PERCEPTUAL LEARNING OF SPEECH SOUNDS: A BIAS FOR 'SIPPER' OVER 'ZIPPER'?

Molly Babel [*1], **Zoe Lawler** [†1], **and Carolyn Norton** [‡1]
[1]Department of Linguistics, University of British Columbia

1 Introduction

Perceptual learning has been proposed as a cognitive mechanism that can account for listeners' adeptness at understanding a wide range of accents and spectral degradations. Our focus is on lexically-guided perceptual learning, whereby a listener's existing phonetic categories are updated with the support of clues from the lexical context [1]. In lexically guided perceptual learning experiments, listeners are presented with, for example, a word like *castle* pronounced not with a canonical /s/, but a fricative that is ambiguous between /s/ and /ʃ/. After exposure to such items, researchers find modifications in the distribution of listeners' /s/ categories, such that more /ʃ/-like sounds are categorized as /s/ or higher word recognition rates for words containing the ambiguous fricative. These changes in behaviour are considered evidence of perceptual learning.

While perceptual learning is a robust phenomenon (see [2] for a review), little is understood about its limits. In this study we test whether listeners show improved learning of phonetically natural shifts in pronunciation. Specifically, we investigate how listeners adapt in the face of a phonetically natural fricative devoicing pattern (e.g., [bɪzi] to *[bɪsi]) compared to a phonetically less-natural fricative voicing pattern (e.g., [mɛdɪsɪn] to *[mɛdɪzɪn]). The aerodynamic constraints of voicing make fricative devoicing patterns more common cross-linguistically.

2 Method

2.1 Subjects

Forty-eight adult volunteers between the ages of 19 - 60 were recruited from the metro Vancouver community ; the majority were students at the University of British Columbia (n = 45). All participants were native speakers of English and reported no speech or hearing impairments.

2.2 Stimulus Materials

Sixty filler sentences and twenty critical sentences containing either the shifted (experimental) or unshifted (control) critical sibilant were determined to be highly-predictable following an online cloze task to assess semantic predictability, and were used as the exposure phase materials. All multisyllabic critical words contained intervocalic alveolar fricatives (/s/ or /z/) and were in sentence-final position. Materials for the lexical decision task consisted of 100 filler words, 60 phonotactically-legal filler nonwords, 20 critical words that were exposed to the listener during the sentence exposure phase, and 40 novel critical words. Efforts were made to ensure that no other sibilant fricatives other than the critical sibilants [s] and [z] were present in any of the materials. However, four words containing an additional sibilant fricative were later identified. With the exception of [h] and [θ], no other fricatives were present in the selected stimuli.

All stimuli were naturally produced by a young adult male speaker of the local Vancouver dialect ; the critical shifted words used in the exposure sentences and in the lexical decision task were not artificially synthesized, following [3]. All materials were recorded using a professional-grade microphone and normalized to 72 dB. No discrepancies in voicing quality were recorded in any of the materials containing the critical sibilant when transcribed by two trained linguists naive to the goals of the study.

2.3 Procedure

Exposure Phase

Half of the /z/ condition participants were assigned to the shifted-/z/ group and exposed to high-predictability sentences containing a sentence-final word that would typically contain [z], (e.g., 'dizzy' [dɪzi]), but was produced with an [s] in place of the [z], rendering [dɪsi] (hereafter referred to as shifted-/z/.) The remaining half of /z/ condition participants were assigned to the control-/z/ group, in which [z] in the critical words in the exposure sentences was produced with the standard pronunciation. The same procedure was used for the shifted-/s/ group and the control-/s/ group. The manipulated variable between subjects was exposure to either the shifted (experimental) or unshifted (control) sibilant.

Eighty randomly ordered sentences (20 critical, 60 filler) were presented consecutively to participants over headphones at a comfortable listening level. Participants were informed to listen carefully to the sentences, but otherwise were not required to do anything while listening to the presented stimuli. Each sentence was separated by a 2000 ms pause.

Lexical Decision Task

A lexical decision task immediately followed the exposure. Participants were presented with a single item over headphones (i.e., nonword, word, exposed shifted-critical word, or novel shifted-critical word) and were asked to classify the item as a 'word' or 'not a word' by 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 orthographically on a computer monitor. Participants were given up to 1500ms to respond. All items were randomized across participants and the total number of trials (220) corresponded with the total number of word

*. molly.babel@ubc.ca
†. zlawler@alumni.ubc.ca
‡. carolyn.norton@alumni.ubc.ca

and nonword items. All critical words presented during the exposure phase were presented in the lexical decision task. Novel words containing either intervocalic /s/ or /z/ were presented with the shifting pattern (voiced to voiceless, or voiceless to voiced) corresponding to the participant ?s condition. Crucially, critical items were consistently presented in their shifted form during the lexical decision task.

