

Effects of syllable stress in adaptation to altered auditory feedback in vowels

Sarah Bakst, and Caroline A. Niziolek

Citation: [The Journal of the Acoustical Society of America](#) **149**, 708 (2021); doi: 10.1121/10.0003052

View online: <https://doi.org/10.1121/10.0003052>

View Table of Contents: <https://asa.scitation.org/toc/jas/149/1>

Published by the [Acoustical Society of America](#)

ARTICLES YOU MAY BE INTERESTED IN

[Sensorimotor adaptation of speech depends on the direction of auditory feedback alteration](#)

[The Journal of the Acoustical Society of America](#) **148**, 3682 (2020); <https://doi.org/10.1121/10.0002876>

[Individual variability in auditory feedback processing: Responses to real-time formant perturbations and their relation to perceptual acuity](#)

[The Journal of the Acoustical Society of America](#) **148**, 3709 (2020); <https://doi.org/10.1121/10.0002923>

[The effect of Mandarin listeners' musical and pitch aptitude on perceptual learning of Cantonese level-tones](#)

[The Journal of the Acoustical Society of America](#) **149**, 435 (2021); <https://doi.org/10.1121/10.0003330>

[Validating a psychoacoustic model of voice quality](#)

[The Journal of the Acoustical Society of America](#) **149**, 457 (2021); <https://doi.org/10.1121/10.0003331>

[Reviewers of Manuscripts, 2020](#)

[The Journal of the Acoustical Society of America](#) **149**, 737 (2021); <https://doi.org/10.1121/10.0003388>

[Temporal cues to onset voicing contrasts in Australian English-speaking children](#)

[The Journal of the Acoustical Society of America](#) **149**, 348 (2021); <https://doi.org/10.1121/10.0003060>



**Advance your science and career
as a member of the**

ACOUSTICAL SOCIETY OF AMERICA

LEARN MORE



Effects of syllable stress in adaptation to altered auditory feedback in vowels

Sarah Bakst^{a)} and Caroline A. Niziolek^{b)}

Communication Sciences and Disorders, University of Wisconsin–Madison, Madison, Wisconsin 53705, USA

ABSTRACT:

Unstressed syllables in English most commonly contain the vowel quality [ə] (schwa), which is cross-linguistically described as having a variable target. The present study examines whether speakers are sensitive to whether their auditory feedback matches their target when producing unstressed syllables. When speakers hear themselves producing formant-altered speech, they will change their motor plans so that their altered feedback is a better match to the target. If schwa has no target, then feedback mismatches in unstressed syllables may not drive a change in production. In this experiment, participants spoke disyllabic words with initial or final stress where the auditory feedback of *F*₁ was raised (Experiment 1) or lowered (Experiment 2) by 100 mels. Both stressed and unstressed syllables showed adaptive changes in *F*₁. In Experiment 1, initial-stress words showed larger adaptive decreases in *F*₁ than final-stress words, but in Experiment 2, stressed syllables overall showed greater adaptive increases in *F*₁ than unstressed syllables in all words, regardless of which syllable contained the primary stress. These results suggest that speakers are sensitive to feedback mismatches in both stressed and unstressed syllables, but that stress and metrical foot type may mediate the corrective response. © 2021 Acoustical Society of America. <https://doi.org/10.1121/10.0003052>

(Received 5 June 2020; revised 10 December 2020; accepted 11 December 2020; published online 28 January 2021)

[Editor: Anders Lofqvist]

Pages: 708–719

I. 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 auditory feedback in real time, speakers change their speech in opposition to the alteration. When the auditory feedback is altered, speakers *compensate* by adjusting their speech acoustics to counteract the alteration within a single syllable (Tourville *et al.*, 2008). When auditory feedback is altered consistently during an experiment, speakers *adapt*, learning to adjust their motor plans in a temporary remapping that persists even after auditory feedback is returned to normal (Houde and Jordan, 1998). Compensation and adaptation are thus evidence that auditory information is important for assessing the attainment of a speech target: when speakers hear themselves producing speech that does not match their expectations, they change their articulation so that their perceived productions are a better match to the target.

Experiments using the altered auditory feedback paradigm, where speakers hear their own speech altered in near-real time, have shown that speakers are sensitive to a variety of acoustic features of the speech target, including formant frequencies (Houde and Jordan, 1998; Tourville *et al.*, 2008) and relationships between syllable timing and formant frequencies (Cai *et al.*, 2011). Speakers respond to formant perturbations in different vowels, though not necessarily equally across the vowel space (Lametti *et al.*, 2018;

Mitsuya *et al.*, 2015). Here, we use this paradigm to investigate how speakers assess their productions of the English vowel [ə] (schwa). 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 (e.g., *atom* vs *atomic*). Unstressed vowels are typically produced closer to the center of the vowel space and have shorter durations than stressed vowels (Lehiste, 1976), and many of these vowels are pronounced as [ə] (Flemming and Johnson, 2007).

Unstressed schwa is unlike other vowels in that its phonetic and phonological representations are debated cross-linguistically. Phonetically, schwa is observed to be highly variable in Dutch and English, and much of this variability is due to coarticulation (Fowler, 1981; van Bergem, 1994), particularly with respect to the second formant frequency (Koopmans-van Beinum, 1994), 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] (Kondo, 1994). However, there is some evidence that, at least in certain contexts, schwa may have a specified target. X-ray data of the 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 of all other vowels produced by a speaker (Browman and Goldstein, 1994). However, the same study found that phonetic context did not predict tongue-body position for any individual schwa token, and that the tongue-body position

^{a)}Electronic mail: sbakst@wisc.edu, ORCID: 0000-0003-3502-6158.

^{b)}ORCID: 0000-0002-6085-1371.

for this vowel appeared to be “warped by an independent schwa component.” The articulation of schwa is therefore not purely the product of coarticulation. In some dialects of English, schwa has been observed to have multiple targets depending on word position, where word-final schwas may be more central and non-final schwas may be higher (Flemming and Johnson, 2007), further suggesting that some identifiable target(s) may exist for this vowel.

The questions surrounding the acoustic target for schwa seem to apply only to the unstressed variant. American English also has a central vowel that occurs in stressed positions, the STRUT vowel (“strut,” “above”) (Wells, 1982). Typically this is denoted with [ʌ] in the international phonetic alphabet (IPA), but here we will refer to it as *stressed* [ə] to highlight the overlap in acoustic space between the STRUT vowel and unstressed schwa (Szigetvári, 2018).

Higher-level phonological factors are known to modulate the size of the opposition to an auditory feedback shift in the formant frequency domain, but it has not been established whether stress is among them. For example, speakers compensate more when a perturbation threatens to produce a different phonological category, indicating that a speaker’s categorical perceptual boundaries modulate responses to altered auditory feedback (Niziolek and Guenther, 2013). Further, speakers’ adaptation differs depending on their native language, even for similar vowels (Mitsuya *et al.*, 2011), supporting the idea that language-level phonological factors influence adaptation. 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.

Stress is a suprasegmental property of multi-syllabic words or phrases. Speakers are sensitive to suprasegmental features in their speech, including amplitude (Bauer *et al.*, 2006; Patel *et al.*, 2015) and pitch (Burnett *et al.*, 1998; Patel *et al.*, 2011). The effect of syllable stress on formant adaptation has not been directly tested, but stress may have an effect on compensation to pitch perturbations. In an altered feedback experiment employing both upward and downward shifts of f_0 (Natke and Kalveram, 2001), native German speakers repeated the nonsense string [tata-tas] 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 speech feedback control is therefore not entirely clear.

While most altered auditory feedback studies consider adaptation or compensation effects in single, isolated monosyllabic words, recent studies consider speech motor learning at the phrase or sentence level. Lametti *et al.* (2018) found that speakers adapted to a greater degree when trained on a randomly-ordered list of words than when trained on sentences. However, expanding into larger speech units introduces factors not relevant to monosyllables, namely, word level, phrasal, or sentential stress. If stress modulates

the degree of adaptation, then sentences may appear to show less adaptation merely because they contain a combination of both stressed and unstressed syllables.

