

RESEARCH ARTICLE

Control of Movement

Sensorimotor adaptation to a nonuniform formant perturbation generalizes to untrained vowels

 Benjamin Parrell,*  Caroline A. Niziolek,* and  Taijing Chen

University of Wisconsin-Madison, Madison, Wisconsin, United States

Abstract

When speakers learn to change the way they produce a speech sound, how much does that learning generalize to other speech sounds? Past studies of speech sensorimotor learning have typically tested the generalization of a single transformation learned in a single context. Here, we investigate the ability of the speech motor system to generalize learning when multiple opposing sensorimotor transformations are learned in separate regions of the vowel space. We find that speakers adapt to a nonuniform “centralization” perturbation, learning to produce vowels with greater acoustic contrast, and that this adaptation generalizes to untrained vowels, which pattern like neighboring trained vowels and show increased contrast of a similar magnitude.

NEW & NOTEWORTHY We show that sensorimotor adaptation of vowels at the edges of the articulatory working space generalizes to intermediate vowels through local transfer of learning from adjacent vowels. These results extend findings on the locality of sensorimotor learning from upper limb control to speech, a complex task with an opaque and nonlinear transformation between motor actions and sensory consequences. Our results also suggest that our paradigm has potential to drive behaviorally relevant changes that improve communication effectiveness.

altered auditory feedback; generalization; sensorimotor adaptation; speech production

INTRODUCTION

When speakers learn to change the way they produce a speech sound, how much does that learning generalize to other speech sounds? Generalization is a critical consideration in speech therapy, as treatment strategies that target the pronunciation of a given speech sound aim to maximize beneficial learning in untreated as well as treated targets. However, much of what we understand about the transfer of learning in speech comes from studies in which a single transformation is learned in a single context (for example, learned changes to the pronunciation of a single word), limiting the expected scope of generalization. Here, we investigate the ability of the speech motor system to generalize learning when multiple opposing sensorimotor transformations are learned in separate regions of the vowel space.

When exposed to errors in sensory feedback, humans update their motor commands in a process called sensorimotor adaptation. For example, an alteration to the auditory feedback of speech, e.g., shifting the formant frequencies of

“bed” to resemble those of “bad,” induces an adaptive change in the opposite direction, i.e., toward “bid” (1). This change persists into subsequent utterances and can be measured after a single exposure to auditory error (2). Additionally, speakers can simultaneously learn distinct adaptive changes to separate movements, for example different words (3). In previous work, we capitalized on this ability to demonstrate that participants can learn to increase their global acoustic contrast when exposed to a nonuniform perturbation field designed to reduce the perceived contrast between vowel sounds (4). Specifically, the perturbation field altered the auditory feedback of all vowels toward the center of the vowel space, making all vowels sound more like schwa. The present study examines the degree to which adaptation to such a nonuniform perturbation field can transfer to intermediate vowels not experienced during training.

On one hand, an extensive literature examining sensorimotor adaptation in limb movements has provided evidence that learning is primarily local, with generalization observed only when movements have considerable overlap: learning

*B. Parrell and C. A. Niziolek contributed equally to this work.

Correspondence: B. Parrell (bparrell@wisc.edu); C. A. Niziolek (cniziolek@wisc.edu).

Submitted 7 June 2024 / Revised 29 August 2024 / Accepted 27 September 2024

decays with distance (5), consistent with a local remapping of sensory-to-motor space in the region of the perturbation (6). Similar gradients of generalization have been reported in speech training with a single stimulus word, in which the transfer of learning to new spoken contexts is conditioned on the acoustic similarity between the trained and transfer words (7). On the other hand, when training movements encompass a wider workspace as in the nonuniform perturbation field used in Parrell and Niziolek (4), with multiple instances of local learning, transfer is readily observed within the workspace, with participants interpolating between extreme end points (5, 6, 8–14). Furthermore, when considering movements that share direction but differ in extent, learning generalizes from larger- to smaller-amplitude movements (15), though not the other way around. In vowel production, large movement amplitudes correspond to extreme degrees of constriction in the vocal tract (“corner” vowels), with noncorner vowels produced with lesser degrees of constriction at the same limited set of constriction locations (16). These results suggest that, contrary to local learning when adapting to perturbations of a single word or vowel, sensorimotor adaptation to a complex vowel-dependent perturbation field will show complete generalization to untrained vowels when the trained movements encompass the extremes of the articulatory workspace.

Our past work used words containing the English corner vowels (/i/, /æ/, /ɑ/, /u/) as training stimuli, maximizing the possibility of generalization by interpolating over the vowel space. In the present study, we replicate our past finding of increased acoustic contrast, and we additionally find evidence of strong generalization, in which learned changes to the production of corner vowels transfer to neighboring vowel sounds. These changes were largely specific to formant frequencies and were not accompanied by increases in duration or voice pitch, arguing against a strategy-based change to speaking style (i.e., a “clear speech” mode). Our results additionally show that the magnitude and direction of the generalization are locally conditioned by the nearby training vowels, suggesting that generalization occurs through the combination of local learning fields rather than through a global remapping of motor and sensory systems. This ready generalization is promising for the potential application of sensorimotor adaptation to increase acoustic contrast in speech sounds.

METHODS

Participants

Twenty-two individuals participated in the present study (14 female, 7 male; mean age $\pm$ SD: 28.1 $\pm$ 11.5 yr). All participants self-identified as native speakers of American English and reported no history of neurological, speech, or hearing disorders. All participants gave written informed consent before participation in the study and were compensated either monetarily or with course credit. All procedures were approved by the Institutional Review Board of the University of Wisconsin-Madison.

