Lexically Guided Perceptual Learning of a Vowel Shift in an Interactive L2 Listening Context

E. Felker, M. Ernestus, M. Broersma

Radboud University Nijmegen, NL

`e.felker@let.ru.nl, m.ernestus@let.ru.nl, m.broersma@let.ru.nl`

Abstract

Lexically guided perceptual learning has traditionally been studied with ambiguous consonant sounds to which native listeners are exposed in a purely receptive listening context. To extend previous research, we investigate whether lexically guided learning applies to a vowel shift encountered by non-native listeners in an interactive dialogue. Dutch participants played a two-player game in English in either a control condition, which contained no evidence for a vowel shift, or a lexically constraining condition, in which onscreen lexical information required them to re-interpret their interlocutor's /ɪ/ pronunciations as representing /ɛ/. A phonetic categorization pre-test and post-test were used to assess whether the game shifted listeners' phonemic boundaries such that more of the /ɛ/-/ɪ/ continuum came to be perceived as /ɛ/. Both listener groups showed an overall post-test shift toward /ɪ/, suggesting that vowel perception may be sensitive to directional biases related to properties of the speaker's vowel space. Importantly, listeners in the lexically constraining condition made relatively more post-test /ɛ/ responses than the control group, thereby exhibiting an effect of lexically guided adaptation. The results thus demonstrate that non-native listeners can adjust their phonemic boundaries on the basis of lexical information to accommodate a vowel shift learned in interactive conversation.

Index Terms: speech perception, lexically guided perceptual learning, non-native listening, interaction

1. Introduction

Given the inherent variability in the acoustic realization of speech sounds, both within and across speakers and dialects, the speech perception system needs to be able to adjust phonemic boundaries dynamically in order to make speech input interpretable. It has been shown that listeners can make use of high-level information from the lexicon to modify their phonemic boundaries in a process known as lexically guided perceptual learning [1]. For example, when an ambiguous fricative /?/ in the spectrum between /f/ and /s/ is repeatedly substituted for word-final /f/ sounds, e.g., replacing /bə'lif/ (from *belief*) with /bə'li?/, listeners are more likely to label the ambiguous sound as /f/ in a subsequent phonetic categorization task, but if the same ambiguous fricative replaces word-final /s/ sounds, e.g., replacing /'notəs/ (from *notice*) with /'notə?/, listeners tend to subsequently categorize the sound as /s/ [2].

To date, most studies on lexically guided perceptual learning focused on ambiguous consonant sounds presented to native listeners in receptive tasks such as passive listening or lexical decision paradigms [see reviews 3, 4]. However, the extent to which this type of learning applies more generally to other classes of speech sounds and in more cognitively demanding listening conditions remains an open question. We present an experiment that investigates whether lexical information can also retune phonemic boundaries for vowel perception in non-native (L2) listening during a task-based dialogue, thereby extending previous research to a different class of speech sounds, a lower-proficiency listener group, and a more naturalistic communicative setting.

Vowels are an interesting test case for lexically guided learning because differences in vowel sounds distinguish many dialects [e.g., 5], making adaptation to vowel variation crucial for communication. Despite this, few studies have specifically tested lexically driven adaptation to vowels. It has been shown that Dutch listeners can use lexical information to retune their perception of an ambiguous Dutch vowel [6], though the learning effects in a phonetic categorization task were highly sensitive to the presentation order of various testing blocks. As for a full vowel shift, one study showed that English listeners exposed to 20 minutes of synthesized speech with systematic front vowel lowering adapted their lexical decision judgments in accordance with the vowel change [7]. Another study showed that both Dutch and native English listeners adapted to a series of lowered front vowel shifts heard in 72 training items in a manipulated English accent [8]. Whether adaptation to a vowel shift can also occur with more limited exposure than in [7] and [8], given the relative instability of the perceptual adaptation in [6], remains to be seen. In theory, we expect a vowel shift to be learnable on the basis of lexical information, but the fact that vowel perception is less categorical than consonant perception may make the observable adaptation effect more subtle.

Relatively little research has studied lexically driven perceptual adaptation in L2 listeners, though it has been demonstrated for Dutch listeners in English with an ambiguous sound between English /l/ and /ɹ/ [9]. Lexically driven perceptual adaptation may be generally more difficult for L2 listeners for numerous reasons: not only because incomplete L2 vocabulary knowledge may lead to differently balanced patterns of lexical activation than for native listeners, but also because increased lexical competition arises from words in their native language during word recognition [10]. In such circumstances, relying on lexical information to disambiguate competing interpretations of a speech sound may be relatively ineffectual. Therefore, while we predict that lexically guided learning will be possible for L2 listeners, the effect may be smaller than what has typically been shown for native listeners.

