Increased speech contrast induced by sensorimotor adaptation to a non-uniform auditory perturbation

Benjamin Parrell[‡] & Caroline A. Niziolek[‡]
Waisman Center, University of Wisconsin–Madison, 1500 Highland Ave, Madison, WI, 53705, USA
Department of Communication Sciences & Disorders, University of Wisconsin–Madison, 1975 Willow Dr, Madison, WI, 53706, USA
[‡] these authors contributed equally to this work

Running head: Increased speech contrast induced by sensorimotor adaptation

Correspondence: Benjamin Parrell, Waisman Center, University of Wisconsin–Madison, 1500 Highland Ave, Madison, WI, 53705, USA
bparrell@wisc.edu

Supplemental Material available at: https://doi.org/10.6084/m9.figshare.13244378

Abstract

When auditory feedback is perturbed in a consistent way, speakers learn to adjust their speech to compensate, a process known as sensorimotor adaptation. While this paradigm has been highly informative for our understanding of the role of sensory feedback in speech motor control, its ability to induce behaviorally-relevant changes in speech that affect communication effectiveness remains unclear. Because reduced vowel contrast contributes to intelligibility deficits in many neurogenic speech disorders, we examine human speakers' ability to adapt to a non-uniform perturbation field which was designed to affect vowel distinctiveness, applying a shift that depended on the vowel being produced. Twenty-five participants were exposed to this "vowel centralization" feedback perturbation in which the first to formant frequencies were shifted towards the center of each participant's vowel space, making vowels less distinct from one another. Speakers adapted to this non-uniform shift, learning to produce corner vowels with increased vowel space area and vowel contrast to partially overcome the perceived centralization. The increase in vowel contrast occurred without a concomitant increase in duration and persisted after the feedback shift was removed, including after a 10-minute silent period. These findings establish the validity of a sensorimotor adaptation paradigm to increase vowel contrast, showing that complex, non-uniform alterations to sensory feedback can successfully drive changes relevant to intelligible communication.

New & Noteworthy

To date, the speech motor learning evoked in sensorimotor adaptation studies has had little ecological consequences for communication. By inducing complex, non-uniform acoustic errors, we show that adaptation can be leveraged to cause an increase in speech sound contrast, a change that has the capacity to improve intelligibility. This study is relevant for models of sensorimotor integration across motor domains, showing that complex alterations to sensory feedback can successfully drive changes relevant to ecological behavior.

Keywords

speech motor control, sensorimotor adaptation, motor learning, speech intelligibility, vowel contrast

Introduction

Real-time alterations of auditory feedback have been used extensively over the past 20 years to probe the sensorimotor control system for speech. In these studies, speech is recorded, processed, altered, and played back to participants in near real-time, enabling experimental manipulation of acoustic parameters. For example, altering the first and second resonant frequencies of the vocal tract, or formants (F1/F2), can cause the perception of a different vowel (Houde and Jordan 1998, 2002; Lametti et al. 2018; Munhall et al. 2009; Parrell et al. 2017; Purcell and Munhall 2006; Rochet-Capellan and Ostry 2011; Villacorta et al. 2007). Over the course of many repetitions, participants learn to alter their speech to oppose the perturbation, a process known as *sensorimotor adaptation*. The ability of sensorimotor adaptation to cause rapid changes in speech without conscious control or awareness (Munhall et al. 2009) has made it a promising potential avenue for rehabilitation in motor disorders, where current treatments rely on intensive training over an extended time period. However, this promise remains unfulfilled, as most current adaptation paradigms lack behaviorally-relevant outcomes (Roemmich and Bastian 2018). In speech, the vast majority of studies to date have examined a single vowel (often a single word), with a single perturbation in F1/F2 space. While this paradigm can cause local changes in the production of a particular vowel, these changes have little ecological consequences for communication.

Critically, the intelligibility impairments prevalent in many speech disorders have been linked to acoustic parameters that adaptation can target: formant contrast between different vowels (Ansel and Kent 1992; Bang et al. 2013; Kim et al. 2011; Neel 2008; Skodda et al. 2011; Tjaden et al. 2005; Weismer et al. 2001). An increase in this contrast is often taken as an outcome metric in research on speech rehabilitation (Lam and Tjaden 2016; Sapir et al. 2007; Tjaden et al. 2013; Tjaden and Wilding 2004; Whitfield and Goberman 2014). Because increasing contrast in acoustic space relies on changing the produced formant frequencies for different vowels, this is a particularly promising avenue for the application of sensorimotor adaptation, though it would necessarily entail more complex perturbation paradigms. A few studies have begun to move beyond the common paradigm of single perturbations applied to a single vowel. Recent work has shown that speakers will adapt to a constant perturbation applied to read sentences (Lametti et al. 2018), and can simultaneously adapt to two opposite perturbations that are applied to different monosyllabic words, even when these words share the same vowel (Rochet-Capellan and Ostry 2011). Together, these results indicate that more complex paradigms, such as those that would be needed to enhance vowel contrasts globally, are learnable.

