DOES SCHWA HAVE AN AUDITORY TARGET? AN ALTERED AUDITORY FEEDBACK STUDY

Sarah Bakst and Caroline A. Niziolek

University of Wisconsin–Madison
sbakst@wisc.edu, cniziolek@wisc.edu

ABSTRACT

Schwa is cross-linguistically described as having a variable target (e.g. Koopmans-van Beinum 1994). The present study examines whether speakers are sensitive to auditory feedback when producing schwa. When speakers hear themselves producing a version of their speech where formants have been altered, they will change their motor plans online so that their altered feedback is a better match to the target (e.g. Houde & Jordan 1998). If schwa has no target, then feedback mismatches may not drive a change in production.

In this experiment, participants spoke disyllabic iambs and trochees where the auditory feedback of F1 was raised by 100 mels. Both stressed and unstressed syllables showed compensatory decreases in F1, with comparable levels of adaptation in schwa and stressed vowels. However, across different vowels, the magnitude of adaptation in schwa was highly dependent on that of the heterosyllabic vowel, consistent with the idea that schwa is highly assimilatory.

Keywords: phonological categories, stress, reduction, speech production.

1. INTRODUCTION

We listen to ourselves while we are talking, and we use this auditory feedback to ensure that our productions match our auditory expectations. In experiments that alter subjects' auditory feedback in real time, speakers change their speech in opposition to the alteration. When the feedback alteration is unpredictable, speakers *compensate* by adjusting their speech acoustics to counteract the alteration within a single syllable [24]. When feedback is altered consistently, speakers *adapt*, learning to adjust their motor plans in a temporary remapping that persists even after feedback is returned to normal [11]. Compensation and adaptation are thus evidence that the speech target is to some degree acoustic or auditory: when speakers hear themselves producing speech that does not match the target, they change their articulation so that their productions are a better match to the target.

Speakers are sensitive to a variety of acoustic features of the speech target, including amplitude [2], pitch [6], formant frequencies [11, 24], and relationships between syllable-timing and formant frequencies [8]. Further, the degree to which speakers may compensate for altered feedback is not determined solely by the magnitude of acoustic difference from the auditory target. Speakers adapt more as the perturbation threatens to produce a different category, indicating that a speaker's categorical perceptual boundaries modulate response to altered feedback [21]. Further, speakers may not respond equally to perturbations in all vowels [15, 22]. Taken together, these results indicate that adaptation, and more generally the way that speakers gauge whether they have reached their speech targets, is at least in part dependent on phonology.

Here, we test whether syllable stress affects adaptation, particularly in schwa ([ə]). In English, unstressed vowels reduce, or take on a phonetic form that is qualitatively different from the full form that emerges if the vowel occurs in a stressed or more prominent position in a word. Unstressed vowels are typically produced closer to the center of the vowel space and have shorter durations than stressed vowels [17], and many unstressed vowels surface as [ə].

The phonetic and phonological representation of schwa is debated cross-linguistically. Phonetically, schwa is observed to be highly variable in Dutch [14, 3] and English, possibly due to coarticulation [10] rather than random variation. This variability has provided evidence for the phonological underspecification of schwa; a study of British English determined schwa to be specified for [height] but not [backness] [13]. However, there is some evidence that, at least in certain contexts, schwa may have a specified target. X-ray data of articulation of schwa in non-words of the form [ˈpVpəpVp] suggests that it may be possible to define a specific *average* articulatory target for schwa by calculating a mean tongue-body position [5]. The same study found that phonetic context did not predict tongue-body position for any given schwa, and that the position appeared to be "warped by an independent schwa component." Further, in some dialects of English, schwa's target may depend on word position, where word-final schwas may be more central and non-final schwas may be higher [9].

While the effect of stress on adaptation to an altered formant has not been directly tested, stress may have an effect on compensation in the suprasegmental tier. In an altered auditory feedback experiment employing both upward and downward shifts of f_0 [19], native German speakers repeated the nonsense string [tatatas] with primary stress on either the first or second syllable. Compensation was highly dependent on syllable position, but initial syllables only showed compensation when they were stressed, so this may have been an effect of the additional length that accompanies stress. The effect of stress on adaptation is therefore not entirely clear.