To our knowledge, the phenomenon of lexically guided perceptual learning has never before been studied in a conversational context. On the one hand, we might expect any kind of perceptual adaptation, including adaptation driven by lexical information, to occur in task-based interaction even more readily than in passive listening since listeners engaged in dialogue may be more motivated to understand their interlocutor and may therefore expend more conscious effort to

comprehend an unfamiliar accent. On the other hand, conversational interaction may be more cognitively demanding as it engages the speech production system; moreover, recent evidence suggests that producing speech during perception training can interfere with learning new sound representations [11]. Therefore, to show that lexically guided perceptual learning still occurs in an interactive task-based dialogue, when exposure to accented speech input is repeatedly interrupted by listeners' own speech production processes, would further attest to the robustness of this type of learning.

In the present experiment, native Dutch-speaking participants played an interactive puzzle-solving computer game called "Code Breaker" with another player whose speech was pre-recorded in order to control phonetic exposure [12, p. EL308]. The pre-recorded speech, belonging to a native English speaker, exhibited an unexpected vowel shift, such that /ɛ/ was pronounced as /ɪ/. As both sounds are part of the Dutch listeners' native phonetic inventory, we expect the vowel shift to be salient. The interactive game was an information gap task that required participants to alternate between solving pattern recognition puzzles aloud and then following their partner's oral instructions to click on certain members of phonological minimal pairs displayed onscreen, the target words being linked to the puzzle shapes. As previous research has shown that as few as ten training items can trigger perceptual learning [e.g., 13], in this experiment we limited the evidence for the vowel shift to a small number of critical target words containing the vowel shift (e.g., "lesson" pronounced as /lɪsən/).

In the lexically constraining condition, the relevant minimal pair contained the target word and a competitor that differed only in a consonant sound (e.g., "lesson" and "lemon"), such that the target word remained the most plausible match to the phonetic input despite the vowel mismatch, thereby promoting perceptual learning. In the control condition, the very same target words were instead displayed alongside an /ɪ/-containing member of the minimal pair (e.g., "listen"), such that the competitor was a perfect match for the phonetic input. In the absence of any corrective feedback about their responses, participants could simply choose the /ɪ/-containing competitor, and no perceptual adaptation to their partner was necessary. A phonetic categorization pre-test and post-test, featuring vowels along a 12-step continuum between the same speaker's /ɛ/ and /ɪ/, was used to evaluate how participants' phonemic boundaries shifted as a result of their experience in the interaction.

2. Method

2.1. Participants

Thirty native Dutch speakers (8 male) aged 18 to 28 years (M = 20.7, SD = 2.8) participated in exchange for course credit or financial compensation. They were raised monolingually and had intermediate to advanced L2 English proficiency.

2.2. Materials

2.2.1. Phonetic categorization in pre-test and post-test

A female native speaker of Middlesbrough English recorded the pseudowords /fɛf/ and /fɪf/, from which the /ɛ/ and /ɪ/ vowels were extracted (/ɛ/: F1 = 734 Hz, F2 = 2036 Hz; /ɪ/: F1 = 557 Hz, F2 = 2186 Hz). In Praat [14], a 12-step continuum ($v1 … v12$) was created between the two endpoints with source-filter vowel resynthesis in which F1, F2, and F3 varied in evenly spaced steps along the continuum (F3 was allowed to vary as it

improved the sound quality). The duration of all resynthesized vowels was set to 164 ms, as in the speaker's original /ɛ/.

The speaker also recorded 9 consonant-group carrier frames (/f_pt/, /f_sk/, /p_f/, /p_ft/, /sp_f/, /sp_p/, /sp_ʃ/, /θ_ʃ/, and /t_ʃ/), each of which was pronounced in 3 versions: surrounding the /ɛ/ vowel (e.g., /fɛpt/), surrounding the /ɪ/ vowel (e.g., /fɪpt/), and preceded by the stressed syllable /pɒp/ and surrounding a schwa (e.g., /'pɒp.fəpt/). All frames were phonotactically legal pseudowords, whether surrounding /ɛ/, /ɪ/, or /ə/. The vowels were then removed from the frames, and the 12 resynthesized vowels of the continuum were spliced into the 9 frames such that for each combination of vowel step and consonant group, 2 of the 3 frame versions were used as carriers, resulting in 216 (12 x 9 x 2) total items. Which 2 of the 3 frame versions were used was systematically shifted throughout the stimuli in a counterbalanced manner (e.g., $v1$ was spliced into the /f_pt/ frames made from /fɛpt/ and /'pɒp.fəpt/, $v2$ was spliced into the /f_pt/ frames made from /fɪpt/ and /fɛpt/, and $v3$ was sliced into the /f_pt/ frames made from /'pɒp.fəpt/ and /fɪpt/). As a result, each of the 9 consonant groups and 3 frame versions occurred the same number of times for each step on the continuum.