In the current study, we implemented an auditory perturbation explicitly designed to enhance the acoustic contrast between vowels. We accomplished this through a non-uniform formant perturbation paradigm that pushes all vowels towards the center of the vowel space (Figure 1A), making them less distinct from one another. We measured how participants adapt their speech to oppose this perturbation, and test whether this leads to increased vowel contrast in the four "corner" vowels of English (/i/ as in *bead*, /æ/ as in *bad*, /ɑ/ as in *bod*, and /u/ as in

booed). We additionally tested how these changes were retained immediately after the perturbation was removed as well as after a 10-minute delay. To control for any changes in speech over the course of the experiment, all participants completed an additional *control* session with the same structure as the main *adapt* session, only with no auditory perturbation applied. We include two global measures of speech articulation as outcome variables. *Vowel space area* (VSA), the area of the irregular quadrilateral formed by the four corner vowels (Neel 2008), is the most common metric of global vowel contrast and allows us to compare our results to previous work. We also include a measure of *average vowel spacing* (AVS), the average of the pairwise distances between the four vowels, which may be more sensitive to changes in vowel contrast and a better predictor of intelligibility (Neel 2008). To disambiguate whether changes in speech production are caused by sensorimotor adaptation or by a general hyperarticulation mechanism that occurs in contexts that encourage clearer speech, we include measures of duration and several other speech parameters associated with clear speech (pitch range, maximum pitch, and amplitude).

--- Figure 1 ---

Materials and Methods

Participants

Twenty-five participants were tested in the current study (21 female/4 male, mean age ± standard deviation: 20.4 ± 2.9 years). This is slightly larger than previous studies on sensorimotor adaptation in speech, most of which have used 10-20 participants (e.g., Lametti et al. 2018; Munhall et al. 2009; Villacorta et al. 2007). All participants were native speakers of American English, without any reported history of neurological, speech, or hearing disorders. Participants gave informed consent prior to 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

Auditory feedback was recorded, altered, and played back to participants using Audapter (Cai et al. 2008; Tourville et al. 2013). Speech was recorded at 16 kHz via a head-mounted microphone (AKG C520), digitized with a Focusrite Scarlett sound card, and sent to a desktop workstation. The speech signal was then perturbed using Audapter, which identifies the vowel formants using linear predictive coding (LPC) and filters the speech signal to introduce a shift to those formants (details of the applied shift are given below). If no shift is applied, Audapter outputs the unmodified input signal with the same processing delay. The output of Audapter was played back to participants via closed-back circumaural headphones (Beyerdynamic DT 770) through all phases of the experiment. The measured latency of audio playback on our system was ~18 ms. Speech was played back at a volume of approximately 80 dB SPL and mixed with speech-shaped noise at approximately 60 dB SPL. The noise served to mask potential perception of the

participants' own unaltered speech, which may have otherwise been perceptible through air or bone conduction.

We employed a modified version of Audapter that is able to specify formant perturbations as a function of the current values of F1 and F2. A participant-specific *perturbation field* was calculated such that all vowels were pushed towards the center of that particpant's vowel space (Fig. 1A). This central point was defined as the centroid of the quadrilateral formed by the four corner vowels of English (/i/, /æ/, /ɑ/, /u/). The magnitude of the perturbation was defined as a percentage of the distance between the currently produced vowel formants and the center of the vowel space. The magnitude varied across the experiment, ramping up to 50% of the distance between the current formant values and those at the center of the vowel space (Fig. 1B). Acoustically, this had the effect of centralizing all produced vowels.

Stimuli and trial structure

Stimuli consisted of four English words containing the four corner vowels in a /bVd/ context: *bead, bad, bod,* and *booed* (containing the vowels /i/, /æ/, /ɑ/, and /u/, respectively). 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. The interstimulus interval was randomly jittered between 0.75-1.5 s. All participants produced the stimuli with the intended vowels. Stimuli were randomly ordered within groups of four trials during the experiment, such that each word was repeated once per group.