Auditory Perturbation

Participants’ speech was recorded with a head-mounted microphone (AKG C520), digitized at 48 kHz with an external

sound card (Focusrite Scarlett 2i2), and downsampled to 16 kHz and processed with Audapter (17, 18). Audapter identifies and shifts speech formants, using linear predictive coding (LPC) and signal filtering techniques. The filtered output of Audapter was played back to participants over closed-back circumaural headphones (Beyerdynamic DT 770). The latency of this process was $\sim$ 18 ms as measured following Kim and colleagues (19). The same process was used for all spoken trials whether or not a perturbation was applied, such that the latency was consistent throughout the experiment.

For most trials, participants’ speech (either veridical or altered) was played back to them at $\sim$ 80 dB SPL (volume based on productions during a preexperiment calibration phase, see below) mixed with speech-shaped noise at $\sim$ 60 dB SPL. The noise served to partially mask the participants’ unaltered speech that could otherwise be perceived through air or bone conduction. The volume of the speech signal varied dynamically with participants’ produced speech but was, in all cases, louder than the produced speech. On a subset of trials (see *Procedure*), participants did not receive feedback of their own speech; rather, they received only speech-shaped noise (*masking*). The amplitude of this noise was modulated by the participants’ speech, such that habitual speech volume resulted in a masking noise level of $\sim$ 80 dB SPL, and no noise was played while they were not speaking. The use of this amplitude-modulated noise severely limited participants’ ability to hear any feedback of their own speech, while avoiding some changes in speech typically caused by speaking in the presence of background noise (20–22).

A modified version of Audapter (4) was used to deliver auditory perturbations that depended on the current values of the first (F1) and second (F2) formants. A participant-specific perturbation field was created such that F1 and F2 were pushed toward the center of the vowel space, defined as the centroid of the quadrilateral formed by the four corner vowels of American English [/i/, /æ/, /ɑ/, and /u/ (Fig. 1A)] produced by each speaker during a preexperiment calibration phase (see below). The magnitude of the perturbation was 50% of the distance between the produced F1/F2 values and this center point (Fig. 1C). Although it is possible that the precise center of this space may vary slightly over time, the perturbation will nonetheless push all productions toward the general center of the working vowel space, resulting in reduced acoustic contrast between the vowels.

Procedure

Stimuli were presented on an LED computer screen, with one word presented per trial. Participants were instructed to read each word out loud as it appeared. Each stimulus was presented for 1.5 s, with a randomly jittered interstimulus period between 0.75 and 1.5 s. There were two sets of stimuli used. The *training* stimuli were presented during the baseline and exposure phases (Fig. 1C, shown in red). These consisted of the words “bead,” “bad,” “bod,” and “booed” containing the four corner vowels of American English (/i/, /æ/, /ɑ/, and /u/, respectively). The *generalization* stimuli (“bid,” “bayed,” “bed,” “bud,” “bode”) contained noncorner vowels (/ɪ/, /e/, /ɛ/, /ʌ/, and /o/, respectively). These stimuli

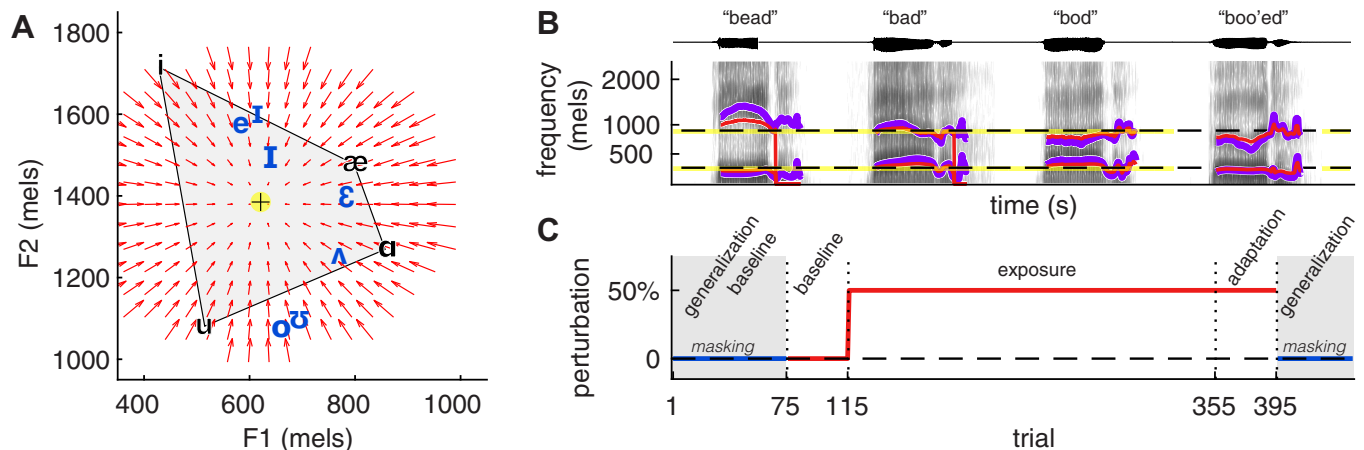


Figure 1. Experimental protocol. **A:** example perturbation field used in the exposure phase. All formants are perturbed toward the center of the vowel space (yellow) for training vowels (black). Vowels used in generalization phases (blue) are shown for illustration only and were not produced under perturbation. **B:** example spectrograms of the 4 training words showing the produced formants (purple), perturbed formants (red), and vowel space center (black dashed line over yellow). **C:** trial structure. Trials with generalization words are shown in blue and trials with training words in red. Loud masking noise (~80 dB SPL) was played during both the generalization baseline phase and generalization phase.

were presented during the generalization baseline and generalization phases (Fig. 1C, shown in blue). Stimuli were randomly ordered within groups of four (training stimuli) or five (generalization stimuli) during the experiment, with each word occurring a single time per group. Before the main experiment, participants completed a brief calibration phase, where they produced 10 repetitions of each training stimulus. These productions were used by the experimenter to determine participant-specific values for the LPC order used by Audapter in the main experiment. In the case where the default LPC order for the participant resulted in mistracked formants (visually represented using a custom GUI), a different LPC order that resulted in tracks without errors was chosen.