In the current study, we investigate two problems surrounding schwa and stress: (1) ambiguity with regard to the acoustic target space for unstressed schwa and (2) the effect of syllable stress on how speakers assess the attainment of their speech targets. We disentangle the effects of syllable position and stress by measuring responses to altered auditory feedback during the production of counterbalanced disyllabic words. When speakers adapt to altered auditory feedback, they change the way they speak so that they hear something closer to what they expect: 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 Experiment 1, 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. In Experiment 2, we use the same paradigm to test whether a decrease in $F1$ feedback drives an increase in produced $F1$. The direction of the $F1$ shift would shift the vowels in the experiment across different category boundaries; we tested two shift directions to help isolate the effect of stress from other phonetic effects.

II. METHODS

Procedures for Experiments 1 and 2 were identical except for the direction of the formant alteration (Fig. 1, top panels); combined methods are presented here. All procedures were approved by the Institutional Review Board at the University of Wisconsin–Madison.

A. Participants

All participants reported no history of hearing loss or neurological disorder. Participants provided informed consent and were paid \$15 or received course extra credit. Seventeen (fifteen female; aged 19–30, mean = 22) native speakers of American English, primarily students at the University of Wisconsin–Madison, participated in Experiment 1, and no data were excluded from this group. Twenty-three (20 female, aged 18–22, mean = 19) participants, none of whom participated in Experiment 1, were recruited for Experiment 2. One was excluded from further analysis for excessive yawning during speech production (> 5% of tokens).

B. Stimuli

Stimuli were chosen to address both foot type (location of primary stress) and vowel quality: the pair “beta” [ˈbeɪrə] and “abate” [əˈbeɪt], the pair “meta” [ˈmɛrə] and “adept” [əˈdɛpt], and to disentangle the effect of vowel quality apart from stress, “above” [əˈbʌv]. We use the phonetic symbol [ə] in both syllables of “above” to indicate that these vowels are acoustically similar (<50 Hz/37 mel difference

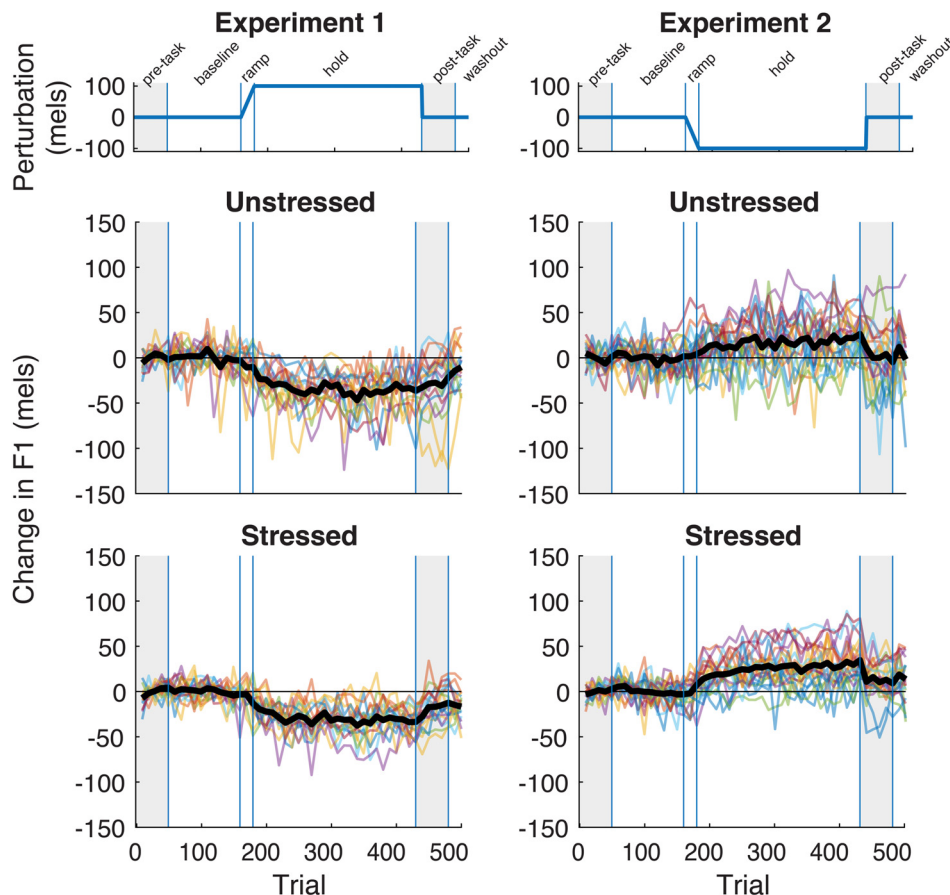


FIG. 1. (Color online) Top: schematic of perturbation magnitudes in Experiment 1 (left) and Experiment 2 (right). Adaptation shown for all subjects in unstressed (above) and stressed (below) syllables. Average across subjects overlaid in black.

in $F1$ and <25 Hz/13 mel in $F2$ in our data; unstressed [ə] is higher and more front than stressed [ə]). The local dialect of English has no words of the form '(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 containing four repetitions of each word.

C. 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. Every 20 trials, participants received an optional self-timed break. Typically, breaks lasted less than 30 s. Participants were not given explicit instructions about utterance length or volume. The experiment lasted 40 min total.

We used Audapter (Cai *et al.*, 2008; Tourville *et al.*, 2013) to alter the auditory feedback participants received, manipulating their vowel formant frequencies and playing them back in near-real-time (delay of ~ 18 ms measured on our system). Speech was recorded at a sampling rate of 48 000 Hz (downsampled to 16 000 Hz). We used the preset defaults in Audapter, including a preemphasis factor of 0.98 and an linear predictive coding order based on participant-reported gender (order = 15 for female, 17 for male).¹

The alteration occurred in six phases. In the *pre-task* phase (50 trials), participants heard loud noise (77 dB) that masked their auditory feedback. In the *baseline* phase (110 trials), participants heard their own unaltered (beyond any small perturbations introduced during resynthesis) feedback over headphones. 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 auditory feedback. In the *baseline*, *ramp*, *hold*, and *washout* phases, moderate speech-shaped noise (55 dB) was mixed with feedback so that participants could hear their feedback, but their unaltered feedback, both through air and bone conduction, was partially masked. During trials that shifted $F1$, the shift was maintained for the entire trial.

D. Analysis

1. Acoustic 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 3 ms using

Praat (Boersma and Weenink, 2017) via the wave_viewer analysis package (Niziolek, 2015), as these hand-corrected formant tracks were more reliable than the online estimates provided by Audapter. 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. We examined the final 75 trials in the hold phase in order to calculate a stable estimate for each word (~ 15 tokens per word). A separate baseline was computed for each word and syllable (e.g., separate baselines for the first and second syllables of “abate”) so that $F1$ change in each stress and vowel context could be individually assessed. 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.

2. Statistical modeling and effects of stress and word

For each experiment, we ran linear mixed-effects models using the lme4 package (Bates et al., 2015) in R (R Core Team, 2019) to estimate the coefficients of the factors under investigation. We performed analysis of variance (ANOVA) over those models to estimate the significance of those effects in the model. We additionally report unitless effect sizes (Cohen’s d) for categorical predictor variables. Models predicted change in $F1$ relative to baseline values. For all analyses in this study, we planned to add random slopes by speaker for each predictor variable in the model, as well as random intercepts for each speaker. Adding random slopes by speaker treats the speaker as the unit of analysis rather than the trial, decreasing the chance of a type I error by not treating each trial as independent, but instead using multiple trials to calculate a more reliable estimate of behavior. The random slopes allow the model to assess the overall effect of an independent variable, while also accounting for the variability between speakers.

In the majority of the models here, including random slopes for each independent variable resulted in models that were too complex to be reliable. Removing random slopes, however, often results in an overstating of the significance of a variable. In order to estimate the size and reliability for each independent variable, we built multiple models, all with random intercepts by subject, but with complementary random slopes by subject, so that each factor could be tested with a corresponding random slope. We estimated the size and significance of effects of the factor(s) that had a corresponding random slope included in that model.