Experimental Procedures

The experiment consisted of six phases (below). Participants received auditory feedback through headphones in all phases of the experiment.

1. A 40-trial *calibration* phase, which was used to define a participant-specific LPC order for formant tracking in Audapter. No perturbation was given during the calibration phase.
2. A 60-trial *baseline* phase, which was used to measure the participants' baseline formant values for the four corner vowels of English. No perturbation was given during the baseline phase. These values were used to calculate the participant-specific perturbation field.
3. A 40-trial *ramp* phase. During the ramp phase, the magnitude of the perturbation was increased by 5% of the distance to the vowel space center at the start of each group of four trials, up to 50%.
4. A 320-trial *hold* phase. During the hold phase, the magnitude of the perturbation was held at 50% of the distance to the vowel space center.
5. A 40-trial *washout* phase. No perturbation was given during the washout phase.
6. A 40-trial *retention* phase. A 10 minute break was given in between the washout and retention phases. Participants were allowed to read during this time, but not to talk. No perturbation was given during the retention phase.

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

To control for possible changes in vowel space area that may occur over the course of producing 500 words, each participant also completed a *control* session, which had the same structure as the adapt session but without any auditory perturbations. All participants completed both adapt and control sessions, with at least one week between sessions (mean time between sessions ± standard deviation: 8.36 ± 3.3 days). The order of the sessions was counterbalanced across participants.

After the end of the second session, participants completed a brief questionnaire that assessed their awareness of the perturbation as well as any potential strategies they used during the study. Participants were initially told there were two groups: a group that received a perturbation to the auditory feedback in both sessions and a group that did not receive a perturbation in either session (all participants received a perturbation in one session). They were asked which group they thought they were in and, if they selected the perturbed group, what they thought the perturbation was. Subsequently, participants were asked if they adopted any strategies during either session of the experiment.

Quantification and statistical analysis

Formant data were tracked using wave_viewer (Niziolek and Houde 2015), which provides a MATLAB GUI interface to formant tracking using Praat (Boersma and Weenink 2019). LPC order and pre-emphasis values were set individually for each participant. Vowels were initially automatically identified by locating the samples which were above a participant-specific amplitude level. Subsequently, all trials were hand-checked for errors. Errors in formant tracking were corrected by adjusting the pre-emphasis value or LPC order. Errors in the location of vowel onset and offset were corrected by hand-marking these times using the audio waveform and spectrogram. A small number of trials were excluded due to errors in production (i.e., the participant said the wrong word), disfluencies, or unresolvable errors in formant tracking. Across participants, 1.6% of trials were excluded (0-8%). Single values for F1 and F2 were measured for each trial as the average of these formants during the middle 50% of the vowel (steady-state portion). These formant values were converted to mels for calculating AVS and VSA. Because vowel production is inherently variable, both AVS and VSA were calculated in bins of 40 trials, using the average formants from 10 repetitions of each stimulus word. AVS and VSA were normalized by dividing these raw values by the values measured during the 60-trial baseline phase, giving a measure of the percentage change from baseline across the experiment.

Normalized measures were calculated for each participant during the last 40 trials of the hold phase (*adaptation*), during the washout phase, and during the retention phase for both adapt and control sessions. Because the control session accounts for any overall change in production during the course of 500 trials, a repeated-measures ANOVA with main effects of phase and session, as well as their interaction, was used to test for differences between adaptation, washout, and retention, as well as between adapt and control sessions. Post-hoc tests were conducted using the Tukey-Kramer method with $\alpha = 0.05$ to correct for multiple comparisons. To additionally test

for changes in absolute value from the baseline within each session, two-tailed t-tests were used to determine whether the values in each phase differed from the baseline phase (where normalized AVS and VSA were, by definition, equal to 1). Holm-Bonferroni corrections were used to maintain an overall session-wise α of 0.05.

In addition to these global measures of vowel contrast, we evaluated the changes in each of the four corner vowels independently for both the adapt and control sessions. To do this, we calculated each trial's Euclidean distance in F1/F2 space from the center of the vowel quadrilateral in the baseline phase (see Fig. 1A). Adaptation magnitude was then defined as the change in this distance-from-the-center between the baseline phase and each of the three test phases (*adaptation*, *washout*, and *retention*). This procedure was performed for both the adapt and control sessions. We evaluated adaptation magnitude through repeated-measures ANOVAs with fixed effects of vowel, phase, and session, as well as their interactions. Post-hoc tests were conducted using the Tukey-Kramer method with $\alpha = 0.05$ to correct for multiple comparisons. Two-tailed t-tests, with Holm-Bonferroni corrections, were used to determine if these values differed from the baseline value of 0.