The main experiment consisted of four phases (Fig. 1C):

- 1) The generalization baseline phase, where participants produced 15 repetitions of each of the generalization stimuli. Participants received masking noise with no playback of their own speech in this phase to provide similar conditions as in the generalization phase, described below.
- 2) The second baseline phase, where participants produced 10 repetitions of each of the training stimuli. Participants heard their own speech over the headphones, but no perturbation was applied. Formant values from this phase (taken from the middle 50% of the vowel) were used to determine the participant-specific perturbation field used in the following phase (see *Auditory Perturbation*).
- 3) The exposure phase, where participants produced 60 repetitions of each training stimulus. Participants heard their own speech with formants shifted toward the center of the vowel space. The magnitude of the shift was held constant at 50% of the distance between the produced formant values and the vowel space center.
- 4) The generalization phase, where participants produced 10 repetitions of each generalization stimulus. As in the first baseline phase, participants received masking noise

over the headphones without playback of their own speech. This masking noise served to prevent the deadadaptation that occurs when participants are exposed to veridical feedback after exposure to altered auditory feedback.

A short, self-timed break was given every 30 trials.

Data Analysis

Data from one participant were excluded from analysis because of equipment failure during the final portion of the training phase, leaving a total of 21 analyzed participants.

Formants were tracked with wave_viewer (23), a MATLAB GUI interface to formant tracking with Praat (24). Vowel onset and offset were initially determined using a participant-specific amplitude threshold. These events were then hand-checked and corrected as needed, using landmarks from the spectrogram: vowel onset was determined as the time when formants were first visible, and vowel offset was determined as the time where vowel formants, particularly F2 and higher formants, were no longer visible. Within this window, F1 and F2 were tracked using participant-specific values for LPC order and preemphasis. Errors in formant tracking were corrected by adjusting the LPC order and/or preemphasis. A small number of trials with unresolvable formant tracking errors or speech production errors (saying the wrong word, coughing, etc.) were discarded (1.5% in total, 0–6.5% across participants). Single values for F1 and F2 for each trial were calculated as the mean values in the middle 50% of the vowel and converted to the mel scale.

Our primary outcome measure was the average of the distance in F1/F2 space between each pair of stimuli vowels, or *average vowel spacing* (AVS). We chose AVS for two reasons. First, AVS can be used with any set of vowels, including the noncorner vowels in our generalization stimuli, whereas alternative metrics of vowel contrast, such as *vowel space area* (VSA), only take the four corner vowels into account. Second, our previous work (4) has shown that AVS is a more sensitive metric of vowel spacing for measuring adaptation to the current vowel centralization perturbation compared to VSA.

We calculated AVS separately for each block of 10 stimulus repetitions (50 trials per block for the 5 generalization stimuli and 40 trials per block for the 4 training stimuli). The first 25 trials of the first baseline block were excluded from this analysis to give participants time to adjust to speaking under masking noise. Subsequently, each block's AVS value was normalized by dividing the AVS value in the corresponding baseline block (containing the same stimuli).

To measure the degree of adaptation to the auditory perturbation, we used the normalized AVS values calculated in the last block of the exposure phase. To measure generalization of learning, we used the normalized AVS value from the generalization phase. Two-tailed *t* tests were used to determine whether AVS values in each phase differed from 0. To examine the relationship between adaptation and generalization, we used Spearman's rank correlation. This allows us to assess whether, across participants, increased learning was related to increased generalization, without making assumptions about the nature of the relationship between AVS calculated over different sets of vowels. Inferential statistical analysis was conducted in R (25), except where noted.

In addition to the global AVS metric, we additionally measured vowel-specific adaptive changes. Because the perturbation was toward the center of the vowel space, we defined vowel-specific adaptation as movement that increased the distance from this center point. To do this, the average F1 and F2 for each vowel were calculated in the baseline phase as well as in the end of the exposure (adaptation stimuli) or generalization (generalization stimuli) phases. From these average values, the average Euclidean distance in F1-F2 space to the center of the vowel distribution was calculated. We examined whether the magnitude of adaptation differed between the baseline and exposure and baseline/generalization phases, using a mixed-effect ANOVA with magnitude as the dependent variable, phase and vowel as the independent variables, the interaction between these factors, and a random effect of participant. A Greenhouse–Geisser correction was automatically applied when sphericity was violated. Post hoc comparisons for each vowel across conditions were conducted by paired *t* tests with a false discovery rate correction for *P* values (26).