2.2.2. Code Breaker interactive game

The Code Breaker game featured 16 critical target words, each spelled with "e" and featuring /ɛ/ in standard English pronunciation. In the lexically constraining condition, the phonological competitor for each target word formed a minimal pair with the target by differing in one consonant sound (e.g., target "set" with competitor "pet"). In the control condition, the competitor differed from target in one vowel by replacing the /ɛ/ with an /ɪ/ (e.g., target "set" with competitor "sit"). In addition to the critical minimal pairs, both conditions contained 64 filler minimal pairs comparable to the critical pairs in word length and frequency and exhibiting various other contrasts: 16 "i"-spelled targets (pronounced with /aɪ/) vs. non-"e"-spelled competitors, 16 non-"i"-spelled targets vs. "e"-spelled competitors, and 32 consonant minimal pairs (16 word-initial and 16 word-final differences) with non-experimental vowels.

Each Code Breaker trial comprised four words: one target word and its phonological competitor (the "foreground" minimal pair) and two unrelated competitors (the "background" minimal pair). Throughout the game, each minimal pair appeared once as the foreground pair and once as the background pair. Fifteen pseudo-randomized stimuli lists were generated, each with different combinations of foreground and background pairs and with trial orders pseudo-randomized such that any trials with critical foreground pairs were spaced at least two trials apart. Four additional fixed word quadruplets were appended to the start of every stimuli list as a practice block such that each list contained 84 trials in total (16 critical trials + 64 filler trials + 4 practice trials). Each list was used once in the lexically constraining condition and once in the control condition, the only difference between the two list versions being the critical minimal pairs.

In addition to the word quadruplets, the Code Breaker game included 84 unique puzzles, one for each trial. Each puzzle was a sequence of five shapes followed by a question mark in place of a missing sixth shape, whose identity could be determined by a pattern in the preceding sequence (e.g., alternating colors). Four puzzles were used for the practice block. The other 80 puzzles were distributed randomly across the trials in each list.

All scripted, pre-recorded speech for the Code Breaker game was recorded by the same speaker as in the phonetic

categorization task. Crucially, a short front vowel shift was introduced in her accent such that the /ɛ/ vowel was pronounced /ɪ/, entailing that all critical target words were pronounced with /ɪ/. This effect was achieved by replacing the target /ɛ/-words in the speaker's script with their phonological /ɪ/-competitors.

The pre-recorded utterances included, for each target word, a multi-word instruction telling the participant which word to click on, sometimes with a disfluency to make the speech sound more natural (e.g., "That's, uh, *chase*", "Okay, so you want *tab*"). There were also several categories of utterances that could be played at any time to react to participants' questions, including affirmative and negative responses, non-lexical backchanneling to indicate listening, reassuring remarks, and task-related phrases. No scripted phrases contained any words with the /ɛ/ or /ɪ/ vowels, whether in their standard or accented pronunciation, to ensure that the only evidence for the vowel shift was the word options displayed onscreen in the lexically constraining condition. Words with /i/ were also excluded from the speech stimuli in order to leave open the possible interpretation from the listener's perspective that the /ɛ/-to-/ɪ/ shift was part of a chain shift rather than a vowel merger.

2.3. Procedures

At the start of the experimental session, participants were told they would be playing two games with a smart computer player that could verbally interact with them. Participants sat in a separate testing booth so they would not notice that the experimenter was controlling the computer player's speech. The session began with the Code Breaker practice block, followed by the phonetic categorization pre-test, the main Code Breaker game, and the phonetic categorization post-test.

