Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Talker-specific perceptual learning about lexical stress: stability over time

Giulio G.A. Severijnen [1], *Verena M. Gärtner* [2], *Runa F.E. Walther* [2], *James M. McQueen* [1,3]

[1] Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen, The Netherlands
[2] Radboud University Nijmegen, The Netherlands
[3] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
giulio.severijnen@donders.ru.nl, verena.gaertner@ru.nl, runa.walther@ru.nl
james.mcqueen@donders.ru.nl

Abstract

Talkers vary in how they speak, resulting in acoustic variability in segments and prosody. Previous studies showed that listeners deal with segmental variability through perceptual learning and that these learning effects are stable over time. The present study examined whether this is also true for lexical stress variability. Listeners heard Dutch minimal pairs (e.g., *VOORnaam* vs. *voorNAAM,* 'first name' vs. 'respectable') spoken by two talkers. Half of the participants heard Talker 1 using only F0 to signal lexical stress and Talker 2 using only intensity. The other half heard the reverse. After a learning phase, participants were tested on words spoken by these talkers with conflicting stress cues ('mixed items'; e.g., Talker 1 saying *voornaam* with F0 signaling initial stress and intensity signaling final stress). We found that, despite the conflicting cues, listeners perceived these items following what they had learned. For example, participants hearing the example mixed item described above who had learned that Talker 1 used F0 perceived initial stress (*VOORnaam*) but those who had learned that Talker 1 used intensity perceived final *stress* (*voorNAAM*). Crucially, this result was still present in a delayed test phase, showing that talker-specific learning about lexical stress is stable over time.

Index Terms: lexical stress, perceptual learning, talker-specific learning, stability of learning

1. Introduction

Word recognition requires correct perception of segmental and suprasegmental information in the acoustic signal. For example, for listeners to correctly perceive the word 'FORbear' (capitalization indicates lexical stress), they must perceive the vowels and consonants in that word, but also identify the correct stress pattern. That is, perceiving the wrong stress pattern would lead to perception of the unintended word ('forBEAR'). While this is a seemingly straightforward process, a challenge for listeners is to deal with differences in how talkers speak, which leads to acoustic variability in segmental and suprasegmental information. The present study examine how listeners deal with such variability in lexical stress, through perceptual learning, and whether these learning effects are stable over time.

In free-stress languages, such as English and Dutch, polysyllabic words contain stress patterns which can distinguish between segmentally overlapping words (e.g., 'FORbear' vs. 'forBEAR'). In most languages, among which Dutch, talkers signal these stress patterns by producing the stressed syllable with a longer duration, larger intensity and higher mean F0 (in accented words) [1]. Listeners, in turn, use these stress patterns in perception to guide word recognition [2], [3], [4], [5], [6].

However, correct perception is challenged by the acoustic variability caused by how different talkers produce lexical stress. For example, in Italian, there are larger mean F0 and duration differences between stressed and unstressed syllables in women compared to men [7]. Also, Tseng et al. [8] found differences in F0 usage for stress production in English between L1 speakers of Taiwan Mandarin and L1 Beijing Mandarin.

On top of these language-specific and gender-based differences, individual talkers of the same gender within the same language and region also vary in how they produce lexical stress [9]. More specifically, Severijnen et al. [9] showed that Dutch participants weighted different cues that signal lexical stress (e.g., F0, intensity, duration, spectral tilt) differently, illustrating the large amount of variability between individual talkers. What's more, talkers seemed to cluster into different groups, depending on which cue was their strongest cue. That is, there were groups of primarily intensity-users and duration-users in the way unaccented words were realized. For accented words, there were groups of primarily intensity-users and F0-users. In sum, these experiments showcase the acoustic variability in lexical stress that listeners have to deal with.