For example, the first models investigated the effects of stress and word on change in $F1$. The planned model was $F1_{change} \sim stress + word + (stress|subject) + (word|subject) + (1|subject)$, which includes the independent variables of stress and word, as well as random slopes by subject for each of these variables, and random intercepts by subject to account for overall differences in adaptation for each

subject. As this model was not stable, we split the random slopes into two models: (1) $F1_{change} \sim stress + word + (stress|subject) + (1|subject)$ and (2) $F1_{change} \sim stress + word + (word|subject) + (1|subject)$. Results of (1), which included a random slope for stress by subject, were used to estimate the effect of stress, and results of (2), which included a random slope for word by subject, were used to estimate the effect of word. The size of the effects tended to be consistent across the models with complementary random slopes (differences typically < 1); it was the degrees of freedom and reliability (p -values) of those effects that differed. We report estimates for stress from model 1 and for word from model 2 in order to more faithfully represent the change in behavior of individual speakers rather than individual trials. Similarly, we estimate effect sizes (Cohen’s d) for stress from model 1 and for word from model 2.

We split models into as few as were necessary to achieve a reliable model. Some models had more than two independent variables, but for some of these cases it was only necessary to split into two models to achieve convergence (for example, one model with two random slopes and the other with the remaining random slope). If it was impossible for a random slope to be added to any configuration of a model, it is noted in the text, and the degrees of freedom are much larger than one less than the number of subjects.

3. Effects of syllable position

In addition to the effects of stress and word, we also investigated the effects of syllable position: when producing the second syllable of a two-syllable word, speakers have already been exposed to their altered auditory feedback for an entire syllable. This would allow extra time to plan and execute an online feedback correction in addition to the adaptive response, potentially causing a larger $F1$ change in final syllables. For this analysis, we excluded “above” in order to control for vowel effects; there is no counterpart to “above” with the same stressed vowel quality as in the pairs “meta” / “adept” and “beta” / “abate.” For each experiment, we report results from two models which predicted change in $F1$ with independent variables of stress, foot type (initial vs final stress), and vowel quality. One model contained random slopes for stress and foot, and the other contained a random slope for vowel. *Post hoc* Tukey tests using the lsmeans package (Lenth, 2018) investigated differences in initial and final syllables in the different metrical foot types.

4. Adaptation in masked auditory feedback

To determine the effects of longer-term learning, we ran an analysis investigating the change in $F1$ during the noise-masked post-task phase. Because there were substantial differences in $F1$ between masked (pre/post-task) and non-masked conditions, the pre-task phase served as a baseline for these trials, which were also produced in noise. Productions were re-baselined by word and syllable by subtracting the pre-task baseline, just as in the calculation for adaptation during the hold phase. These models included the

independent variables stress and word, just as in the first model described in this section. Similarly, we were not able to include random slopes for both stress and word in the same model for Experiment 1, so we ran two separate models, one for each random slope. However, for Experiment 2, even models with complementary random slopes were not stable. In order to reduce the random effects terms in these models, the data were split by metrical foot type, so that these models were run over initial-stress and final-stress words separately. For the Experiment 2 post-task phase, there were four models: for each stress pattern, there were two models, one estimating the effect of stress, and the other estimating the effect of word. We ran an additional analysis with a subset of the data that again excluded productions of “above,” investigating effects of learning separately by syllable and foot type. Comparing these results with the same models in the non-masked hold phase allowed us to establish the contribution of the effect of online compensation over the course of a single word. The structure of these models was the same as for the syllable position analysis during the hold phase, considering the independent variables stress, foot type, and vowel. Again, multiple models were required. In Experiment 1, one model included random slopes for foot type and vowel, and the other for stress. In Experiment 2, one model included random slopes for stress and vowel, and the other for foot type.

5. Effects of syllable duration

Syllable position and duration have the potential to explain similar variance in adaptation as a function of within-trial exposure to altered auditory feedback. Stressed vowels tend to be longer than unstressed vowels, and different vowel qualities have systematic differences in length (Lehiste, 1970). Longer syllables are likely to show greater $F1$ change relative to shorter syllables because the longer a syllable is on a given trial, the more time the speech motor system has to detect an error in the auditory feedback and correct for that error online. A positive correlation between $F1$ change and vowel duration could in this way be indicative of effects of online compensation beyond any learned adaptation. To determine whether duration and syllable position independently accounted for $F1$ change, we also planned to run models (again excluding “above”) for each experiment with independent variables of vowel quality in the stressed syllable, stress, metrical foot type (initial vs final stress), and a continuous predictor variable of duration, with random slopes for each of these four factors. However, no models including the effect of vowel were stable. Given that the purpose of these models was to assess the effects of foot type versus duration, and given that the effects of vowel were negligible in previous models, we omitted the vowel factor from these models. In both experiments, two separate models were required, one including a random slope for duration, and the other with random slopes for foot and stress.

6. Relationship between stressed and unstressed syllables in the same word

Finally, given possible coarticulatory or carryover effects that could influence adaptation across syllables within a word, we investigated the relationship between adaptation in the stressed syllable and adaptation in the unstressed syllable within the same utterance. We established the possibility of a relationship by running Pearson’s correlation tests for each subject over the adaptation in each syllable. We calculated Fisher’s z -transformed correlation coefficients, computed the mean, and then inverse z -transformed the mean back to a mean correlation coefficient. We additionally ran models predicting the change in the unstressed syllable with the change in the heterosyllable as a continuous predictor variable, and with metrical foot type and vowel as categorical predictors. In order to attain stable estimates, we split the data by foot type and calculated the size of the effect of adaptation in the heterosyllable in each type. To test the directionality of the relationship between heterosyllables by stress, we ran models with the same factors, but with unstressed syllables predicting change in the stressed syllable. There were in all eight models. For most models, random slopes for adaptation in the heterosyllable by subject did not result in a stable model, so only the random slope for vowel was included. The analysis focuses on differences in the size of the effect as a means to understand differences in metrical foot type rather than on significance estimates.

III. RESULTS

The results reported in this section reflect models with corresponding random factors. Full tables of results from ANOVAs are given in the Appendix. All significance levels here were tested at an α level of 0.05.

A. Adaptation in stressed and unstressed syllables

Normalized $F1$ values are displayed by stress type in Fig. 1. Over the course of Experiment 1, participants decreased their $F1$ by an average of 33 mels in opposition to the $F1$ increase in their auditory feedback. Similarly, participants increased their $F1$ by an average of 31 mels in opposition to the $F1$ decrease in Experiment 2. In both experiments, these adaptive shifts in vowel production were found for both stressed and unstressed vowels. Subject means by word and stress for each experiment in the hold phase are shown in Fig. 2. Participants decreased their $F1$ by an average of 23 mels in the post-task phase and 13 mels in the washout phase in Experiment 1. Participants increased their $F1$ by an average of 11 mels in the post-task phase and 17 mels in the washout phase in Experiment 2.

In an ANOVA predicting $F1$ change in Experiment 1 hold phase, only the effect of word was significant [$F(4, 15.9) = 8.7, p < 0.001$]. Two words with final stress, “abate” and “adept,” showed less adaptation than the reference word with final stress, “meta” (adaptation in “abate” was 19 mels less ($d.f. = 16, p < 0.01, d = 1.19$), and adaptation in “adept” was 15 mels less ($d.f. = 16, p < 0.01$,

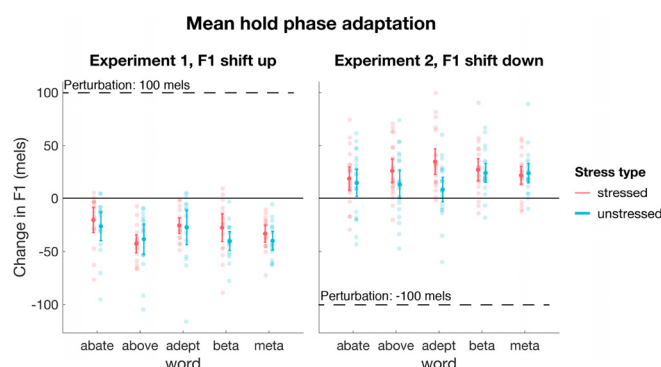


FIG. 2. (Color online) Distribution of subjects' mean change in $F1$ by word and stress for Experiment 1 (left) and Experiment 2 (right). Error bars show 95% confidence intervals.