We additionally tested whether the magnitude of adaptation transfer in each of the generalization stimuli was related to its distance from each of the four training vowels. We assumed that the influence of each training vowel on a particular generalization vowel would scale inversely with the distance in mels between the two (7). To estimate the relationship between generalization magnitude and acoustic distance, we fit the following linear model, where the magnitude of adaptation in the generalization vowels (mag_{gen}) is predicted by the magnitude of adaptation in the training vowels (mag_{iy} , mag_{ae} , mag_{aa} , mag_{uw}) and the distance from each generalization vowel to each of the training vowels (dist_{iy} , dist_{ae} , dist_{aa} , dist_{uw}):

$$\text{mag}_{\text{gen}} \sim B_0 + B_1 * (\text{mag}_{\text{iy}}/\text{dist}_{\text{iy}} + \text{mag}_{\text{ae}}/\text{dist}_{\text{ae}} + \text{mag}_{\text{aa}}/\text{dist}_{\text{aa}} + \text{mag}_{\text{uw}}/\text{dist}_{\text{uw}})$$

A mixed-effects model with a random intercept by participant returned singular fits, so the linear model without this term was used.

In addition to the analysis of the magnitude of adaptation, we additionally examined the direction of this change in F1/F2 space, calculated as the angle between the vector connecting the average F1/F2 values in the baseline and exposure/generalization phases and the *x*-axis, such that 0° reflects an increase in F1 with no change in F2. These analyses were conducted with the *CircStat* software package in MATLAB (27). First, we used a nonparametric multisample test for equal medians (*circ_cmtest*) to test whether the angle of adaptation differed between the vowels, using a model with adaptation angle as the dependent variable and vowel category as the independent variable. Pairwise post hoc tests were conducted using the same test, with *P* values adjusted using the false discovery rate. Second, we tested whether the distance between training and generalization vowels had an effect on the angle of change in the generalization vowels by constructing a circular regression function with the adaptation angle as the dependent variable and the circular weighted mean angle of the training vowels as the independent variable, where the weights were set to the inverse of the distance between each particular generalization vowel and each training vowel ($w_{\text{iy}} = 1/\text{dist}_{\text{iy}}$, etc.).

To assess whether changes observed in response to the auditory perturbation could potentially be related to a change in speaking style, we measured several speech parameters that undergo increases in clear speech, including vowel duration, peak amplitude, maximum pitch, and pitch range (4). For each of these measurements, repeated-measures ANOVAs were conducted with factors of vowel and phase (baseline, end of exposure/generalization) and the interaction between these two factors.

Finally, we calculated cross-correlations between adaptation magnitude (linear correlations) and angle (circular correlations) across all nine stimulus vowels (4 training vowels and 5 generalization vowels). Where applicable, effect sizes [partial eta-squared (η_p^2), Hedges' *g*] were calculated with the *effectsize* package (28). Summary statistics are reported as means and standard errors.

RESULTS

We examined whether exposure to an auditory feedback perturbation that reduces vowel contrast in the four corner vowels of English (/i/, /æ/, /a/, /u/) would 1) lead to an increase in vowel contrast in these “trained” vowels and 2) additionally cause an increase in vowel contrast in five non-corner “generalization” vowels (/ɪ/, /e/, /ɛ/, /ʌ/, /o/). Participants strongly adapted to the auditory perturbation in the exposure phase, producing a mean increase in AVS of $6.1 \pm 1.7\%$ at the end of this phase [$t(20) = 60.1$, $P < 0.0001$, $g = 12.62$; Fig. 2B]. This increased contrast transferred to the generalization stimuli, which were produced with an AVS $7.3 \pm 1.7\%$ higher in the generalization phase than in the generalization baseline [$t(20) = 63.2$, $P < 0.0001$, $g = 13.27$; Fig. 2B]. Across participants, there was a strong positive relationship between the AVS increase in the training vowels and that of the transfer vowels ($\rho = 0.54$, $P = 0.014$; Fig. 2D).

We additionally examined how individual vowels were affected by the centralization perturbation (Fig. 3A). We found that vowels were produced significantly farther from center of the vowel space at the end of the exposure/

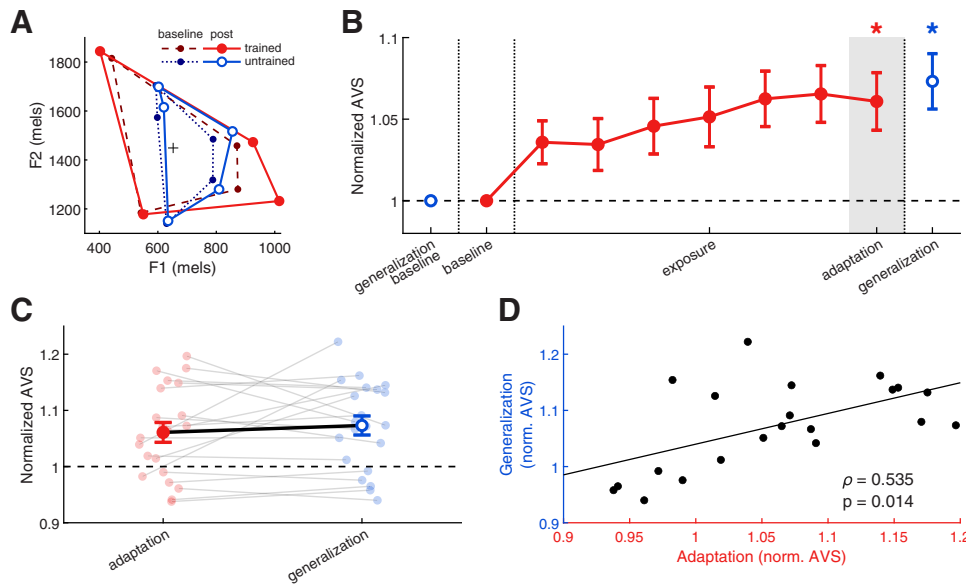


Figure 2. Participants' ($n = 21$) generalized changes from trained to untrained vowels. **A:** vowel space from a representative participant, showing an expansion from baseline (dashed lines, small markers) to postexposure productions (solid lines, large markers). Untrained (generalization) vowels are shown in blue and trained vowels in red. **B:** normalized average vowel spacing (AVS) over the course of the experiment. Normalization is specific to each set of vowels. *Value different from 0 ($P < 0.0001$). **C:** AVS at the end of the exposure phase (adaptation, red) and in the generalization phase (blue). Lighter dots connected by gray lines represent individual participants. Darker dots connected by a black line represent group means with SE. **D:** correlation between adaptation and generalization.