Listeners can in fact deal with such talker differences through perceptual learning [10], [11]. More specifically, in Severijnen et al. [11], Dutch participants completed a learning experiment in which they heard two talkers producing minimal stress pairs (e.g., *VOORnaam* vs. *voorNAAM*, 'first name' vs. 'respectable'). In the learning phase, the two talkers produced the stress pairs using talker-specific cues to lexical stress (e.g., Talker 1 used only F0 and Talker 2 used only intensity; talker-cue mappings were counterbalanced across participants), allowing listeners to learn the talker-specific details specifying which talker used which cue. In a subsequent test phase, participants heard the same talkers producing these minimal stress pairs, but now the words contained two conflicting cues to lexical stress ("mixed items"). For example, one cue (e.g., F0) signaled initial stress (Strong-Weak, SW; e.g., *VOORnaam*) while the other cue signaled final stress (Weak-Strong, WS; e.g., *voorNAAM*). Despite the conflicting information about the stress pattern in the word, participants perceived these words in line with the talker-specific knowledge they had acquired. That is, if participants learned that Talker 1 mainly used F0, and they heard a word produced by Talker 1 in which F0 signaled SW and intensity WS, they were more likely to perceive a SW pattern. In contrast, participants who learned that Talker 1 used intensity were more likely to perceive a WS pattern in the acoustically identical stimulus.

10.21437/SpeechProsody.2024-133

An open question remains whether these learned talker representations regarding lexical stress are stable over time. Evidence from perceptual learning studies on *segments* has previously shown this is the case for fricatives [12], [13], [14]. However, it might be the case that these findings would not translate to perceptual learning of lexical stress since, in Dutch (and e.g. English), there are very few minimal stress pairs. This might lower the necessity for listeners to store these learned representations in memory compared to segments.

The present study therefore replicated the design in Severijnen et al. [11] but also examined the stability of the observed effect. Participants performed the same learning and test phase as in Severijnen et al. [11], with one crucial difference: the test phase was split into two shorter test phases, separated by a 25-minute delay. We chose 25 minutes based on Kraljic & Samuel [13]. To our knowledge, ours is the first study examining stability of perceptual learning about lexical stress. We therefore opted for a relatively short delay. We predicted that we would observe the same findings as in Severijnen et al. [11]: Participants who learned that Talker 1 used F0 were more likely to perceive a SW pattern when the mixed items in the test phase contained F0 signaling SW and intensity signaling WS, and *vice versa* for participants who learned that Talker 1 used intensity. Moreover, based on the findings of stability in segmental learning, we expected this effect to still be present in the second test phase, suggesting that the learned representations are stable over the time delay of 25 minutes.

2. Methods

2.1. Participants

We recruited 37 native talkers of Dutch from the Radboud University participant pool (33 female, 4 male, age range: 18-28, M_{age} = 21.8, SD_{age} = 2.95). All participants gave informed consent and received a monetary reward or course credits. No participant reported having a speaking and/or reading problem. The sample size was estimated based on a power analysis [15], which estimated a power of 1 with 30 participants.

2.2. Materials

We used the stimuli from Severijnen et al. [11]: four bisyllabic (e.g., *VOORnaam* vs. *voorNAAM*, 'first name' vs. 'respectable') and four trisyllabic (e.g., *VOORkomen* vs. *voorKOmen*, 'to appear' vs. 'to prevent') minimal stress pairs spoken by two male talkers. These appeared in sentence-final position in a carrier sentence (*Het woord is*…, 'The word is…'). Moreover, there were two feedback sentences (*Goed, het woord is…* 'Correct, the word is…'; *Fout, het woord is…,* 'Wrong, the word is…'). The stimulus list is available at https://osf.io/z3579/?view_only=1008323ffd564c1a9f1b7e048 0874071.

The word pairs were manipulated by Severijnen et al. [11] using Praat [16], to create two types of stimuli. First, there were 'control items', in which only one cue signaled lexical stress. That is, only F0 or intensity signaled lexical stress while the other cues were set to ambiguous values. These manipulations were performed for both talkers (i.e., Talker 1 using only F0, Talker 2 only intensity, and *vice versa*). There was thus a total of 64 control items: 8 word pairs with 2 patterns (one SW token and one WS token), spoken by 2 talkers, in 2 talker-cue mappings (Talker 1 using F0, Talker 2 using intensity and *vice versa*). These items were used in the learning phase with the aim for listeners to learn talker-specific cue mappings.