$d = 0.92$) than that in “meta”). No other words were significantly different from “meta.”

For Experiment 2, the ANOVA showed that word was not a significant predictor [$F(4, 20.9) = 1.36, p = 0.28$] but there was an effect of stress [$F(1, 20.8) = 9.0, p = 0.007$]. Adaptation in *stressed* syllables was on average 9.3 mels greater ($d.f. = 20.8, p = 0.007, d = 0.73$) than in the accompanying unstressed syllable (Fig. 2, right panel).

B. Effects of syllable position in adaptation

In Experiment 1, initial-stress words showed greater adaptation than final-stress words. In order to understand the extent to which the presence of initial vs final stress may have contributed to differences in adaptation between words, as well as the extent to which each syllable accounted for these differences, we re-analyzed the hold phase productions, adding the effect of metrical foot type. In these analyses, productions of “above” were excluded to balance the data for syllable position and vowel quality. The following models included stress, metrical foot type, and vowel as independent variables and subject as a random factor.

An ANOVA found that only foot type [$F(1, 16.1) = 16.0, p = 0.001$] predicted change in $F1$ in Experiment 1. Words with final stress (“abate,” “adept”) adapted 15 mels less than words with initial stress (“beta,” “meta”) ($d.f. = 16.1, p = 0.001, d = 1.16$).

Post hoc Tukey tests on the model including random slopes for stress and foot showed 15 mels greater change in $F1$ in stressed vowels that occurred in initial position compared to final position ($d.f. = 16, p = 0.005$). Unstressed syllables in initial position showed changes in $F1$ that were 15 mels less than when they occurred in final position ($d.f. = 16, p = 0.005$). Unstressed syllables that occurred in final position showed 20 mels greater adaptation than stressed syllables that occurred in the same position ($d.f. = 31, p = 0.001$). In summary, unstressed final syllables adapted more than both unstressed *initial* syllables as well as *stressed* final syllables, but not different from stressed initial syllables.

In Experiment 2, the same model found that only stress was a significant factor [$F(1, 20.9) = 9.9, p = 0.005$]. In this model, stressed syllables adapted about 9.2 mels more than unstressed syllables did ($d.f. = 20.9, p = 0.005, d = 0.72$). In Experiment 2, unlike in Experiment 1, there was an effect of stress, but this effect was indifferent to whether the stressed syllable was initial or final. *Post hoc* Tukey tests with the same model specifications showed that in initial-stress words, the initial syllable adapted 9.2 mels more than the final syllable; in final-stress words, the final syllable adapted more than initial syllables by the same amount (both tests $d.f. = 21, p = 0.02$). Initial syllables in initial-stress words adapted 14 mels than initial syllables in final-stress words ($d.f. = 40.4, p = 0.02$).

C. Effects of syllable duration

By the same rationale that later syllables might exhibit greater shift-opposing behavior than initial syllables, we hypothesized that *longer* syllables would have a similar effect on change in $F1$. Further, duration and foot type might have accounted for similar variance in the data. In order to test the respective effects of duration and syllable order, ANOVAs including duration, stress, and metrical foot type as independent variables and subject as a random factor predicted change in $F1$.

In Experiment 1, an ANOVA found that duration was not a significant predictor in the model [$F(1, 16.2) = 0.001, p > 0.9$]. The only significant main-level effect was metrical foot type [$F(1, 17) = 15.51, p = 0.001$], where initial-stress words adapted 15 mels more than final-stress words ($d.f. = 16.9, p = 0.001, d = 1.16$). In this model, *post hoc* Tukey tests showed the same effects by stress and syllable as when duration was not included in the model: stressed syllables in initial-stress words showed 15 mels greater adaptation than in final-stress words ($d.f. = 16.8, p = 0.005$), unstressed final syllables showed 20 mels greater adaptation than stressed final syllables ($d.f. = 31, p = 0.002$), and unstressed final syllables showed 15 mels greater adaptation than initial unstressed syllables ($d.f. = 16.8, p = 0.005$). In the same model for Experiment 2, there was again neither effect of duration [$F(1, 20.8) = 2.93, p = 0.1$] nor of any other effect in the model. This may be because stress and duration are correlated, and the effect of stress may have been shared across the two factors. We concluded that duration did not add explanatory power and excluded it from further models.

D. Adaptation in masked auditory feedback

Comparing changes in production between the pre- and post-task phases, when no auditory feedback was available, allowed for assessing pure effects of adaptation in the absence of any online compensation. In both experiments, there was evidence of adaptation in the masked post-task phase, indicative of sensorimotor learning (Fig. 3).

In Experiment 1, the mean post-task change in $F1$ across all syllables was 23 mels in the direction opposing

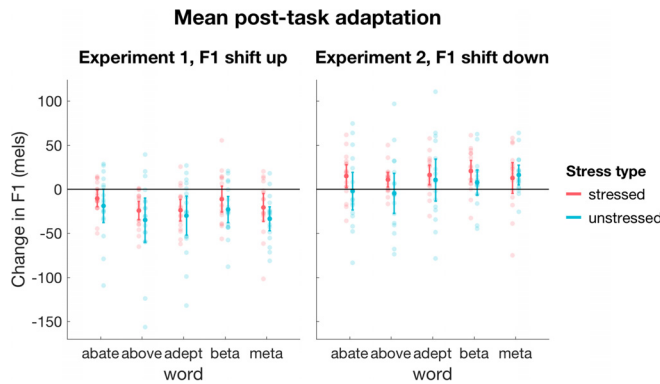


FIG. 3. (Color online) Difference in mean adaptation between unstressed and stressed syllables during the post-task phase relative to the pre-task for Experiment 1 (left) and Experiment 2 (right). Adaptation during masking noise suggests effects of learning.

perturbation, 70% of the change seen in the hold phase. In the ANOVA associated with the linear model including stress and word, there was an effect of stress that was not evident in the hold phase, with 10 mels greater adaptation ($d=0.82$) in unstressed syllables than stressed [$F(1,16) = 4.7, p = 0.046$]. The effect of word was maintained [$F(4,16) = 3.6, p = 0.03$], with productions of “abate” showing showing 12.5 mels ($d=0.98$) less $F1$ change than the reference “meta.”

In Experiment 2, the mean post-task change in $F1$ across all syllables was 7 mels, 22% of the change seen in the hold phase. No model with a random slope for word converged, so the data were split into initial-stress and final-stress words to reduce the number of random slope terms in the model predicting the effect of word. In these models, the ANOVA showed that neither stress nor word effects were significant.

In the Experiment 1 hold phase, there had been an effect of metrical foot type, as well as three significant differences in marginal means, where unstressed final syllables and stressed initial syllables showed the greatest adaptation. We tested whether any of these effects could have been due to online compensation by considering these differences in the noise-masked post-task phase, again excluding productions of “above.” Differences between initial and final position for stressed and unstressed syllables by phase of the experiment are shown in Fig. 4. The effect of foot did not persist, but as in the post-task analysis that included the factor of word, there was an effect of stress [$F(1,15.9) = 6.2, p = 0.02$], with unstressed syllables showing adaptive decreases in $F1$ that were 10 mels greater than in stressed syllables ($d.f. = 16, p = 0.02, d = 0.43$). There was also an effect of vowel [$F(1,15.9) = 11.8, p < 0.01$], with words containing [ei] in the stressed syllable showing 11.6 mels ($d=0.84$) less adaptive change in $F1$ than words containing [ε]. *Post hoc* Tukey tests showed that none of the three effects by stress and foot type survived into the post-task phase (all $p > 0.05$). In the Experiment 2 post-task phase, there was no significant effect of foot, stress, or vowel.