generalization phases compared to the baseline phases [mean increase of $19.6 \pm \text{SE } 2.6$ mels, $F(8,359) = 81.8$, $P < 1e-15$, $\eta_p^2 = 0.65$]. Whereas vowels varied in their distance to the center of the vowel space, as expected [$F(1,359) = 11.0$, $P = 0.001$, $\eta_p^2 = 0.03$], there was no difference in the magnitude of their adaptive movement away from the center [interaction $F(8,359) = 0.6$, $P = 0.79$, $\eta_p^2 = 0.01$]. Post hoc tests on individual vowels showed that the difference between baseline and exposure/generalization phases was significant for /i/, /e/, /æ/, /a/, and /ʌ/ (all corrected $P < 0.01$).

In terms of the angle of change in formant space, most vowels showed changes that moved in the direction opposite

the vowel center (Fig. 3B). For the training vowels, this movement opposed the applied perturbation. The exception is /u/, for which minimal changes were observed. Overall, there was a significant difference between the vowels in the angle of this change ($P = 28.2$, $P < 0.0001$). Individually, the angle for /i/ differed from the angles in /a/, /æ/, and /ʌ/, and the angle for /i/ differed from the angles in /æ/ and /e/ (corrected $P < 0.05$). These results are consistent with vowel-specific adaptive changes in response to the different perturbations delivered to the four training vowels.

Finally, we examined whether the distance, in F1/F2 space, between the generalization and training vowels influenced

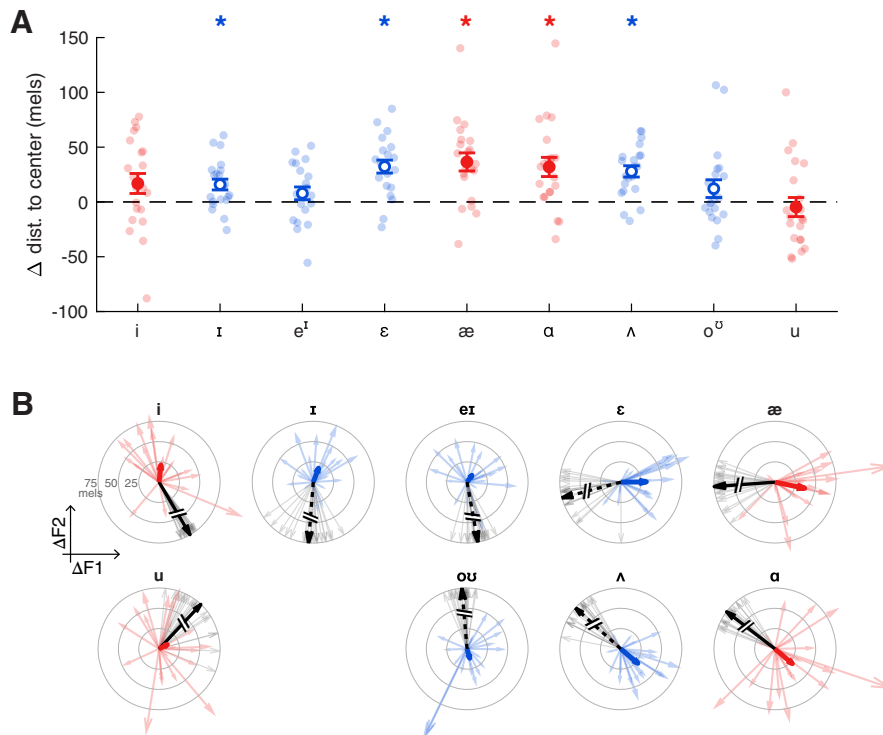


Figure 3. Change in individual vowels ($n = 21$ participants). Training vowels are shown in red (filled circles) and generalization vowels in blue (open circles). **A:** change (Δ) in distance to the center of the vowel space, with positive values indicating an increase in this distance compared to the baseline phase. Small dots represent individual participants. Large dots represent group means with SE. *Significant difference from 0 (corrected $P < 0.05$). **B:** F1-F2 change vectors from baseline to the end of the exposure phase (adaptation, red) or the generalization phase. Thin lighter-colored lines represent individual participants; thick darker-colored lines represent the group response (mean change vector). Thin light gray arrows indicate the participant-specific angle of the auditory perturbation; thick black arrows represent the group mean angle. Note that the magnitudes of auditory perturbations varied across participants and trials and were much larger than shown; only the angles are represented here. Dotted arrows indicate the average direction of the perturbations that would have been applied for generalization vowels.

the magnitude and direction of changes in the generalization vowels. Previous work has suggested that the magnitude of transfer of learning following adaptation of a single vowel is related to the acoustic similarity between training and generalization tokens (7). In the present study, we found a significant relationship between the magnitude of adaptation for each generalization vowel and a weighted sum of the adaptation in the training vowels, where weights decrease with increased distance to the generalization vowel ($\beta = 28.3$, $t = 4.7$, $P < 0.0001$, adjusted $R^2 = 0.17$). This indicates that the magnitude of transfer was indeed related to the distance between the generalization and training vowels. We similarly found a significant positive circular correlation between the angle of adaptation in the generalization vowels and a measure of adaptation angle in the training vowels, weighted by distance ($r = 0.32$, $P < 0.0002$).