Second, there were 'mixed items', which contained two conflicting cues to lexical stress. There were two different patterns of items: words in which F0 signaled a SW pattern while intensity signaled a WS pattern ('*F0_Intensity*') and *vice versa* ('*Intensity_F0*). In total, there were 32 mixed items: 8 word pairs with 2 patterns (*F0_Intensity* and *Intensity_F0*), spoken by 2 talkers. These items were used in the test phase to examine how words with ambiguous stress patterns would be affected by the talker-cue mappings. For details on the stimulus manipulations and stimulus piloting, see Severijnen et al. [11].

2.3. Procedure

The experiment was built and hosted on the Gorilla Experiment Builder (www.gorilla.sc), an online experiment platform, but it was an in-house experiment. We used the Gorilla Experiment Builder to use the same experimental software as in [11].

We followed the same procedure as Severijnen et al. [11]. Participants were randomly assigned to one of the talker-cue mappings (half of the participants heard Talker 1 using F0, Talker 2 using intensity and *vice versa* for the other half). The experiment consisted of a familiarization phase, a learning phase, and a test phase.

2.3.1. Familiarization phase

In the familiarization phase participants were visually presented with the orthographic word forms of the target words, the meaning of the target words, and an example sentence with the target word. Moreover, we auditorily presented the control items of the words (following the talker-cue mapping to which the participant was assigned). This promoted familiarity with the target words before continuing with the experiment.

2.3.2. Learning phase

During the learning phase participants heard the control items embedded in the carrier sentence (e.g., *Het woord is …*), and had to indicate which word they had heard. Participants first saw a fixation cross in the middle of the screen. After 500 ms, we visually presented a cartoon image of the respective talker in that trial to strengthen the link between speech from that talker and the talker's identity. After 300 ms, we presented the two response options (i.e., the two members of the minimal pair) in the lower left lower right corner of the screen (position counterbalanced across participants). Next, 200 ms after seeing the response options (which remained on the screen), participants were auditorily presented with the carrier sentence and target word and instructed to respond at sound offset with button presses ([Z] or [M] for the left and right option respectively) which word they had heard. If participants did not respond within 5 s from sound offset, the trial was recorded as a missing data point. Participants then received feedback (e.g., *Goed, het woord is VOORnaam* or *Fout, het woord is VOORnaam*) and instructed to press the correct button again based on the feedback. After this final response, they heard the correct word once again in isolation. The next trial started 1 s from sound offset of the final word in isolation. The learning phase consisted of 192 experimental trials (8 words × 2 stress patterns × 2 talkers × 6 repetitions) and 8 practice trials that were excluded from statistical analyses.

2.3.3. Test phase

After the learning phase, participants completed the test phase, during which we assessed participants' perception on the mixed items. Crucially, the test phase was divided into two blocks (Block 1: 'immediate test'; Block 2: 'delayed test') with a 25-minute delay between the two. During that delay, participants listened to orchestral classical music and were given line drawings that they could color. We chose these activities because the stimuli did not contain any linguistic information.

The trial structure was similar to that in the learning phase, except for participants not receiving feedback on their responses. The entire test phase (across immediate and delayed test) contained 96 trials with mixed items (8 target words × 2 stress patterns × 2 talkers × 3 repetitions). Additionally, to still provide solid anchors of unambiguous items, we also presented control items on 96 trials (8 target words × 2 stress patterns × 2 talkers × 3 repetitions). Importantly, the control words were consistent with the talker-cue mapping in the learning phase. In contrast, all participants, irrespective of the talker-cue mapping, heard the same mixed items.