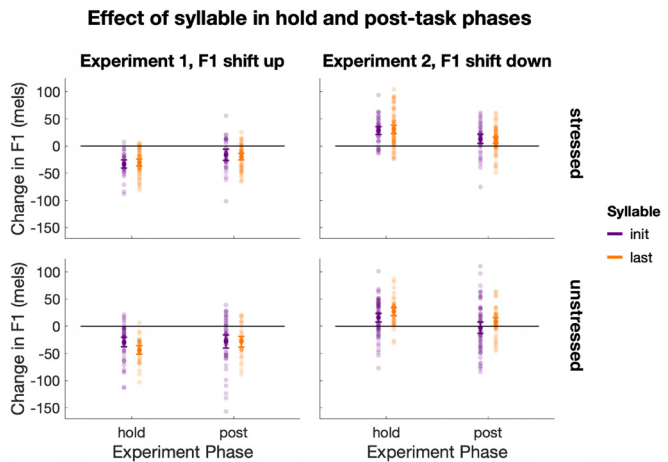


FIG. 4. (Color online) Difference in mean adaptation between initial and final syllables by stress type during the post-task phase relative to the hold phase for Experiment 1 (left) and Experiment 2 (right). Figure and analysis exclude productions of “above.”

E. Relationship between stressed and unstressed syllables in the same word

One possible theory explaining differences in adaptation between stressed and unstressed syllables considers whether adaptation in the unstressed syllable is planned independently or whether it is contingent on adaptation in the stressed syllable. We tested this idea by investigating whether there was a relationship between adaptation in both syllables within a single trial (Fig. 5). In baseline phase productions, the mean correlation (back-transformed from the mean Fisher z -transformed coefficients) across subjects between baseline-normalized $F1$ in the two syllables was 0.30 across all words in Experiment 1 and 0.17 in Experiment 2. The mean correlation across subjects between adaptation in the two syllables during the hold phase was 0.16 across all words in Experiment 1 and 0.11 in Experiment 2.

To further test which factors affected the strength of this relationship, we additionally considered the role of foot type, i.e., whether the stressed syllable was first or last. Productions of the word “above” were excluded from this analysis to balance the stimulus set for foot type. Linear models that included factors of change in $F1$ in the heterosyllable, foot type, and vowel did not converge, so models were split by foot type. There were eight models in all, with models predicting each syllable by the other, and with splits by metrical foot type. Data for all eight models are shown in the Appendix.

For three of the eight models, random slopes by subject for the adaptation in the heterosyllable failed to converge, increasing the possibility of type I error. Therefore, we will focus on the size of the coefficient rather than the significance of the effect. In no model was there a significant effect of vowel.

In predicting change in the unstressed syllable by the change in the stressed syllable in Experiment 1, there was a greater coefficient in initial-stress words (0.34, $d.f. = 14.8, p < 0.0001$) than in final-stress words (0.08, $d.f. = 18.8, p = 0.44$). That is, the stressed syllable was more predictive

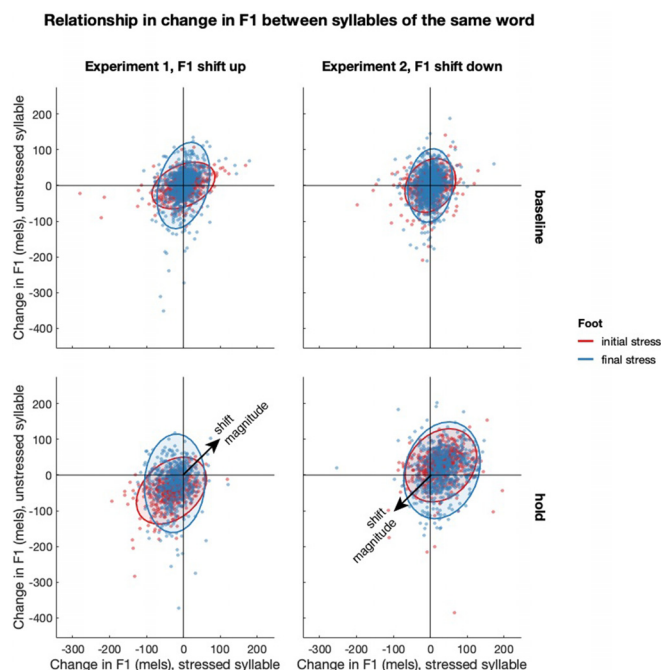


FIG. 5. (Color online) Relationships between change in $F1$ in each syllable of the same word in the baseline phase (top panels) and hold phase (bottom panels), 95% confidence ellipses shown for each foot type. Data shown exclude productions of “above.”

when it preceded the unstressed syllable. In Experiment 2, the coefficient associated with change in the unstressed syllable was only slightly greater in initial-stress words (0.17 , $d.f. = 15.5$, $p = 0.02$) than in final-stress words (0.15 , $d.f. = 11.52$, $p = 0.04$). However, as shown in the top panels in Fig. 5, even the baseline phase in Experiments 1 (at left) and 2 (at right) differed in the extent to which $F1$ values in one syllable were predicted by values in the heterosyllable.

In predicting change in the stressed syllable by the change in the unstressed syllable, Experiments 1 and 2 showed similar patterns. The effect was greater in initial-stress words, with a coefficient of 0.34 ($d.f. = 499.6$, $p < 0.0001$) in Experiment 1 and 0.22 ($d.f. = 21.9$, $p < 0.001$) in Experiment 2, than in final-stress words, which had a coefficient of only 0.06 ($d.f. = 479$, $p < 0.01$) in Experiment 1 and 0.07 ($d.f. = 577.6$, $p = 0.006$) in Experiment 2.

F. Reliability of online feedback shifts

We investigated the results of Audapter’s online tracking algorithm to verify that the shift had worked as intended for both stressed and unstressed syllables. Audapter recorded formant values equal to zero for 5.5% of unstressed syllables in Experiment 1 and 7.0% of unstressed syllables in the hold phase of Experiment 2. For both experiments, zeros were recorded for 1% or fewer stressed syllables during the hold phase. With these zeros excluded, the correlation in $F1$ values (mels) between Audapter’s online estimates and our hand-corrected analysis with Praat was 0.71 in Experiment 1 and 0.83 in Experiment 2.

However, these zeros were not evenly distributed across participants. While most participants experienced fewer than 20

unshifted trials during the combined ramp and hold phases (out of 290 trials that should have been shifted), there were four participants from each experiment who experienced 50–100 trials that were not shifted. For most of these participants, adaptation was still in the range of those who experienced the shift in all trials. Further, the unshifted trials were largely very short (less than 40 ms) and may therefore not have contributed much to the response even if they had been tracked. For all of these reasons, we did not exclude these participants.

IV. DISCUSSION

Speakers showed robust adaptation to $F1$ shifts in auditory feedback, changing $F1$ in both stressed and unstressed syllables in opposition to the feedback shift. Adaptation in syllables of both stress types was elicited by both upward (Experiment 1) and downward (Experiment 2) shifts of $F1$, but stress interacted with adaptation magnitude in complex ways: the size and direction of the effect of stress was not preserved across experiments and was not consistent across words.

A. Adaptation suggests a target for schwa

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 on par with stressed vowels when $F1$ was shifted up (Experiment 1), and they showed more adaptation than stressed vowels in the post-task phase testing effects of learning. Conversely, stressed syllables adapted more than unstressed syllables when $F1$ was shifted down (Experiment 2), but there was still significant adaptation in unstressed syllables; the difference between stressed and unstressed syllables did not persist in the post-task phase.

The adaptation observed in unstressed syllables suggests that schwa does indeed have a target. However, a possible alternative explanation is that the size of adaptation in a given word may be planned at the word level rather than individually for each syllable or vowel. Under this theory, perhaps only the stressed syllable would determine both the amount of adaptation for a word and the target for the unstressed vowel, as adaptation may be dependent on the vowel quality in the stressed syllable (Lametti *et al.*, 2018; Mitsuya *et al.*, 2015). This theory is supported by the fact that adaptation in the unstressed syllable was fairly well-predicted by that in the stressed syllable, but for Experiment 1, this relationship was limited to initial-stress words. A theory positing adaptation planning at the word level would not predict an effect of stress placement.