To further investigate how generalization related to patterns of adaptation, we calculated regression coefficients (Pearson's r) between each pair of vowels for both the magnitude of change from baseline as well as the angle of change (Fig. 4). As expected, we saw the highest correlations for vowels near to one another in F1/F2 space. Front generalization vowels (/i/, /e/, /ε/) were most similar to front training vowels (/i/, /æ/); low generalization vowels (/ε/ and particularly /Λ/) were most similar to the low training vowels (/æ/, /a/), while the back generalization vowel /o^u/ was most similar to the back training vowel /u/. Overall, the pattern of correlations was similar for adaptation angle, though somewhat less reliable.

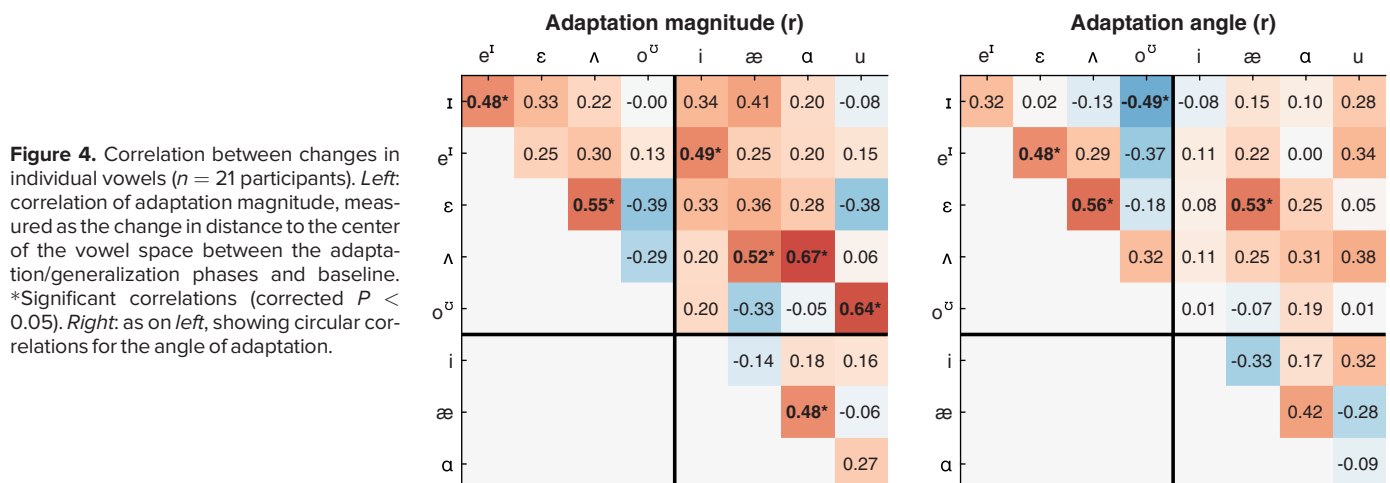
Finally, we confirmed that the changes we observed in these vowels are unlikely to be due to participants adopting a clear speech style. Repeated-measures ANOVAs showed that vowel duration, pitch range, and maximum pitch did not differ significantly between the baseline/generalization baseline phases and the end of exposure/generalization phases (main effect of phase and interaction of phase with vowel all $P > 0.14$). Although we did find a minimal, but significant, increase in intensity in the latter phases [$F(1,342) = 7.4$, $P = 0.007$; baseline: 11.4 ± 2.5 arbitrary units (a.u.); exposure/generalization: 14.8 ± 3.2], this increase was similar in magnitude to the increase in intensity we observed in a previous study using the same adaptation paradigm that

occurred both when the perturbation was applied as well as in a control session with no perturbation (an increase of ~ 3 a.u. in Ref. 4). Thus, this increase in intensity in the present study likely reflects a change due to repeated production of the stimuli used here rather than an increase in speech clarity per se.

DISCUSSION

Auditory feedback perturbations elicit speech motor learning, driving changes to pronunciation that reduce sensory error. In the present study, we confirmed that a nonuniform perturbation field can drive differential changes to the production of different vowel sounds, resulting in an increase in the average spacing between vowels (AVS) and thus in acoustic contrast. Crucially, we show that this learning generalizes to vowels not produced during the training, giving rise to an increase in contrast of a similar magnitude. The increase in contrast was achieved without a concomitant increase in syllable duration, strongly suggesting that it is the result of successful sensorimotor learning and not the result of a clear speech mode. The magnitude and direction of adaptation in the generalization words were highly predicted by those of the trained words, both at the participant level and at the level of individual vowels, suggesting the generalization of learning to neighboring vowel sounds.

These findings are consistent with studies of speech learning and generalization employing a single uniform auditory feedback transformation, which reported greater generalization to utterances that were closer in acoustic space (7, 29). In the present study, the nonuniform *centralization* perturbation field additionally introduced conflicts between the required adaptive changes for vowels occupying different regions of formant space (e.g., a raising of the tongue for /i/ and a lowering of the tongue for /æ/); nevertheless, the adaptive changes in untrained vowels were appropriate to the centralization field, resembling the formant changes of their neighbors (e.g., /i/ patterned with /i/ while /ε/ patterned with /æ/). The generalization seen here is expected given similar findings in limb control, in which changes in an untrained region of a spatial workspace reflect the interpolation of local learning at multiple surrounding regions



(11). This pattern of learning can be explained by a model in which speakers learn multiple local sensorimotor transformations that decay with distance away from the observed perturbation in an extrinsic coordinate system (e.g., radial basis functions in vowel formant space), with effects at untrained regions of vowel space defined by the superposition of the local transformations (6). In other words, participants interpolate between transformations learned at different parts of the workspace (30).