In order to facilitate comparison with previous work (Lametti et al. 2018), we additionally measured the magnitude of the changes in vowel production that directly opposed the perturbation (*compensation*). We first calculated the difference in F1/F2 space between the average values in each test phase and the average values in the baseline phase. This difference vector, representing change from baseline, was then projected onto the inverse of the vector defining the average perturbation for that vowel, calculated from all trials in the baseline phase, to yield the measure of compensation (Figure S4A). In other words, compensation is the component of the deviation that directly opposed the formant shift. We measured compensation both in mels and as a percentage of the perturbation. This procedure was performed for both the adapt and control sessions.

We additionally evaluated the variability of vowel production. For each participant, we measured the standard deviation of each vowel in both F1 and F2 separately for each phase in both sessions. We then used these standard deviation measurements in separate repeated-measures ANOVAs to analyze variability in F1 and F2. Each model had fixed factors of vowel, session, and phase, as well as all two-way interactions between these terms. Post-hoc tests were conducted using the Tukey-Kramer method with $\alpha = 0.05$ to correct for multiple comparisons.

In order to test for any differences in baseline vowel contrast between the two sessions, we examined raw (non-normalized) AVS and VSA values. Repeated-measures ANOVAS with fixed factors of session and session order (adapt first vs. control first) were used to assess any potential differences.

Lastly, we evaluated any potential changes in a broad range of speech parameters that are associated with hyperarticulation and clear speech. In order to adapt to the vowel-centralization perturbation, participants must produce more extreme versions of vowels that lie farther away from the center of their vowel space. A similar increase in vowel space occurs when speakers pronounce words more clearly than they are normally pronounced, often referred to as

hyperarticulation (Baker and Bradlow 2009; Lindblom 1990). Importantly, all circumstances known to induce hyperarticulation also result in increases in vowel duration and often affect other parameters of speech as well. We measured vowel duration (in ms), maximum intensity (as measured from the root mean square signal from Audapter in arbitrary units), maximum vocal pitch (in Hz), and pitch range (in Hz). All of these measures were normalized by subtracting the average value in the baseline from the remaining trials. Repeated-measures ANOVAs were used to evaluate statistical significance.

Data and Code Availability

Analysis code is available on GitHub at https://github.com/blab-lab/vsaCentralize. Some functions rely on additional code available at https://github.com/carrien/free-speech.

Results

Speakers adapt to vowel centralization by increasing global contrast

Speakers responded to the centralization perturbation by expanding VSA in the adapt session relative to the control session (Figures 2, 3A, 6A), indicated by a main effect of session ($F(1,48) = 5.32$, $p = 0.03$). In the adapt session, VSA remained high until the end of the hold phase (*adaptation* phase), dropping closer to baseline values in the washout and retention phases. In the control session, VSA fell slightly below baseline in the adaptation, washout, and retention phases. These changes are reflected in significant effects of phase ($F(2,48) = 3.81$, $p = 0.03$) and the interaction between session and phase ($F(2,48) = 4.44$, $p = 0.02$). During the adaptation phase, VSA was significantly greater in the adapt session (9.7% ± 4.2%) than in the control session (-4.3% ± 3%, $p < 0.05$), though neither session was significantly different from baseline after correction for multiple comparisons. Neither the washout nor retention phases differed between sessions, nor were any of these values different from the baseline (all $p > 0.07$).

--- Figure 2 ---

--- Figure 3 ---

AVS showed similar results (Figures 3B, 6B). In all phases after the baseline, AVS was larger in the adapt session than the control session ($p < 0.05$, main effect of session: $F(1,48) = 11.02$, $p = 0.003$). The difference between sessions was greatest in the adaptation phase, where participants increased the spacing between vowels, relative to baseline, by an average of 6.3 ± 1.7% in the adapt session and decreased it by -1.4 ± 1.4% in the control session. A significant difference was maintained throughout both the washout (3.2 ± 1.5% vs. -1.6 ± 1.2%) and retention (1.5% ± 1.5% vs. -1.5 ± 1.2%) phases. The change in AVS across phases was shown by a main effect of phase ($F(2,48) = 7.25$, $p = 0.002$) as well as an interaction between phase and session ($F(2,48) = 5.33$, $p = 0.009$). Only the adaptation phase in the adapt session was significantly different from baseline ($t(24) = 3.73$, $p = 0.001$) after correction for multiple

comparisons. The only significant within-session difference was between the adaptation and retention phases in the adapt session ($p < 0.05$).