2.3.1. Phonetic categorization in pre-test and post-test

In this task, participants had to decide which of two pseudowords the computer player was pronouncing in each trial. At the start of each trial, one audio stimulus was played through a set of headphones. The item's ɛ-representation (e.g., "poptesh") was spelled out on the left side of the screen and its ɪ-representation (e.g., "poptish") on the ride side. Once the participant made a choice by pressing either the left or right button of a button box, the next trial began after a randomly determined interval of 450 to 650 ms. The same 216 stimuli were played in a different randomized order for the pre-test and post-test according to 15 trial lists, each of which was used for two participants: one in the control condition and one in the lexically constraining condition.

2.3.2. Code Breaker interactive game

In each Code Breaker trial, the participant's screen displayed the puzzle sequence above the set of four words. The correct answer to the puzzle and three distractor shapes appeared on the experimenter's screen, with each shape displayed above one of the same four words. The correct-answer shape always appeared together with the target word for that trial.

In each trial, participants' first task was to figure out the pattern in their sequence of shapes and to state what sixth shape would be needed to complete the sequence. In response, the experimenter played a pre-recorded utterance telling the participant which word to click on; this was always the trial's target word, regardless of what shape the participant asked for. Using different numeric keys linked to different categories of utterance types, the experimenter could play pre-recorded

phrases as needed to respond to requests for help, repetition, or clarification, thereby making the game more interactive.

3. Results

3.1.1. Code Breaker interactive game

As expected, participants almost always clicked on the critical target word in the lexically constraining condition, despite the vowel mismatch (mean target word responses = 98.8%, SD = 11.1%), while they almost never chose the critical target word in the control condition (mean target word responses = 4.6%, SD = 21.0%). This difference between conditions was significant: $t(363.96) = 61.48$, $p < 0.001$. Thus, the game effectively caused listeners in the lexically constraining condition to actively choose /ɛ/-words (e.g., "lesson") when hearing /ɪ/ pronunciations (e.g., /lɪsən/) from their partner.

3.1.2. Phonetic categorization in pre-test and post-test

To assess whether the lexically constraining condition led to a shift in phonemic boundaries between /ɛ/ and /ɪ/, we analyzed participants' responses along the 12-step vowel continuum in the phonetic categorization pre-test and post-test using mixed-effects logistic regression models with the binomial link function in the lme4 package in R [15]. Response was the binary dependent variable, participant and consonant frame were random effects, and test time (pre vs. post), condition (control vs. lexically constraining), and vowel step (continuous 1–12) were fixed effects; no random slopes were included in the final model due to lack of convergence.
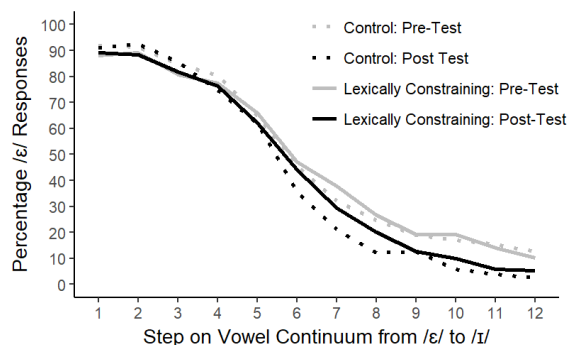

Figure 1: *Phonetic categorization responses.*

The mean responses are shown in Figure 1. The proportion of /ɛ/ responses was significantly higher toward the /ɛ/ end of the continuum ($\beta = -0.74$, SE = 0.02, $p < 0.001$), but contrary to our expectation, fewer /ɛ/ responses were made in the post-test than the pre-test for both groups ($\beta = -0.68$, 0.18, $p < 0.001$). There were significant interactions between vowel step and test time ($\beta = 0.20$, SE = 0.03, $p < 0.001$) and between vowel step and condition ($\beta = 0.08$, SE = 0.03, $p < 0.05$); these indicate that the shift toward /ɪ/ responses in the post-test, as well as the difference between conditions, was greater for vowels closer to the /ɪ/ end of the continuum. The three-way interaction between vowel step, test time, and condition was significant ($\beta = -0.08$, SE = 0.04, $p < 0.05$). Thus, the tendency to shift toward /ɪ/ responses in the post-test for vowels on the /ɪ/ end of the continuum was reduced in the lexically constraining condition relative to the control condition; in other words, listeners in the experimental group were more likely than control-group listeners to label the /ɪ/-like items as /ɛ/ items in the post-test.

