

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*(2), 204–238. DOI: https://doi.org/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: https://doi.org/10.1080/23273798.2015.1081703

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. MacNeilage (Ed.), *The Production of Speech* (pp. 189–216). New York, NY: Springer-Verlag. DOI: https://doi.org/10.1007/978-1-4613-8202-7_9

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 101–140). Berlin, Germany: De Gruyter Mouton. DOI: https://doi.org/10.1515/9783110197105.1.101

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46*(2–3), 115–154. DOI: https://doi.org/10.1177/00238309030460020501

Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). *Buckeye corpus of conversational speech* (2nd release) [Corpus]. Columbus, OH: Department of Psychology, Ohio State University.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America, 118*(4), 2561–2569. DOI: https://doi.org/10.1121/1.2011150

Psychology Software Tools, Inc. (2012). E-Prime 2.0 [Computer software].

Reinisch, E., & Holt, L. L. (2014). Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance, 40*(2), 539. DOI: https://doi.org/10.1037/a0034409

Reinisch, E., & Mitterer, H. (2016). Exposure modality, input variability and the categories of perceptual recalibration. *Journal of Phonetics, 55*, 96–108. DOI: https://doi.org/10.1016/j.wocn.2015.12.004

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: https://doi.org/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: https://doi.org/10.1016/j.wocn.2014.04.002

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110*(4), 474–494. DOI: https://doi.org/10.1016/j.wocn.2014.04.002

Samuel, A. G., & Larraza, S. (2015). Does listening to non-native speech impair speech perception? *Journal of Memory and Language, 81*, 51–71. DOI: https://doi.org/10.1016/j.jml.2015.01.003

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

Schmale, R., Cristia, A., & Seidl, A. (2012). Toddlers recognize words in an unfamiliar accent after brief exposure. *Developmental Science, 15*(6), 732–738. DOI: https://doi.org/10.1111/j.1467-7687.2012.01175.x

Schmale, R., Seidl, A., & Cristia, A. (2015). Mechanisms underlying accent accommodation in early word learning: Evidence for general expansion. *Developmental Science, 18*(4), 664–670. DOI: https://doi.org/10.1111/desc.12244

Schuhmann, K. S. (2014). Perceptual learning in second language learners (Doctoral dissertation, Stony Brook University). Stony Brook Theses and Dissertations Collection. https://ir.stonybrook.edu/xmlui/handle/11401/77746

Smith, C. L. (1997). The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics, 25*(4), 471–500. DOI: https://doi.org/10.1006/jpho.1997.0053

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

Stan Development Team (2021). Stan modeling language users guide and reference manual, Version 2.26. https://mc-stan.org

Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., & Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *The Journal of the Acoustical Society of America, 91*(5), 2979–3000. DOI: https://doi.org/10.1121/1.402933

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

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language, 60*(4), 487–501. DOI: https://doi.org/10.1016/j.jml.2009.01.001

Todd, S., Pierrehumbert, J. B., & Hay, J. (2019). Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition, 185*, 1–20. DOI: https://doi.org/10.1016/j.cognition.2019.01.004

Tzeng, C. Y., Nygaard, L. C., & Theodore, R. M. (2021). A second chance for a first impression: Sensitivity to cumulative input statistics for lexically guided perceptual learning. *Psychonomic Bulletin & Review*. DOI: https://doi.org/10.3758/s13423-020-01840-6

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics, 71*, 147–161. DOI: https://doi.org/10.1016/j.wocn.2018.07.008

Veatch, Thomas (1989). Word-final devoicing of fricatives in English. Paper presented at the Linguistic Society of America meeting, Washington, DC.

Warner, N., Good, E., Jongman, A., & Sereno, J. (2006). Orthographic vs. morphological incomplete neutralization effects. *Journal of Phonetics*, *34*(2), 285–293. DOI: https://doi.org/10.1016/j.wocn.2004.11.003

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics*, *32*(2), 251–276. DOI: https://doi.org/10.1016/S0095-4470(03)00032-9

Weatherholtz, K. (2015). Perceptual learning of systemic cross-category vowel variation (Doctoral dissertation). Retrieved from OhioLINK ETD. (osu1429782580)

Wedel, A. B. (2006). Exemplar models, evolution and language change. *The Linguistic Review*, *23*(3), 247–274. DOI: https://doi.org/10.1515/TLR.2006.010

Weirich, M., & Simpson, A. P. (2014). Differences in acoustic vowel space and the perception of speech tempo. *Journal of Phonetics*, *43*, 1–10. DOI: https://doi.org/10.1016/j.wocn.2014.01.001

Wells, J. C. (1982). *Accents of English* (Vol. 1). Cambridge, United Kingdom: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511611759

White, K. S., & Aslin, R. N. (2011). Adaptation to novel accents by toddlers. *Developmental Science*, *14*(2), 372–384. DOI: https://doi.org/10.1111/j.1467-7687.2010.00986.x

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, *30*(5), 945–982. DOI: https://doi.org/10.1207/s15516709cog0000_89

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: https://doi.org/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: https://doi.org/10.1037/xhp0000285

Zhang, X., & Samuel, A. G. (2014). Perceptual learning of speech under optimal and adverse conditions. *Journal of Experimental Psychology: Human Perception and Performance, 40*(1), 200–217. DOI: https://doi.org/10.1037/a0033182

Zheng, Y., & Samuel, A. G. (2019). The relationship between phonemic category boundary changes and perceptual adjustments to natural accents. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(7), 1270–1292. DOI: https://doi.org/10.1037/xlm0000788