When speakers adapt to altered feedback, they change the way they are speaking so that they hear themselves producing what they expected to hear: the altered feedback introduces an "error" that speakers correct for. If schwa has a variable target, then adaptation may not occur. Without a stable acoustic or articulatory target, the altered feedback may not produce a mismatch that speakers must correct for. In this study, we use a sensorimotor adaptation paradigm to examine whether an increase in F1 feedback drives a decrease in the produced F1 of schwa that is comparable to that of stressed vowels.

2. METHODS

2.1. Participants

Seventeen (fifteen female) students at the University of Wisconsin–Madison participated in this study. All procedures were approved by the Institutional Review Board at the University of Wisconsin–Madison. Participants were paid $10/hour or received course extra credit. The experiment lasted 45 minutes on average.

2.2. Stimuli

Stimuli were chosen to address foot type, syllable order, and vowel quality: the pair "beta" [ˈbeiɾə] and "abate" [əˈbeit˺], the pair "meta" [ˈmɛɾə] and "adept" [əˈdɛpt], and to disentangle the effect of vowel quality apart from stress, "above" [əˈbəv]. Here, we use the phonetic symbol [ə] in both syllables of "above" to indicate that these vowels are acoustically similar ($<$ 40 Hz difference in F1 and F2 in our data). The local dialect of English has no words of the form ˈ(C)əCə, so it was not possible to additionally extricate the role of syllable position for this vowel. Stimuli were randomized within 20-trial blocks.

2.3. Procedure

Participants were seated in front of a computer and wore a head-mounted microphone and circumaural headphones. In each trial, one of the five stimulus words was pseudorandomly selected and displayed on the screen, and participants read it aloud. As they spoke, they heard their feedback over the headphones. After each 20-trial block, participants received a self-timed break.

The alteration occurred in six phases. In the first baseline phase (*pre-task*: 50 trials), feedback was not shifted, and participants heard noise that masked their feedback (77 dB). In the second baseline phase (*baseline*: 110 trials), noise (55 dB) was mixed with feedback so that participants could hear their own unaltered feedback over headphones, but their bone conductive and ambient hearing was masked. During the *ramp* phase (20 trials), a +5 mel perturbation was applied to F1, which linearly increased throughout the phase so that speakers were gradually acclimated to a +100 mel shift. During the *hold* phase (250 trials), the +100 mel perturbation was sustained. A post-perturbation noisy phase (*post-task*: 50 trials) identical to the pre-task phase tested how much speakers had adapted their motor plans by again masking auditory feedback. Finally, a *washout* phase (20 trials) identical to the baseline phase re-acclimated speakers to their normal feedback. All feedback resynthesis was done in Audapter [7, 23].

2.4. Analysis

For each spoken trial, vowel onsets and offsets were manually marked to delineate the first and second syllables. In each vowel, F1 and F2 were tracked every 5 ms using Praat [4] via the wave_viewer analysis package [20]. Formant values from the middle 20% of each vowel were averaged to obtain single values to represent the vowel.

The analysis considers how F1 changed during the hold phase in comparison with the baseline values. A separate baseline was computed for each word and syllable (e.g. separate baselines for the first and second syllables of "abate"). The baselines were calculated as the F1 mean across trials of a given word and syllable in the baseline phase. The means were subtracted from their matching vowels in the hold phase to determine the change in F1 that occurred for each vowel in each utterance. For example, the normalized F1 of the schwa in "abate" was determined by subtracting the mean of all schwas occurring in the word “abate” in the baseline phase from the computed F1 values of each schwa occurring in the word “abate” in the hold phase.

3. RESULTS

F1 values normalized to the baseline phase are displayed in Figures 1 and 2. Over the course of the experiment, participants decreased their F1 by an average of 39 Hz in opposition to the F1 increase in their auditory feedback. This adaptive shift in vowel production was found for both initial and final syllables (Figure 1) as well as for both stressed and unstressed vowels in both syllable positions (Figure 2).