3 Results

A glmer model comparing word endorsement rates in the lexical decision task for the /s/-shifted condition found no differences in critical words relative to control. This is shown in Figure 1. In this figure, filler words, exposure critical words, novel critical words, and filler nonwords are presented with the mean proportion of time these items were responded to as real words. Filler words and filler nonwords receive high and low word endorsement rates, respectively. The critical items pattern between the two.

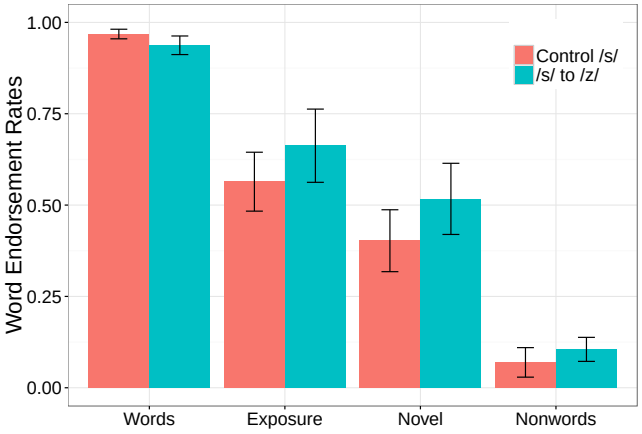

Figure 1: Word endorsement rates by lexical status for shifted /s/-words.

An equivalent model was run for the /z/-shifted condition and found evidence for learning : critical items received higher word endorsement rates in the /z/-shifted condition relative to the control group ($B = 1.31, SE = 0.57, z = 2.3, p = 0.02$). These results are shown in Figure 2. A comparison of performance on exposure words compared to novel words found no significant differences : listeners readily generalized the novel pronunciation shift to items they were not exposed to in the sentence exposure phase.

4 Discussion

Perceptual learning was found in the shifted-/z/ condition, with participants accepting both exposed and novel shifted forms of [z]→[s] items as words after exposure to shifted pronunciations compared to the /z/-shifted control condition. There was no effect of exposure noted for the shifted-/s/ condition, supporting a bias for perceptually learning devoiced fricatives over voiced fricatives. This research offers insight into the limitations of the perceptual learning process.

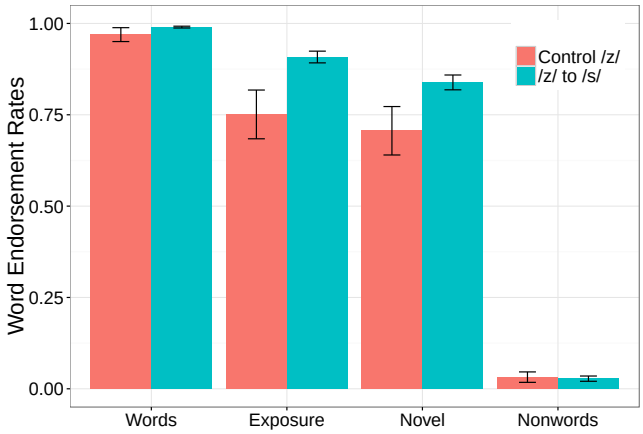

Figure 2: Word endorsement rates by lexical status for shifted /z/-words.

Listeners appear to adapt more readily to phonetically more natural changes, potentially tapping into pre-existing shifts in experienced phonetic distributions. Listeners likely have real-world experience perceiving underlying voiced fricatives as devoiced, as devoicing patterns are common, while the reverse pattern may be less likely ; we intend to confirm this intuition in a corpus study. In the context of this experiment, adaptation to the /z/ to /s/ change may be exploiting those phonetic memory traces.

5 Conclusions

These results suggest that perceptual learning processes may be guided by previous perceptual experiences or phonetic naturalness. Listeners do not learn all ambiguous pronunciations, at least not after a short exposure phase in a lab setting. Understanding the limits and constraints on perceptual learning mechanisms is important in determining which features or aspects of the perceptual learning process are wholly perceptual in nature, as opposed to post-perceptual decision biases. Moreover, understanding which features of the speech signal, elicit more or less robust learning serves as a stepping-stone for researchers to investigate how linguistic factors interact with a multitude of indexical and social cues that broaden or bias our ability to accommodate phonetic variation.

Acknowledgments

We would like to thank the Speech In Context Lab, UBC, and Chase Lawler for your patient dedication to this project.

References

[1] D. Norris, J. M. McQueen, and A. Cutler. Perceptual learning in speech. *Cognitive Psychology.*, 2 :204–238, 2003.

[2] Sven L Mattys, Matthew H Davis, Ann R Bradlow, and Sophie K Scott. Speech recognition in adverse conditions : A review. *Language and Cognitive Processes*, 27(7-8) :953–978, 2012.

[3] Kodi Weatherholtz. *Perceptual learning of systematic cross-category vowel variation.* PhD thesis, The Ohio State University, 2015.