Another possibility is that, rather than acquiring a specific target and adaptation magnitude at the word level, schwa is attracted to nearby stressed vowels through coarticulation, as schwa is subject to coarticulatory pressures from either a preceding or following stressed vowel (Fowler, 1981). Under this second theory, the adaptation that occurs in schwa is not due to a mismatch between auditory feedback and an intrinsic target, but rather that schwa

shifts in the direction of the preceding or upcoming adapted vowel. That is, adaptation is planned at the level of the stressed heterosyllable, but the adaptation itself is not applied to schwa; schwa shifts in the direction of the adapted heterosyllable *via* coarticulatory processes unrelated to error-correction. Given potential differences in perseveratory and anticipatory coarticulation, this explanation might account for differences in adaptation by stress placement. However, such an explanation might have predicted a stronger effect of word or vowel quality and also would not predict the finding in Experiment 1, where there was greater adaptation in unstressed schwa relative to stressed vowels in the post-task phase.

B. Effects of phonetic categories on adaptation

There were overall differences in adaptation depending on the word, but these between-word differences did not persist across experiments. The effect of word was significant in Experiment 1, where the final-stress words “adept” and “abate” had significantly less adaptation than in “meta”; this pattern did not extend to Experiment 2. Given that both of these words have different vowels in the stressed syllable but similar stress patterns, the significance seems to highlight an effect of foot type rather than of phonetic context. An effect of the vowel identity emerged, but only in the post-task phase: in Experiment 1, words with [ei] showed less adaptation than words with [ɛ]. There was no consistent effect of word or vowel quality across the two experiments. Importantly, there was no difference between adaptation in “above” and other words during the hold phase in either experiment, suggesting that any differences in unstressed [ə] during this phase were not due to vowel quality effects alone.

The adaptation in unstressed schwa differed between the two experiments: in Experiment 1, adaptation in unstressed syllables during the hold phase was on par with that in stressed syllables, and evidence from the post-task phase suggests that speakers may have learned a pattern where unstressed syllables adapted more than stressed syllables. In Experiment 2, adaptation in unstressed syllables was less than that of stressed syllables during the hold phase, but this difference was eliminated during the post-task phase. Still, the difference in direction of the effect across the two experiments was unexpected, and it is not predicted by any of the theories offered so far.

Given that speakers respond more strongly to altered auditory feedback signals that result in the perception of a vowel category that differs from the target (Niziolek and Guenther, 2013), the organization of the vowel space in the $F1$ dimension may explain some of these differences in schwa adaptation across experiment conditions, since the $F1$ perturbation shifted vowels towards different phonetic categories depending on the direction of the shift. Raising $F1$ in [ə] in Experiment 1 caused speakers to hear a vowel closer to [a], which is a contrastive vowel in English. In Experiment 1, therefore, in both the stressed and unstressed schwa, the $F1$ perturbation shifted participants’ productions into a

contrasting vowel category. Lowering $F1$ in Experiment 2 caused speakers to hear a vowel closer to [ɪ]. Producing [ɪ] for [ə] is not acceptable in a stressed context in English (*[əbɪv]). However, in an unstressed context, speakers may have judged the $F1$ perturbation to be more acceptable, as unstressed [ə] may be produced [ɪ] in certain contexts, such as in the -ed suffix (e.g., *wretched*, *blurred*) or plural morpheme following sibilants (*roses*, see Flemming and Johnson, 2007). This may be evidence that the upper $F1$ category boundary may be comparable between stressed and unstressed [ə], but the lower $F1$ boundary may extend lower (i.e., encompass a greater range of acceptability) for the unstressed vowel.

An overall category size that is greater for unstressed [ə] than stressed [ə] is also supported by accounts that this unstressed vowel is more variable in production than other vowels (Magen, 1984). This greater range of acceptable $F1$ values for unstressed [ə] would predict that there could be a greater range of *produced* $F1$ values in the unstressed variant as well. In productions of “above” from the baseline phase (unaltered auditory feedback) in both experiments, $F1$ values produced during the stressed vowel were a subset of those produced during the unstressed vowel. Indeed, the unstressed [ə] vowel contained lower values of $F1$ than those produced in the stressed vowel (Fig. 6).

An additional factor that may explain differences in adaptation between schwa and other vowels is the different somatosensory feedback experienced during the production of these vowels. When speakers change their articulations in response to the altered auditory feedback, they may perceive their altered vowels to be a better *auditory* match to the target. However, by producing a vocal tract configuration that is different from the usual articulation for a given word, a mismatch in somatosensory feedback, or tactile and tongue position information, is introduced. Mismatches between expected and actual somatosensory feedback may explain why speakers in altered auditory feedback experiments tend not to oppose the entire magnitude of the shift (Katseff *et al.*, 2012). Schwa is produced with very little constriction in the oral cavity, though there is likely constriction at the pharynx (Gick, 2002). Relatively little somatosensory information in

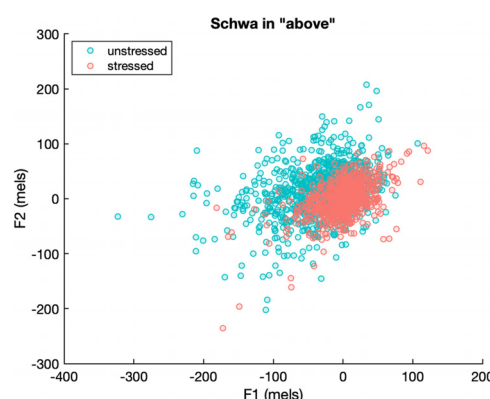


FIG. 6. (Color online) $F1$ and $F2$ values for [ə] in “above” during unaltered baseline phase only, normalized by the second syllable. There is a greater range of low $F1$ values for the unstressed vowel.

schwa might predict that speakers may not be as sensitive to changes in its articulation in comparison with close vowels, which have greater lingual contact with the upper jaw; compensation for perturbations to close vowels may trigger a large somatosensory error, inhibiting compensation to a greater degree than in other vowels (Mitsuya *et al.*, 2015). This finding would predict that unstressed schwa, given its lack of lingual contact, would be free to adapt more than stressed syllables in both experiments. However, because speakers expect less somatosensory feedback from schwa, they may be sensitive to an *introduction* of unexpected somatosensory information caused by the change in vowel articulation. This would predict greater adaptation in stressed syllables than unstressed syllables across the board. We did not find such consistency in the effect of stress in either direction. However, the shifts in each experiment might have had different effects on sensitivity to somatosensory feedback mismatches. When *F1* was shifted up, speakers counteracted the shift in *F1* by producing a vowel with a tongue position that was higher in the oral cavity (more close vowel), but when *F1* was shifted down, speakers counteracted the shift in *F1* by producing a vowel with a lower tongue position (more open vowel). If speakers are more sensitive to the feedback introduced by producing a more close vowel, then we would expect *less* adaptation in unstressed schwa relative to other vowels in Experiment 1 but *more* adaptation in unstressed schwa relative to other vowels in Experiment 2. However, this hypothesis was not borne out: the directionality of the effect of stress was opposite of what this hypothesis would predict. It seems that the directionality of the effect of schwa can not be tied to differences in somatosensory feedback.

C. Effects of timing: Duration and syllable position

We had hypothesized that the longer a speaker hears their altered auditory feedback during continuous speech production, the more time the speaker would have to calculate and execute a motor plan to oppose the alteration. A longer period of time can be measured either as the duration of a given syllable or whether that syllable was first or last in a given trial. We therefore hypothesized that both *longer* syllables and *final* (i.e., later) syllables would show greater *F1* change than shorter or initial syllables.