For this reason, we chose corner vowels as the targets of training. They cover the extent of the available vowel space, representing the most extreme vocal tract configurations (e.g., the tongue is at its highest and most fronted position for /i/ and at its lowest and most backed position for /a/) and consequently resulting in the most extreme formant values. More central vowels, then, can be seen not only as intermediate positions in an extrinsic coordinate space (i.e., formant space) but also as less extreme movements with vocal tract constrictions that differ in extent. Given past findings on generalization across movement amplitudes (15, 31), this is precisely the situation in which learning would be expected to generalize; we capitalized on these circumstances to train a limited number of vowel targets and elicit learned changes across the vowel space.

A perennial question in speech motor control is the nature of the planning or representational space [e.g., are vowels planned in acoustic space or as constrictions in the vocal tract? (32–34)], similar to questions in limb control about extrinsic/visual or intrinsic/joint angle control (12). Similarly, the precise level of representation at which adaptation occurs and generalizes has received little direct evaluation. One study has presented evidence that adaptation in speech may transfer based on distance in acoustic space (7); conversely, recent modeling work has shown that adaptation could occur by modifications to the transformation between the articulatory (muscle/articulator position) and vocal tract constriction levels of representation (35). Given the difficulty in distinguishing between acoustic and constriction-based goals in speech (36), careful experimental work will be needed in the future to resolve the representational basis of both learning and generalization in speech adaptation.

As is typical in speech sensorimotor adaptation, there was considerable variability across participants in the magnitude of adaptation, and not all participants increased their AVS over the course of the experiment. However, prior work has indicated that AVS goes down over time when feedback is veridical (4), which may mask effects of adaptation in the participants who did not appear to increase their AVS.

A perennial problem in speech therapy is maximizing the effect of the limited training available to most patients. The generalization of learned improvements to the production of speech sounds is therefore of considerable interest and utility. In this experiment, we show that exposure to a nonuniform formant perturbation at the extremes of vowel space shows strong generalization to productions at intermediate locations, resulting in increased acoustic contrast. For future clinical applications, further work is needed to explore whether the results found here for isolated word production are also found in running speech (37). Nonetheless, as increased contrast is a predictor of intelligibility, this paradigm has the

potential to drive behaviorally relevant changes that improve communication effectiveness.

DATA AVAILABILITY

All data and analysis scripts can be found at <https://doi.org/10.17605/OSF.IO/78Z93>.

GRANTS

This work was supported by Grants R01 DC019134 (C.A.N. and B.P.), R01 DC017091 (B.P.), and P50HD105353 (Waisman Center) from the National Institutes of Health and BCS 2120506 (C.A.N. and B.P.) from the National Science Foundation.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

B.P. and C.A.N. conceived and designed research; T.C. performed experiments; B.P. and T.C. analyzed data; B.P., C.A.N., and T.C. interpreted results of experiments; B.P. and C.A.N. prepared figures; B.P. and C.A.N. drafted manuscript; B.P. and C.A.N. edited and revised manuscript; B.P., C.A.N., and T.C. approved final version of manuscript.

REFERENCES

1. Houde JF, Jordan MI. Sensorimotor adaptation in speech production. *Science* 279: 1213–1216, 1998. doi:10.1126/science.279.5354.1213.
2. Hantzsch L, Parrell B, Niziolek CA. A single exposure to altered auditory feedback causes observable sensorimotor adaptation in speech. *eLife* 11: e73694, 2022. doi:10.7554/eLife.73694.
3. Rochet-Capellan A, Ostry DJ. Simultaneous acquisition of multiple auditory-motor transformations in speech. *J Neurosci* 31: 2657–2662, 2011. doi:10.1523/JNEUROSCI.6020-10.2011.
4. Parrell B, Niziolek CA. Increased speech contrast induced by sensorimotor adaptation to a nonuniform auditory perturbation. *J Neurophysiol* 125: 638–647, 2021. doi:10.1152/jn.00466.2020.
5. Gandolfo F, Mussa-Ivaldi FA, Bizzi E. Motor learning by field approximation. *Proc Natl Acad Sci USA* 93: 3843–3846, 1996. doi:10.1073/pnas.93.9.3843.
6. Ghahramani Z, Wolpert DM, Jordan MI. Generalization to local remappings of the visuomotor coordinate transformation. *J Neurosci* 16: 7085–7096, 1996. doi:10.1523/JNEUROSCI.16-21-07085.1996.
7. Rochet-Capellan A, Richer L, Ostry DJ. Nonhomogeneous transfer reveals specificity in speech motor learning. *J Neurophysiol* 107: 1711–1717, 2012. doi:10.1152/jn.00773.2011.
8. Atkeson CG. Learning arm kinematics and dynamics. *Annu Rev Neurosci* 12: 157–183, 1989. doi:10.1146/annurev.ne.12.030189.001105.
9. Donchin O, Francis JT, Shadmehr R. Quantifying generalization from trial-by-trial behavior of adaptive systems that learn with basis functions: theory and experiments in human motor control. *J Neurosci* 23: 9032–9045, 2003. doi:10.1523/JNEUROSCI.23-27-09032.2003.
10. Huang VS, Shadmehr R. Evolution of motor memory during the seconds after observation of motor error. *J Neurophysiol* 97: 3976–3985, 2007. doi:10.1152/jn.01281.2006.
11. Malfait N, Gribble PL, Ostry DJ. Generalization of motor learning based on multiple field exposures and local adaptation. *J Neurophysiol* 93: 3327–3338, 2005. doi:10.1152/jn.00883.2004.
12. Mattar AA, Ostry DJ. Modifiability of generalization in dynamics learning. *J Neurophysiol* 98: 3321–3329, 2007. doi:10.1152/jn.00576.2007.