Figure 1: Average of all vowels and stresses across subjects, normalized to baseline and averaged in 10-trial bins in syllable 1 (left) and syllable 2 (right). Vertical lines indicate phase divisions. Green patches indicate trials with masking noise.

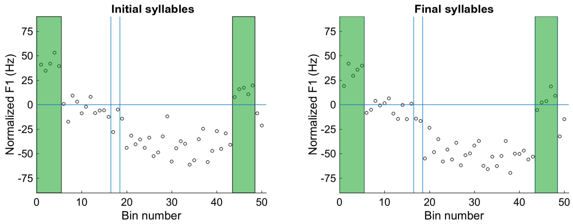
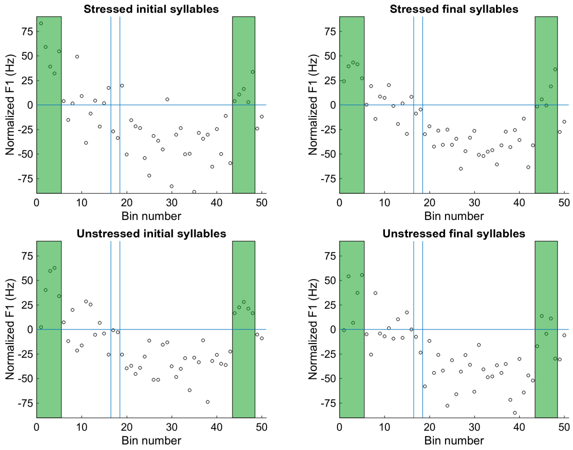


Figure 2: Adaptation by stress and syllable position. Trials were averaged within 50 equal-sized bins. Same patching as in Figure 1.



An ANOVA was run in Matlab [18] predicting baseline-normalized F1 during the entire hold phase with subject as a random factor and with stress, syllable position, and vowel quality as fixed effects. In all analyses, unstressed vowels were labeled with the vowel quality of the stressed vowel in that word; for example, schwa in “adept” was labeled as unstressed [ɛ].

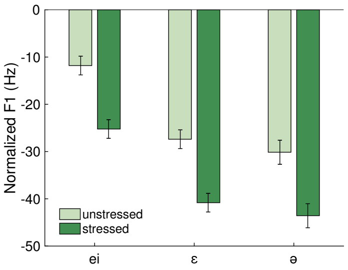
Stress was a significant predictor in the model ($p < 0.001$). Syllable position and quality of the stressed vowel (both $p < 0.001$) were also significant predictors.

Given these main effects, Tukey post hoc comparisons were performed to investigate effect direction as well as within-category differences in marginal means of sub-categories (e.g. stress while holding syllable position and vowel quality constant). Decreases in F1 were significant for both stressed and unstressed vowels in both syllable positions, evidence of across-the-board adaptation to the formant alteration. Participants lowered F1 an average of 6 Hz more in unstressed vowels than in stressed vowels, and 8 Hz more in final than initial syllables. For the three stressed vowels under investigation, the magnitude of F1 lowering in [ei] was found to be 7 Hz less than in [ɛ], and this magnitude in [ɛ] was 7.5 Hz less than in [ə]. Within all three vowel contexts, the unstressed vowel adapted significantly more than the stressed vowel.

While F1 changes differed significantly between stressed and unstressed syllables, the quality of the stressed vowel exerted a greater effect on the magnitude of that change than stress alone: F1 changes in the unstressed vowel patterned with F1 changes in the heterosyllabic vowel. Thus the unstressed vowels in “abate” and “beta” showed evidence of adaptation of a similar magnitude to that of the stressed vowels in those words, and this adaptation was significantly less than the adaptation of the unstressed vowels in [ɛ] words, which was also significantly less than the adaptation of the unstressed vowel in the [ə] word (“above”).