We divided the trials over the immediate and delayed tests such that half of the trials was presented in the immediate test, and the other half in the delayed test. Since one repetition of all possible combinations equals 32 trials, we needed to fit 1.5 repetitions (48 trials) into each part. To achieve this, each test phase contained one full repetition, and one repetition that was divided over the two parts. We selected half of the target words and placed all control and mixed items of those words in the immediate test, the other half in the delayed test. Which words appeared in the immediate or delayed test was rotated in four lists, such that we had four different combinations of which words were repeated twice in each block.

3. Results

We first excluded any missing data points (0.7%) and trials with RTs below 100 ms (0.25%). The latter was done because on trials with RTs below 100 ms, the majority of the first syllable was not perceived yet (shortest syllable duration was 171 ms). This resulted in a total of 14,106 trials (10,595 trials on control items, 3,533 trials on mixed items).

We then analyzed the data in the familiarization phase, which showed that all participants knew at least 14/16 words (87.8% knew 16/16 words while 12.2% knew 15/16 words). Severijnen et al. [11] showed that the results were similar for the full dataset compared to a dataset in which the unknown words were excluded. For that reason we included the entire dataset in the analyses.

3.1. Control items

We calculated the proportion of SW responses, separately for the SW and WS control items in the learning phase (SW: 0.8, WS: 0.24) and test phase (SW: 0.79, WS: 0.22). This suggests that in both phases, participants successfully perceived most of the control items.

3.2. Mixed items

The analysis on the mixed items tested whether perception of the mixed items was affected by the learned talker-specific cue mappings picked up in the learning phase. In other words, whether a mixed item with pattern *F0_Intensity* produced by Talker 1 would be more often perceived as SW when participants learned that Talker 1 used F0, compared to when they learned that Talker 1 used intensity (and *vice versa* for the *Intensity_F0* pattern). This was also tested for Talker 2.

This involved testing a three-way interaction between Mapping (Talker 1 using F0, Talker 2 using intensity and *vice versa*), Pattern (*F0_Intensity* or *Intensity_F0*), and Talker (Talker 1 or Talker 2). To simplify the analyses, we followed the procedure in Severijnen et al. [11], and created a new variable that coded for this three-way interaction.

This variable was called Predicted Response, and coded for the expected response (Predicted SW or Predicted WS) on each trial based on the three factors mentioned above. For example, trials with a mixed item containing the pattern *F0_Intensity* produced by Talker 1 resulted in the expected response "Predicted SW" for participants who learned the mapping Talker 1 using F0, but in "Predicted WS" when they learned that Talker 1 used intensity. This was done for every combination of Mapping, Talker, and Pattern.

We then calculated the proportion of SW responses for each level of Predicted Response (Figure 1), which showed that Predicted SW trials were indeed perceived as SW more often than Predicted WS trials. The same pattern was observed in Block 2, while numerically smaller.

The statistical analyses tested the binomial categorization responses (SW coded as 1, WS coded as 0) using a Generalized Linear Mixed model (GLMM) with a logistic linking function in the *lmerTest* package [17] in R [18]. The final model with the best fit to the data contained the following fixed factors: Predicted Response (categorical factor with two levels, deviance coded with Predicted WS coded as -0.5 and Predicted SW coded as 0.5), Talker (categorical predictor with two levels, deviance coded with Talker 1 coded as -0.5 and Talker 2 coded as 0.5), Pattern (categorical predictor with two levels, deviance coded with *F0_Intensity* coded as -0.5 and *Intensity_F0* coded as 0.5), and the interaction between Predicted Response and Talker. Log-likelihood comparisons showed that a model with Block ($\chi2(1) = 1.008$, $p = .179$) or with the interaction between Predicted Response and Block ($\chi2(2) = 1.060$, $p = .589$) did not improve the model fit to the data. The final model further included by-participant random slopes for Talker and Pattern and by-item random slopes for Predicted Response and Talker. The random structure was optimized using Principal Component Analyzes (PCA), which contained the minimally required factors to explain the largest variance [19].