13. **Thoroughman KA, Shadmehr R.** Learning of action through adaptive combination of motor primitives. *Nature* 407: 742–747, 2000. doi:10.1038/35037588.
14. **Thoroughman KA, Taylor JA.** Rapid reshaping of human motor generalization. *J Neurosci* 25: 8948–8953, 2005. doi:10.1523/JNEUROSCI.1771-05.2005.
15. **Mattar AA, Ostry DJ.** Generalization of dynamics learning across changes in movement amplitude. *J Neurophysiol* 104: 426–438, 2010. doi:10.1152/jn.00886.2009.
16. **Gaines JL, Kim KS, Parrell B, Ramanarayanan V, Nagarajan SS, Houde JF.** Discrete constriction locations describe a comprehensive range of vocal tract shapes in the Maeda model. *JASA Express Lett* 1: 124402, 2021. doi:10.1121/10.0009058.
17. **Cai S, Boucek M, Ghosh S, Guenther FH, Perkell J.** A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin triphthong /iaʊ/. In: *Proceedings of the 8th International Seminar on Speech Production, Strasbourg, France*. 2008, p. 65–68.
18. **Tourville JA, Cai S, Guenther F.** Exploring auditory-motor interactions in normal and disordered speech. *Proc Mtgs Acoust* 19: 060180, 2013. doi:10.1121/1.4800684.
19. **Kim KS, Wang H, Max L.** It's about time: minimizing hardware and software latencies in speech research with real-time auditory feedback. *J Speech Lang Hear Res* 63: 2522–2534, 2020. doi:10.1044/2020_JSLHR-19-00419.
20. **Lombard E.** Le signe de l'elevation de la voix. *Ann Mal Oreille Larynx* 37: 2, 1911.
21. **Summers WV, Pisoni DB, Bernacki RH, Pedlow RI, Stokes MA.** Effects of noise on speech production: acoustic and perceptual analyses. *J Acoust Soc Am* 84: 917–928, 1988. doi:10.1121/1.396660.
22. **Summers WV, Johnson K, Pisoni DB, Bernacki RH.** An addendum to "Effects of noise on speech production: acoustic and perceptual analyses" [*J Acoust Soc Am* 84: 917–928 (1988)]. *J Acoust Soc Am* 86: 1717–1721, 1989. doi:10.1121/1.398602.
23. **Niziolek CA, Houde J.** Wave_Viewer: First release. 2015.
24. **Boersma P, Weenink D.** Praat: doing phonetics by computer (Online). 2019. <http://www.praat.org/>.
25. **R Core Team.** R: a language and environment for statistical computing (Online). R Foundation for Statistical Computing. <http://www.R-project.org/>.
26. **Benjamini Y, Hochberg Y.** Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 57: 289–300, 1995. doi:10.1111/j.2517-6161.1995.tb02031.x.
27. **Berens P.** CircStat: a MATLAB toolbox for circular statistics. *J Stat Softw* 31: 1–21, 2009. doi:10.18637/jss.v031.i10.
28. **Ben-Shachar MS, Lüdtke D, Makowski D.** effectsize: Estimation of effect size indices and standardized parameters. *J Open Source Softw* 5: 2815, 2020. doi:10.21105/joss.02815.
29. **Reilly KJ, Pettibone C.** Vowel generalization and its relation to adaptation during perturbations of auditory feedback. *J Neurophysiol* 118: 2925–2934, 2017. doi:10.1152/jn.00702.2016.
30. **Ghahramani Z, Wolpert DM.** Modular decomposition in visuomotor learning. *Nature* 386: 392–395, 1997. doi:10.1038/386392a0.
31. **Krakauer JW, Pine ZM, Ghilardi MF, Ghez C.** Learning of visuomotor transformations for vectorial planning of reaching trajectories. *J Neurosci* 20: 8916–8924, 2000. doi:10.1523/JNEUROSCI.20-23-08916.2000.
32. **Guenther FH, Hampson M, Johnson D.** A theoretical investigation of reference frames for the planning of speech movements. *Psychol Rev* 105: 611–633, 1998. doi:10.1037/0033-295x.105.4.611-633.
33. **Goldstein L, Fowler CA.** Articulatory phonology: a phonology for public language use. In: *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, edited by Schiller NO, Meyer A. Berlin: Mouton de Gruyter, 2003, p. 159–207.
34. **Perrier P, Fuchs SF.** Motor equivalence in speech production. In: *The Handbook of Speech Production*, edited by Redford M. Hoboken, NJ: Wiley-Blackwell, 2015, p. 225–247.
35. **Kim KS, Gaines JL, Parrell B, Ramanarayanan V, Nagarajan SS, Houde JF.** Mechanisms of sensorimotor adaptation in a hierarchical state feedback control model of speech. *PLoS Comput Biol* 19: e1011244, 2023. doi:10.1371/journal.pcbi.1011244.
36. **Iskarous K.** Vowel constrictions are recoverable from formants. *J Phon* 38: 375–387, 2010. doi:10.1016/j.wocn.2010.03.002.
37. **Lametti DR, Smith HJ, Watkins KE, Shiller DM.** Robust sensorimotor learning during variable sentence-level speech. *Curr Biol* 28: 3106–3113.e2, 2018. doi:10.1016/j.cub.2018.07.030.