Finally, we ran a separate analysis to determine the effects of online compensation by considering the change in F1 during the noise-masked post-task phase. Because there were substantial differences in F1 between pre/post-task and main conditions (see Figures 1 and 2), the pre-task phase served as a base-

Figure 3: Mean change in F1 from pre- to post-task for all vowel and stress conditions. Standard error bars shown.



line for these trials, which were also produced in noise. Productions were re-baselined by word and syllable by subtracting the pre-task baseline. An ANOVA with the same fixed and random effects showed significant main effects of both stress and vowel ($p < 0.001$), but the effect of syllable was eliminated ($p = 0.12$). Post hoc Tukey tests showed that decreases in F1 were, on average, 13 Hz greater in unstressed vowels than in stressed vowels. Again, adaptation in unstressed vowels patterned with the stressed heterosyllabic vowel, and again adaptation was significantly greater in the unstressed vowel in all words. Mean baseline-normalized F1 for all vowel and stress conditions is shown in Figure 3.

4. DISCUSSION

It was hypothesized that if schwa does not have a target, speakers would not adapt to the altered auditory feedback during the production of schwa as much as during the production of a stressed vowel. This experiment found that unstressed vowels adapted significantly more than stressed vowels, albeit by a difference of 6 Hz. The reliability of the adaptation in schwa alone would suggest that schwa does indeed have a target. However, adaptation in schwa was significantly affected by the quality of the vowel in the stressed syllable, suggesting that adaptation in schwa is dependent on phonetic context.

The results suggest that schwa must not be entirely targetless, but they also raise a question about how the target is phonetically defined, given that the adaptation in schwa was dependent on the phonetic environment. Schwa may not have an intrinsic target, but it may acquire a target from surrounding context.

One theory considers lexical effects. During this experiment, a constant 100 mel upward shift was applied to F1, and the same shift was applied to every vowel in every word. It is possible that size of adaptation is planned at the word level rather than vowel level. Under this theory, the stressed vowel determines both the amount of adaptation and the target for the unstressed vowel. This theory alone would not account for greater changes to unstressed than to stressed vowels.

Another possibility is that, rather than acquiring a specific target from surrounding context, schwa is susceptible to coarticulatory pressures or attracted to nearby stressed vowels. Under this second theory, the adaptation that occurs in schwa is not due to mismatch between auditory feedback and intrinsic target, but rather that schwa shifts in the direction of the preceding or upcoming adapted vowel. That is, adaptation is planned in the heterosyllable, and schwa is adapted in that direction. This is consistent with the finding that schwa is subject to coarticulatory pressures from either a preceding or following stressed vowel [10]. However, this theory, too, does not account for the greater adaptation seen in schwa.

When speakers adapt to altered auditory feedback, the *acoustic* mismatch is not the only force determining adaptation. Speakers are also sensitive to somatosensory information [1] and speakers likely differ in how they weight auditory and somatosensory feedback, with some responding more to auditory and some responding more to somatosensory feedback mismatches [12, 16]. Articulation in schwa has little constriction at all, and speakers who change their articulation to adapt to altered auditory feedback may be less sensitive to the somatosensory mismatch that results. If adaptation is determined at the lexical level, and if the somatosensory feedback mismatch is indeed less informative for schwa, then this type of feedback may not interfere with the adaptation in schwa. These differences in tactile feedback may also explain the between-vowel differences observed in the stressed vowels, where the most constricted vowel [ei] had the smallest change in F1, and [ɔ] had the greatest change.

Finally, the elimination of the syllable position effect for words produced in loud masking noise suggests that the increased adaptation in final syllables was due to online compensation that occurred while speaking rather than trial-to-trial adaptation. That is, in the hold phase, participants could take advantage of auditory feedback during the first syllable to adjust their F1 in the second syllable.

5. CONCLUSION

This study found significant adaptation to altered auditory feedback in schwa. The quality of the stressed vowel in the same word was a predictor of the magnitude of adaptation in schwa: adaptation in an unstressed syllable closely followed adaptation in the stressed vowel in the same word. This may suggest either that schwa does not have an intrinsic target, or that schwa is highly assimilatory. Future studies will be designed to distinguish between these theories by applying different alterations to schwa and other syllables, or by using stimuli with multiple stresses.