Indeed, in Experiment 1, unstressed final syllables showed greater adaptation than unstressed initial syllables. Furthermore, the elimination of this difference in the noise-masked post-task phase suggests that the increased adaptation in final syllables may be due to online compensation (opposition to the shift over a single trial) rather than trial-to-trial (learned) adaptation. However, stressed syllables showed the opposite pattern: stressed final syllables showed *less* adaptation than stressed initial syllables, and this effect was eliminated in masking noise. This pattern is not as readily explained as the effect of online compensation. However, it may be partially reconciled by considering the differences in the durations of initial stressed and initial unstressed syllables: as initial unstressed syllables tend to be

shorter than stressed syllables, they provide comparatively less exposure to the altered feedback before the onset of the following syllable. For example, the first vowel in “abate” was approximately 55 ms on average, which provided only one-third of the exposure time afforded by the stressed vowel in “beta,” which lasted approximately 162 ms on average. Therefore, final unstressed syllables may potentially benefit more from the longer exposure time provided by an initial stressed syllable, resulting in greater apparent adaptation. In contrast to Experiment 1, there were no observed effects of foot type or vowel duration in Experiment 2.

The seemingly contradictory results between the experiments and effects of stress type, metrical foot type, and online compensation can be rationalized as an interaction of these three factors that differ in their size and direction, depending on the experiment. Initial-stress words showed greater overall adaptation than final-stress words in Experiment 1. One possibility is that the initial syllable is what triggers adaptation magnitude, and stressed syllables *trigger* greater overall adaptation than unstressed syllables. Unstressed syllables did still adapt in an initial position, suggesting that adaptation is not triggered exclusively by an initial stressed syllable. Foot type may have a larger effect than online compensation, which would be consistent with the observation that stressed syllables in the initial position adapted more than stressed syllables in the final position, despite potential effects of online compensation giving an advantage to final syllables. Further, on a given trial, the adaptation in the unstressed syllable reliably predicted the adaptation in the stressed syllable, but this relationship was largely limited to initial-stress words.

The interaction of stress type, foot type, and online compensation may account for the subtler differences in how the effect of stress emerged in both experiments. In Experiment 1, syllables that are final, unstressed, or that occur in initial-stress words all receive an “advantage” in adaptation; that is, they show greater adaptation relative to syllables that are, respectively, initial, stressed, or in final-stress words. These three separate advantages co-occur in final unstressed syllables, such as the schwas in “meta” and “beta.” Clear effects attributable to each of these three factors did not emerge in Experiment 2. One possibility is that the effects of online compensation and foot type are similar between the two experiments, but these similarities are masked by the fact that the effect of stress occurs in the opposite direction, with stressed syllables having an “advantage” in Experiment 2. The trifecta of advantages in Experiment 1 would not be able to co-occur in Experiment 2: it is impossible for a stressed syllable to be word-final in an initial-stress word. For example, the first syllable of “beta” receives the advantages of being stressed and occurring in an initial-stress word, but not of online compensation because it is not the final syllable. Similarly, the final syllable in “abate” has the advantages of stress and online compensation, but not of occurring in an initial-stress word. These opposing effects may have occluded underlying

effects of online compensation and foot type in Experiment 2 that more clearly emerged in Experiment 1.

While stress, foot type, and online compensation may have had their own independent effects on adaptation, it is difficult to assess each factor independently in disyllables. Longer words with one or more stresses (e.g., ['ɛləfənt], [ædəp'theɪfən]) offer more combinations of stress types to investigate independent effects of stress order and overall stress as well as a longer period of time to observe effects of online compensation.

V. CONCLUSION

This study found significant adaptation to altered auditory feedback in both unstressed and stressed versions of [ə]. Overall, there was no difference between unstressed and stressed vowel adaptation in Experiment 1, when $F1$ was raised, although initial-stress words adapted more than final-stress words, and noise-masking revealed greater adaptation in unstressed vowels. The opposite pattern emerged in Experiment 2, where $F1$ was lowered: there was overall greater adaptation in stressed than unstressed vowels, and this difference disappeared under masking noise. The inconsistency in responses to perturbations in unstressed and stressed versions of [ə] between Experiments 1 and 2 suggests differences in the acoustic location and size of the unstressed and stressed [ə] vowel categories, and these differences are supported by the patterns of variability in $F1$ in the production of these vowel categories in unperturbed speech. Adaptation differences between syllables of the same word suggest intersecting effects of foot type, stress, and online compensation, but it is difficult to confidently assess the independent effects of each because they are all interdependent in disyllables. Experiments considering longer words with multiple stresses may help to determine the independent role of each of these factors. While the *direction* of the effect of stress is not consistent across experiments, the differences in adaptation observed here show that syllable stress, and, importantly, the order of stressed syllables in a word, does impact how speakers correct for errors in their auditory feedback, and this should be taken into consideration in the design of altered auditory feedback experiments considering word units larger than a single syllable.

ACKNOWLEDGMENTS

Many thanks to Daniel Bolt for help with statistical modeling. This work was supported by NIH Grant No. 5T32DC005359-13 to the University of Wisconsin–Madison, NIH F32 training Grant No. DC017653 to S.B., and NIH R00 Grant No. DC014520 to C.N.

APPENDIX

[random]: refers to the varying random effects structure in each model. Multiple models were required to estimate the size and significance of many of the effects in this

study. The values reported in Tables I–III correspond to the model version that included the random slope matching that factor. For example, estimates for stress come from models that included a random slope for stress, but random

TABLE I. ANOVA results for all models. Coefficients shown for model with random effects structure corresponding to that factor (see Sec. II). **Boldface** indicates significance ($p < 0.05$).

Factor	<i>df.</i>	<i>F</i>	<i>p</i>
$F1_{change} \sim stress + word + [random]$			
Expt 1 stress	1,16	0.96	0.34
Expt 2 stress	1,20.8	9.0	0.007
Expt 1 word	4,15.9	8.7	<0.001
Expt 2 word	4,20.9	1.36	0.28
$F1_{change} \sim stress + foot + vowel + [random]$			
Expt 1 stress	1,16	2.7	0.12
Expt 2 stress	1,20.9	9.9	0.005
Expt 1 foot	1,16.1	16.0	0.001
Expt 2 foot	1,21.0	1.83	0.19
Expt 1 vowel	1,15.9	1.05	0.32
Expt 2 vowel	1,20.96	0.002	0.97
$F1_{change} \sim stress + foot + duration + [random]$			
Expt 1 stress	1,18	2.61	0.12
Expt 2 stress _{init}	1,22.44	0.84	0.37
Expt 2 stress _{final}	1,70.38	1.3	0.26
Expt 1 foot	1,17	15.51	0.001
Expt 2 foot	1,21.4	0.12	0.73
Expt 1 duration	1,16.2	0.001	0.97
Expt 2 duration	1,20.8	2.93	0.10
$F1_{change,unstr} \sim F1_{change,str} + vowel + [random]$			
Expt 1 hetsyll _{init}	1,14.8	33.1	<0.0001
Expt 2 hetsyll _{init}	1,15.5	8.3	0.01
Expt 1 vowel _{init}	1,15.9	0.2	0.67
Expt 2 vowel _{init}	1,21.1	0.03	0.87
Expt 1 hetsyll _{final}	1,18.8	0.6	0.44
Expt 2 hetsyll _{final}	1,11.5	5.4	0.04
Expt 1 vowel _{final}	1,15.8	0.09	0.77
Expt 2 vowel _{final}	1,21.1	1.49	0.24
$F1_{change,str} \sim F1_{change,unstr} + vowel + [random]$			
Expt 1 hetsyll _{init}	1,499	63.0	0.0001
Expt 2 hetsyll _{init}	1,22	15.6	0.0007
Expt 1 vowel _{init}	1,15.8	1.3	0.27
Expt 2 vowel _{init}	1,20.94	1.26	0.27
Expt 1 hetsyll _{final}	1,479	6.7	0.001
Expt 2 hetsyll _{final}	1,577.6	7.5	0.006
Expt 1 vowel _{final}	1,16	0.3120	0.58
Expt 2 vowel _{final}	1,20.8	4.25	0.052
$F1_{change,post} \sim stress + word + [random]$			
Expt 1 stress	1,16	4.7	0.046
Expt 2 stress	1,21	3.8	0.07
Expt 1 word	4,16	3.6	0.03
Expt 2 word _{init}	1,22.7	0.007	0.94
Expt 2 word _{final}	2,20.9	2.5	0.11
$F1_{change,post} \sim stress + foot + vowel + [random]$			
Expt 1 stress	1,15.9	6.2	0.02
Expt 2 stress	1,21.1	3.1	0.09
Expt 1 foot	1,16.0	0.1	0.75
Expt 2 foot	1,20.9	1.6	0.23
Expt 1 vowel	1,15.9	11.8	0.003
Expt 2 vowel	1,21.2	0.1	0.75

TABLE II. Coefficients for significant terms in linear models. Terms whose estimates did not reach significance (i.e., $p > 0.05$) not shown.