The model showed a significant effect of Predicted Response ($\beta = 0.255$, $SE = 0.088$, $z = 2.882$, $p < .01$), suggesting more SW responses on the Predicted SW trials. In other words, the learned talker-cue mappings affected responses on the mixed items. Moreover, crucially, this effect was present in Block 1 and in Block 2, as evidenced by the lack of an interaction between Predicted Response and Block. Further, we found a marginally significant effect of Talker ($\beta = 0.308$, $SE = 0.164$, $z = 1.874$, $p = .06$), indicating that participants gave overall more SW responses on trials produced by Talker 2. Moreover, we found a marginally significant interaction between Predicted Response and Talker ($\beta = -1.304$, $SE = 0.707$, $z = -1.884$, $p = .06$).

To further examine this interaction, we performed pairwise comparisons using *emmeans* [20]. Results showed a significant effect of Predicted SW for Talker 1 ($\Delta = -0.907$, $SE = 0.364$, $z = -2.49$, $p < .05$), while it was non-significant and numerically in the other direction in Talker 2 ($\Delta = 0.397$, $SE = 0.365$, $z = 1.09$, $p = .27$). This illustrates that the main effect of Predicted Response is driven by Talker 1.

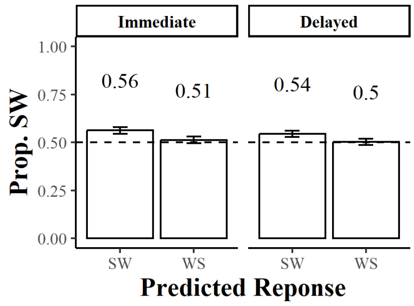

Figure 1: *Categorization responses in the test phase, divided by Predicted Response (SW and WS). The left panel shows the results in the immediate test phase (Block 1), the right panel those in the delayed test phase (Block 2). Error bars represent 95% CIs.*

4. Discussion and conclusions

The present study examined whether learned talker-specific representations about which talker uses which cue to produce lexical stress are stable over time. We found that responses on the mixed items, containing two conflicting cues to lexical stress, were indeed perceived according to the learned cue mappings. The present study thus replicated the results in Severijnen et al. [11], who showed that listeners can use previously learned talker-specific cues to lexical stress in an immediate test phase only. Crucially, the present study not only provided converging evidence for these results but also extended them by showing that the learned representations are stable over a time delay of at least 25 minutes.

These results are in line with previous experiments showing that adaptation to varying productions of segments is stable over longer stretches of time using different paradigms such as lexically guided phonetic retuning [12], [13], [14] and cross-modal priming [21]. The present study shows for the first time that similar learning effects for lexical stress in Dutch are also stable over time. Listeners thus appear to store information in memory not only about how talkers produce segments but also about how talkers signal lexical stress and are able to reactivate those representations at a later point in time.

It should be noted that while the present study replicated the main findings in Severijnen et al. [11], the effect in the present study was mostly driven by Talker 1. That is, the model did not show a significant effect of Predicted Response for Talker 2. In contrast, Severijnen et al. [11] did find a learning effect for both talkers (though it was numerically smaller for Talker 2). While we believe that this difference is probably due to the smaller sample size in the present study, there are two other possible explanations. First, it could be that the stimuli from Talker 2 were less clear (i.e., less successfully induced talker-specific learning of lexical stress). However, while Severijnen et al. [11] reported a numerically smaller effect for Talker 2, this did not differ statistically from the effect for Talker 1, suggesting that in [11] the stimuli were equally successful in inducing perceptual learning. Second, it might be the case that it is more difficult for listeners to store perceptual representations over longer periods of time for two talkers. An argument against this option is that there is no particular reason to believe that listeners should consistently learn the representations only for Talker 1 in the present study (the talker-cue mappings were counterbalanced, trials were presented in randomized order). Therefore, the reason for the difference between the two studies is more likely due to the smaller sample size here.