To clarify the overall effect, Figure 2 illustrates the mean response data collapsed across all steps of the vowel continuum.
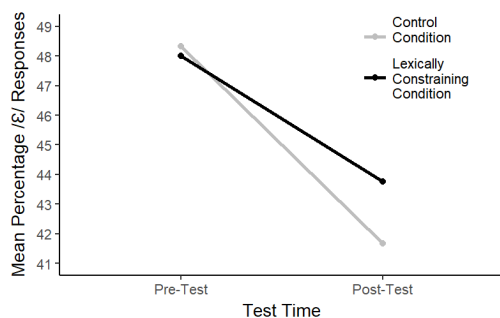

Figure 2: *Mean percentage of /ɛ/ responses by condition and test time.*

Given the unexpected finding that participants in both conditions shifted toward /ɪ/ responses in the post-test, we conducted post-hoc analyses to investigate whether the shift was due to mere exposure to the stimuli. The post-test /ɪ/ shift cannot be explained as a compensation for a pre-test bias in the opposite direction, as the percentage of items labeled as /ɪ/ was already 51.8% in the pre-test (significantly greater than half; $t(6479) = 2.96$, $p < 0.001$). Moreover, when trial number was added as a fixed effect to the model reported previously, it was significant in the opposite direction: in later trials within a test, responses tended more toward /ɛ/ ($\beta = 0.16$, SE = 0.03, $p < 0.001$). Therefore, the overall shift toward /ɪ/ in the post-test appears to result from either the time interval between the tests or from exposure to the speaker's voice during the interaction, rather than from repeated exposure to the vowel continuum.

4. Discussion

The phonetic categorization results include two main findings: first, listeners showed an overall tendency to shift toward /ɪ/ interpretations from the pre-test to the post-test, and second, this effect was attenuated for listeners in the lexically constraining condition: they made more /ɛ/ responses in the post-test than listeners in the control condition, specifically from the middle to the /ɪ/ half of the spectrum. Thus, listeners who were exposed to lexically constraining information show enough perceptual adaptation—learning that their interlocutor's /ɪ/ pronunciations actually represent /ɛ/ words—to partially counteract the larger /ɪ/-directional bias exhibited by both listener groups.

The relative subtlety of the observed lexically guided perceptual learning effect was in line with what we expected, given the continuous nature of vowel perception and the more cognitively demanding listening conditions of an L2 dialogue setting. However, the small effect size may also be due to a combination of the unexpected /ɪ/-shifting bias and several methodological factors. The overall slight bias toward /ɪ/ may reflect a well-documented asymmetry in vowel perception [16]: listeners can more easily discriminate a change from a more central vowel to a more peripheral vowel than vice versa since the more peripheral vowel serves as a perceptual anchor. Thus, when labeling sounds along the /ɛ/-/ɪ/ continuum, it is easier to hear that a given sound is more /ɪ/-like than the previous one, leading to a slight response bias in that direction. Why the preference for the /ɪ/ label increased between the pre-test and post-test might be because of the additional exposure to the speaker's voice during the Code Breaker interaction, in which only 16 /ɪ/ vowels and no /ɛ/ or /i/ vowels were heard. This

manipulation in the set of vowels heard, or simply the additional information about the speaker's realization of other vowels, could have altered how listeners mapped her vowel space.

Several methodological aspects of the present study may also have contributed to the subtlety of the lexically guided adaptation. One is the relatively limited evidence presented for the vowel shift. After a long phonetic categorization pre-test with 216 items spanning the whole spectrum from /ɛ/ to /ɪ/, participants began the Code Breaker game with a well-founded expectation that their interlocutor would produce "e"-words with /ɛ/-like sounds. For participants to change their phonemic boundaries on the basis of just 16 critical accented words in the Code Breaker game thus requires a substantial amount of pre-test exposure to be unlearned. Strengthening the evidence for the vowel shift during the interactive game—whether by increasing the number of critical trials or by incorporating additional accented vowels to form a chain shift as in [7] and [8]—would probably increase the lexically driven adaptation.

Another difference between our methodology and that of traditional lexically guided learning studies is that in our experiment, the learning was driven by lexical constraints built into the task itself, rather than from the listeners' mental lexicon. That is, due to the use of minimal word pairs as the Code Breaker stimuli, all phonetic input listeners received during the game was compatible with real English words, even when those words were absent from the screen. Stronger perceptual learning may occur if listeners were to be exposed to accented words whose /ɪ/ pronunciations did not map onto real words (e.g., "best" pronounced /bɪst/), though this design would preclude the use of a control condition when using a cross-category sound shift rather than an ambiguous sound.