Factor	Est.	error	d.f.	p	Coh. d
$F1_{change} \sim stress + word + [random]$					
Expt 1 abate	18.9	6.0	16	0.006	1.19
Expt 1 adept	14.5	4.3	164	0.004	0.92
Expt 2 unstrs	-9.3	3.1	20.8	0.007	0.73
$F1_{change} \sim stress + foot + vowel + [random]$					
Expt 1 fin.strs	14.9	3.7	16.1	0.001	1.16
Expt 2 unstrs	-9.2	2.9	20.9	0.005	0.72
$F1_{change} \sim stress + foot + duration + [random]$					
Expt 1 foot	14.9	3.8	17	0.001	1.16
$F1_{change.post} \sim stress + word + [random]$					
Expt 1 abate	12.5	4.3	16	0.01	0.98
Expt 1 unstrs	-10.4	4.8	16	0.046	0.82
$F1_{change.post} \sim stress + foot + vowel + [random]$					
Expt 1 unstrs	-10	4.1	15.9	0.02	0.43
Expt 1 [ei]	11.6	3.4	15.9	0.003	0.84

TABLE III. Coefficients in separate models for each stress type. Columns 2 and 3 show the intercept of the model and coefficient of change in heterosyllable, respectively. Effect of vowel was not significant in any model and is not shown. Degrees of freedom shown for effect of heterosyllable; all d.f. for intercepts were approximately 16. Large d.f. resulted when the random slopes by subject for change in $F1$ in the heterosyllable could not be included due to lack of convergence. Significance in these models should be treated with caution. Random intercepts for subject were included in all models. **Boldface** indicates significance ($p < 0.05$).

Stress type	Intercept	coeff _{hetsyll}	d.f. _{hetsyll}	p _{hetsyll}
Predicting change in $F1$ in unstressed syllable				
Expt 1 _{initial}	-40.6	0.34	14.8	<0.0001
Expt 2 _{initial}	26.8	0.17	15.5	0.01
Expt 1 _{final}	-25.6	0.08	18.8	0.44
Expt 2 _{final}	10.5	0.15	11.5	0.04
Predicting change in $F1$ in stressed syllable				
Expt 1 _{initial}	-36	0.34	499.6	<0.0001
Expt 2 _{initial}	24.4	0.22	21.9	<0.001
Expt 1 _{final}	-26	0.06	479	<0.01
Expt 2 _{final}	39.4	0.07	577.6	0.006

slopes for other factors may have been excluded from that model. All models include random intercepts for subject. For more details on each model, see Sec. II.

¹A full list of settings we used can be found in the Audapter code, https://github.com/caisq/audapter_matlab/blob/0b643a241a526868bec159bfb4205db4f9e1787/mcode/getAudapterDefaultParams.m.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). "Fitting linear mixed-effects models using lme4," *J. Stat. Softw.* **67**(1), 1–48.

Bauer, J. J., Mittal, J., Larson, C. R., and Hain, T. C. (2006). "Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude," *J. Acoustical Soc. America* **119**(4), 2363–2371.

Boersma, P., and Weenink, D. (2017). "Praat: Doing phonetics by computer [computer program]," v.6.0.33 (Last viewed 12/1/2017).

Browman, C., and Goldstein, L. (1994). "Targetless" schwa: An articulatory analysis," in *Papers in Laboratory Phonology II Gesture, Segment, Prosody*, edited by G. J. Docherty and D. R. Ladd (Cambridge University Press, Cambridge, UK), pp. 26–67.

Burnett, T. A., Freedland, M. B., and Larson, C. R. (1998). "Voice F0 responses to manipulations in pitch feedback," *J. Acoust. Soc. Am.* **103**, 3153–3161.

Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., and Perkell, J. S. (2008). "A system for online dynamic perturbation of formant frequencies and results from perturbation of the mandarin triphthong /iaiu/," in *Proceedings of the 8th International Seminar on Speech Production*, Strasbourg, France, pp. 65–68.

Cai, S., Ghosh, S. S., Guenther, F. H., and 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," *J. Neurosci.* **31**(45), 16483–16490.

Flemming, E., and Johnson, S. (2007). "Rosa's roses: Reduced vowels in American English," *J. Int. Phon. Assoc.* **37**(1), 83–96.

Fowler, C. A. (1981). "Production and perception of coarticulation among stressed and unstressed vowels," *J. Speech Hear. Res.* **24**, 127–139.

Gick, B. (2002). "An X-ray investigation of pharyngeal constriction in American English schwa," *Phonetica* **59**, 38–48.

Houde, J. F., and Jordan, M. I. (1998). "Sensorimotor adaptation in speech production," *Science* **279**, 1213–1216.

Katseff, S., Houde, J. F., and Johnson, K. (2012). "Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback?," *Lang. Speech* **55**(2), 295–308.

Kondo, Y. (1994). "Phonetic underspecification in schwa," in *Proceedings of the Third International Conference on Spoken Language Processing*, September 18–22, Yokohama, Japan, pp. 311–314.

Koopmans-van Beinum, F. J. (1994). "What's in a schwa?," *Phonetica* **51**, 68–79.

Lametti, D. R., Smith, H. J., Watkins, K. E., and Shiller, D. M. (2018). "Robust sensorimotor learning during variable sentence-level speech," *Curr. Biol.* **28**(19), 3106–3113.

Lehiste, I. (1970). *Suprasegmentals* (MIT Press, Cambridge, MA).

Lehiste, I. (1976). "Suprasegmental features of speech," in *Contemporary Issues in Experimental Phonetics* (Academic Press, New York), pp. 225–239.

Lenth, R. (2018). "lsmeans, 'Least-Squares Means,'" v.2.30-0 (Last viewed 6/1/2020), see <https://cran.r-project.org/web/packages/lsmeans/>.

Magen, H. (1984). "Vowel-to-vowel coarticulation in English and Japanese," *J. Acoust. Soc. Am.* **75**, S41.

Mitsuya, T., MacDonald, E. N., Munhall, K. G., and Purcell, D. W. (2015). "Formant compensation for auditory feedback with English vowels," *J. Acoust. Soc. Am.* **138**(1), 413–424.

Mitsuya, T., MacDonald, E. N., Purcell, D. W., and Munhall, K. G. (2011). "A cross-language study of compensation in response to real-time formant perturbation," *J. Acoust. Soc. Am.* **130**(5), 2978–2986.

Natke, U., and Kalveram, K. T. (2001). "Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables," *J. Speech Lang. Hear. Res.* **44**, 577–584.

Niziolek, C. (2015). "wave_viewer: First release," (Last viewed 6/1/2020).

Niziolek, C. A., and Guenther, F. H. (2013). "Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations," *J. Neurosci.* **33**(29), 12090–12098.

Patel, R., Niziolek, C., Reilly, K., and Guenther, F. H. (2011). "Prosodic adaptations to pitch perturbation in running speech," *J. Speech Lang. Hear. Res.* **54**, 1051–1059.

Patel, R., Reilly, K. J., Archibald, E., Cai, S., and Guenther, F. H. (2015). "Responses to intensity-shifted auditory feedback during running speech," *J. Speech Lang. Hear. Res.* **58**(6), 1687–1694.

R Core Team (2019). "R: A language and environment for statistical computing," Technical Report.

Szigetvári, P. (2018). "Stressed schwa in English," *Even Yearbook* **13**, 81–95.

Tourville, J. A., Cai, S., and Guenther, F. H. (2013). "Exploring auditory-motor interactions in normal and disordered speech," *Proc. Mtg. Acoust.* **9**, 060180.

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

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

Wells, J. C. (1982). *Accents of English 1: An Introduction*, Vol. 1 (Cambridge University Press, Cambridge, UK).