Despite this, the current findings still provide evidence for the stability of perceptual learning for lexical stress in Dutch. In the present study, we opted for a relatively short delay of 25 minutes, in which participants did not receive any information that might affect the learned representations. We chose this particular delay because this was, to our knowledge, the first experiment to examine the stability of learning for lexical stress. Our main goal was to establish that learned representations for lexical stress are stable over a short period of time without introducing complicating factors in the delay.

Having provided evidence for this, the present study opens up interesting avenues for future research to fully understand the limits of these learning effects. For example, future studies could test whether these effects are stable over longer periods of time, and how they are affected by sleep [12]. They could also examine which kinds of information hinder retention of these memories. Kraljic & Samuel [13] showed that when conflicting information from the same talker is presented to participants, learning effects are attenuated. It would be interesting to examine exactly what kind of information affects perceptual representations for lexical stress. For example, are these representations also affected by conflicting information from prosodically similar talkers? Are listeners only affected by conflicting information from minimal stress pairs or also other words that do not necessarily require stress perception?

Another open question concerns what the nature of the learned representations is. More specifically, McQueen et al. [22] showed that perceptual categories for segments are abstract representations that can be applied to other words. In the present study, this would imply that listeners learned that talkers used specific prosodic cues more strongly than others but in a way that abstracted away from knowledge about the specific minimal-pair words that we used. It would therefore be interesting to examine two things. First, is the learning specific to the minimal-pair words or does it also apply to the processing of lexical stress information in other words? Second, is the learning specific to lexical stress or whether it can also be applied to other types of prosody (e.g., sentential focus) or even to the recognition of segments (e.g., processing of vowel F0 or intensity). In other words, do listeners store talker-specific information separately for segments and different types of prosody, or are these different types of information interdependent? Examining what kind information listeners store about how talkers speak will contribute to a better understanding of the speech recognition system.

In sum, the present study showed that previously learned representations about which cues individual talkers use to produce lexical stress in Dutch are stable over time. This extends previous findings on perceptual learning in segments and opens up interesting new avenues for research on word recognition to examine the exact nature of these representations and under which circumstances they are retained in memory.

5. Acknowledgements

This research was funded by a Donders Centre for Cognition grant, awarded to James McQueen and Hans Rutger Bosker. The stimuli were obtained from Severijnen et al. [11], (see https://osf.io/dczx9/?view_only=44f227db3c134685ad1db9cf46e317f7). The data and analysis scripts are available at https://doi.org/10.34973/x4p3-y641.

6. References

[1] M. Gordon and T. Roettger, "Acoustic correlates of word stress: A cross-linguistic survey," *Linguistics Vanguard*, vol. 3, no. 1, p. 20170007, Aug. 2017, doi: 10.1515/lingvan-2017-0007.

[2] A. Cutler and W. Van Donselaar, "Voornaam is not (really) a Homophone: Lexical Prosody and Lexical Access in Dutch," *Lang Speech*, vol. 44, no. 2, pp. 171–195, Jun. 2001, doi: 10.1177/00238309010440020301.

[3] E. Reinisch, A. Jesse, and J. M. McQueen, "Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately," *Quarterly Journal of Experimental Psychology*, vol. 63, no. 4, pp. 772–783, Apr. 2010, doi: 10.1080/17470210903104412.

[4] A. Cutler, "Forbear is a Homophone: Lexical Prosody Does Not Constrain Lexical Access," *Lang Speech*, vol. 29, no. 3, pp. 201–220, Jul. 1986, doi: 10.1177/002383098602900302.

[5] A. Jesse and J. M. McQueen, "Suprasegmental Lexical Stress Cues in Visual Speech can Guide Spoken-Word Recognition," *Quarterly Journal of Experimental Psychology*, vol. 67, no. 4, pp. 793–808, Apr. 2014, doi: 10.1080/17470218.2013.834371.

[6] S. Sulpizio and J. M. McQueen, "Italians use abstract knowledge about lexical stress during spoken-word recognition," *Journal of Memory and Language*, vol. 66, no. 1, pp. 177–193, Jan. 2012, doi: 10.1016/j.jml.2011.08.001.