A strength of the present experimental design is that using a pre-test in addition to a post-test made it possible to assess perceptual adaptation within rather than only between listeners. Moreover, we have shown that employing pre-recorded speech in an interactive game is a fruitful method to study speech processing in a more naturalistic setting. In future research, it would be interesting to expand the present design to be able to directly compare the size of the effect for native and non-native listeners, for different types of sounds within the same speaker, or for listening conditions that differ in cognitive load.

5. Conclusions

The goal of this paper was to determine whether lexical information drives perceptual adaptation to a vowel shift for non-native listeners in an interactive context. To that end, participants played an interactive game containing lexical evidence that the interlocutor's /ɪ/ pronunciation should be interpreted as /ɛ/, and their phonemic category boundaries were assessed with a phonetic categorization pre-test and post-test. Relative to the control condition, listeners in the lexically constraining condition were more likely to interpret /ɪ/-like sounds as /ɛ/ in the post-test, despite the fact that both listener groups were biased in the /ɪ/ direction. This shows that lexically guided perceptual adaptation can indeed occur for a vowel shift, from a relatively small amount of evidence, and within the cognitively demanding setting of an L2 task-based interaction, attesting to the robustness of this type of perceptual learning.

6. Acknowledgements

7. References

[1] D. Norris, J. McQueen, A. Cutler, "Perceptual learning in speech." *Cognitive Psychology*, 2003, vol. 47, no. 2, pp. 204-238.

[2] E. Reinisch, L. L. Holt, "Lexically guided phonetic retuning of foreign-accented speech and its generalization." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 40, no. 2, pp. 539-555, 2014.

[3] A. G. Samuel, T. Kraljic, "Perceptual learning for speech." *Attention, Perception, & Psychophysics*, vol. 71, no. 6, pp. 1207-1218, 2009.

[4] M. Baese-Berk, "Perceptual learning for native and non-native speech." *Psychology of Learning and Motivation*, vol. 68, pp. 1-29, 2018.

[5] E. R. Thomas, *An Acoustic Analysis of Vowel Variation in New World English.* Durham, NC: Duke University Press, 2001.

[6] J. McQueen, H. Mitterer, "Lexically-driven perceptual adjustments of vowel categories," in *Proc. ISCA Workshop on Plasticity in Speech Perception*, pp. 233-236, 2005.

[7] J. Maye, R. N. Aslin, M. K. Tanenhaus, "The weckud wetch of the wast: Lexical adaptation to a novel accent." *Cognitive Science*, vol. 32, no. 3, pp. 543-562, 2008.

[8] A. Cooper, A. Bradlow, "Training-induced pattern-specific phonetic adjustments by first and second language listeners." *Journal of Phonetics*, vol. 68, pp. 32-49, 2018.

[9] P. Drozdova, R. van Hout, O. Scharenborg, "Lexically-guided perceptual learning in non-native listening." *Bilingualism: Language and Cognition*, vol. 19, no. 5, pp. 914-920, 2016.

[10] M. L. G. Lecumberri, M. Cooke, A. Cutler, "Non-native speech perception in adverse conditions: A review." *Speech Communication*, vol. 52, no. 11-12, pp. 864-886, 2010.

[11] M. M. Baese-Berk, A. G. Samuel, "Listeners beware: Speech production may be bad for learning speech sounds." *Journal of Memory and Language,* vol. 89, pp. 23-36, 2016.

[12] E. Felker, A. Troncoso-Ruiz, M. Ernestus, and M. Broersma, "The ventriloquist paradigm: Studying speech processing in conversation with experimental control over phonetic input." *Journal of the Acoustical Society of America*, vol. 144, no. 4, pp. EL304-309, 2019.

[13] K. Poellmann, J. M. McQueen, H. Mitterer, "The time course of perceptual learning," in *Proc. 17th Int. Congr. Phonetic Sciences, 2011*, W.-S. Lee & E. Zee, Eds. Hong Kong, 2011, pp. 1618-1621.

[14] P. Boersma, "Praat, a system for doing phonetics by computer." *Glot international*, vol. 5, no. 9/10, pp. 341-345, 2001.

[15] D. Bates, M. Maechler, B. Bolker, S. Walker, "Fitting linear mixed-effects models using lme4." *Journal of Statistical Software*, vol. 67, no. 1, pp. 1-48, 2015.

[16] L. Polka, O-S. Bohn, "Asymmetries in vowel perception." *Speech Communication*, vol. 41, no. 1, pp. 221-231, 2003.