6. REFERENCES

[1] Abbs, J. H., Gracco, V. L. 1984. Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology* 51(4), 705–723.

[2] Bauer, J. J., Mittal, J., Larson, C. R., Hain, T. C. 2006. Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. *The Journal of the Acoustical Society of America* 119(4), 2363–2371.

[3] van Bergem, D. R. 1994. A model of coarticulatory effects on the schwa. *Speech Communication* 14(2), 143–162.

[4] Boersma, P., Weenink, D. Praat: doing phonetics by computer [computer program].

[5] Browman, C., Goldstein, L. 1994. "Targetless" schwa: an articulatory analysis. In: Docherty, G. J., Ladd, D. R., (eds), *Papers in Laboratory Phonology II Gesture, Segment, Prosody.* Cambridge University Press 26–67.

[6] Burnett, T. A., Freedland, M. B., Larson, C. R. 1998. Voice F0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America* 103, 3153–3161.

[7] Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., Perkell, J. S. Dec. 8-12 2008. A system for online dynamic perturbation of formant frequencies and results from perturbation of the mandarin triphthong /iau/. *Proceedings of the 8th Intl. Seminar on Speech Production* Strasbourg, France. 65–68.

[8] Cai, S., Ghosh, S. S., Guenther, F. H., Perkell, J. S. 2011. Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing. *The Journal of Neuroscience* 31(45), 16483–16490.

[9] Flemming, E., Johnson, S. 2007. Rosa's roses: reduced vowels in American English. *Journal of the International Phonetic Association* 37(1), 83–96.

[10] Fowler, C. A. 1981. Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research* 24, 127–139.

[11] Houde, J. F., Jordan, M. I. 1998. Sensorimotor adaptation in speech production. *Science* 279, 1213–1216.

[12] Katseff, S., Houde, J. F., Johnson, K. 2012. Partial compensation for altered auditory feedback: a tradeoff with somatosensory feedback? *Language and Speech* 55(2), 295–308.

[13] Kondo, Y. 1994. Phonetic underspecification in schwa. *Third International Conference on Spoken Language Processing* 311–314.

[14] Koopmans-vanBeinum, F. J. 1994. What's in a schwa? *Phonetica* 51, 68–79.

[15] Kothare, H., Raharjo, I., Ranasinghe, K., Ramanarayanan, V., Parrell, B., Houde, J., Nagarajan, S. August 2018. Sensorimotor adaptation in speech is sensitive to vowel targets of altered feedback. Society for the Neurobiology of Language.

[16] Lametti, D. R., Nasir, S. M., Ostry, D. J. 2012. Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *The Journal of Neuroscience* 32(27), 9351–9358.

[17] Lehiste, I. 1976. *Contemporary issues in experimental phonetics* chapter Suprasegmental Features of Speech, 225–239. New York: Academic Press.

[18] MATLAB, 2018. *version 9.2 (R2017b).* Natick, Massachusetts: The MathWorks Inc.

[19] Natke, U., Kalveram, K. T. 2001. Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables. *Journal of Speech, Language, and Hearing Research* 44, 577–584.

[20] Niziolek, C. Jan 2015. wave_viewer: First release.

[21] Niziolek, C. A., Guenther, F. H. 2013. Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *The Journal of Neuroscience* 33(29), 12090–12098.

[22] R.Lametti, D., J.Smith, H., E.Watkins, K., M.Shiller, D. 2018. Robust sensorimotor learning during variable sentence-level speech. *Current Biology* 28(19), 3106–3113.

[23] Tourville, J. A., Cai, S., Guenther, F. H. June 2-7 2013. Exploring auditory-motor interactions in normal and disordered speech. *Proceedings of Meeting on Acoustics. 9:060180.* Presented at 165th Meeting of the Acoustical Society of America Montreal, Quebec, Canada.

[24] Tourville, J. A., Reilly, K. J., Guenther, F. H. 2008. Neural mechanisms underlying auditory feedback control of speech. *NeuroImage* 39(3), 1429–1443.