[7] A. Eriksson, P. M. Bertinetto, M. Heldner, R. Nodari, and G. Lenoci, "The Acoustics of Lexical Stress in Italian as a Function of Stress Level and Speaking Style," presented at the Proc. Interspeech 2016, Sep. 2016, pp. 1059–1063. doi: 10.21437/Interspeech.2016-348.

[8] C. Tseng, C. Su, and T. Visceglia, "Levels of lexical stress contrast in english and their realization by L1 and L2 speakers," in *2013 International Conference Oriental COCOSDA held jointly with 2013 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE)*, Gurgaon, India: IEEE, Nov. 2013, pp. 1–5. doi: 10.1109/ICSDA.2013.6709853.

[9] G. G. A. Severijnen, H. R. Bosker, and J. M. McQueen, "How do 'VOORnaam' and 'voorNAAM' differ between talkers? An corpus analysis of individual talker differences in lexical stress in Dutch," presented at the 18th conference on Laboratory Phonology (LabPhon18), 2022.

[10] G. G. A. Severijnen, H. R. Bosker, V. Piai, and J. M. McQueen, "Listeners track talker-specific prosody to deal with talker-variability | Elsevier Enhanced Reader," *Brain Research*, vol. 1769, 2021, doi: https://doi.org/10.1016/j.brainres.2021.147605.

[11] G. G. A. Severijnen, G. Di Dona, H. R. Bosker, and J. M. McQueen, "Tracking talker-specific cues to lexical stress: Evidence from perceptual learning.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 49, no. 4, pp. 549–565, Apr. 2023, doi: 10.1037/xhp0001105.

[12] F. Eisner and J. M. McQueen, "Perceptual learning in speech: Stability over time," *The Journal of the Acoustical Society of America*, vol. 119, no. 4, pp. 1950–1953, Apr. 2006, doi: 10.1121/1.2178721.

[13] T. Kraljic and A. G. Samuel, "Perceptual learning for speech: Is there a return to normal?," *Cognitive Psychology*, vol. 51, no. 2, pp. 141–178, Sep. 2005, doi: 10.1016/j.cogpsych.2005.05.001.

[14] Y. Zheng and A. G. Samuel, "Flexibility and stability of speech sounds: The time course of lexically-driven recalibration," *Journal of Phonetics*, vol. 97, p. 101222, Mar. 2023, doi: 10.1016/j.wocn.2023.101222.

[15] L. Kumle, M. L.-H. Võ, and D. Draschkow, "Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R," *Behav Res*, May 2021, doi: 10.3758/s13428-021-01546-0.

[16] P. Boersma and D. Weenink, "Praat: doing phonetics by computer." Feb. 26, 2019. [Online]. Available: www.praat.org

[17] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest Package: Tests in Linear Mixed Effects Models," *J. Stat. Soft.*, vol. 82, no. 13, 2017, doi: 10.18637/jss.v082.i13.

[18] R Core Team, "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria, 2020. [Online]. Available: https://www.R-project.org/

[19] D. Bates, R. Kliegl, S. Vasishth, and H. Baayen, "Parsimonious Mixed Models," *arXiv:1506.04967 [stat]*, 2015, Accessed: Jul. 23, 2021. [Online]. Available: http://arxiv.org/abs/1506.04967

[20] R. Length, "emmeans: Estimated Marginal Means, aka Least-Square." 2022. [Online]. Available: http://cran.r-project.org/package=emmeans

[21] M. J. Witteman, N. P. Bardhan, A. Weber, and J. M. McQueen, "Automaticity and Stability of Adaptation to a Foreign-Accented Speaker," *Lang Speech*, vol. 58, no. 2, pp. 168–189, Jun. 2015, doi: 10.1177/0023830914528102.

[22] J. M. McQueen, A. Cutler, and D. Norris, "Phonological abstraction in the mental lexicon," *Cognitive Science*, vol. 30, no. 6, pp. 1113–1126, 2006, doi: 10.1207/s15516709cog0000_79.