VSA and AVS values were highly correlated ($r = 0.95$, $p < 0.0001$, Figure S1). Baseline values did not differ between the adapt and control sessions (VSA: $F(1,46) = 0.01$, $p = 0.92$, AVS: $F(1,46) = 0.01$, $p = 0.91$). There was no change from session 1 to session 2 in either metric (VSA: $F(1,46) = 0.15$, $p = 0.70$, AVS: $F(1,46) = 0.02$, $p = 0.89$), and no interaction between session type and order (VSA: $F(1,46) = 0.6$, $p = 0.44$, AVS: $F(1,46) = 0.4$, $p = 0.55$).

Speech contrast increases as duration decreases

In contrast to VSA and AVS, vowel duration decreased slightly over the course of the experiment (Figure 4). In the adapt session, vowels in the adaptation phase were 17 ± 33 ms shorter than the baseline ($p = 0.02$). Duration continued to decrease in the washout (23 ± 37 ms, $p = 0.006$) and retention phases (25 ± 38 ms, $p = 0.004$). Decreases in duration from baseline were smaller in the control session, with no phase significantly shorter than baseline (adaptation: 5 ± 40 ms, $p = 0.53$; washout: 10 ± 34 ms, $p = 0.17$; retention: 11 ± 31 ms, $p = 0.09$). The difference between the sessions did not reach significance ($F(1, 48) = 3.1$, $p = 0.08$). There was no significant effect of phase ($F(2, 48) = 2.2$, $p = 0.12$), nor any interaction between phase and session ($F(2, 48) = 0.9$, $p = 0.92$). We similarly observed only minimal changes in other speech parameters that did not differ across sessions (Table S1; Fig. S2): maximum pitch and peak intensity slightly increased, and pitch range did not change.

--- Figure 4 ---

Speakers simultaneously learn multiple vowel-specific compensatory changes

Speakers in the adapt session achieved these increases in speech contrast by increasing the distance between each vowel and the center of the vowel space (Figures 5, 6), reflected by a main effect of session ($F(1,414) = 9.49$, $p < 0.0001$). The increase in distance to the center was greatest in the adaptation phase (20.9 ± 3.9 mels), and smaller in the washout (10.7 ± 3.4 mels) and retention (4.6 ± 3.4 mels) phases. Adaptation, washout, and retention phases in the adapt session all differed from the control session ($p < 0.05$), where the distances decreased from baseline (adaptation: -2.9 ± 2.9 mels, washout: -3.1 ± 3.3 mels, retention: -4.5 ± 2.8 mels). These changes were reflected in a main effect of phase ($F(2,414) = 8.41$, $p = 0.007$) and an interaction between phase and session ($F(2,414) = 5.47$, $p = 0.005$). There were no significant differences between vowels ($F(3,414) = 2.0$, $p = 0.12$), though there was a significant interaction between vowel and session ($F(3,414) = 3.81$, $p = 0.01$). Post-hoc tests comparing individual vowels between the sessions showed that /i/ was farther from the center in both the adapt and washout phases, and that /ɑ/ was farther from the center in all three test phases (all $p < 0.05$). No other comparison was significant after correction for multiple comparisons. Results were highly similar when examining the portion of compensatory changes that directly opposed the perturbation (Figure S3).

--- Figure 5 ---

Individual variability in adaptation

Although the vowel centralization paradigm resulted in increased speech contrast at the group level, there was substantial variability in response magnitude across participants, both in global measures of vowel spacing (Fig. 6A,B) and in the compensatory movement of individual vowels away from the vowel center (Fig. 6D). Similar inter-individual variability is consistently seen in studies of sensorimotor adaptation in speech (Martin et al. 2018; Munhall et al. 2009; Parrell et al. 2017; Villacorta et al. 2007). We examined whether individual adaptation magnitude was predicted by vowel spacing in the baseline phase; we found no such correlation for either global measure of vowel spacing (VSA: $r = 0.08$, $p = 0.70$; AVS: $r = 0.26$, $p = 0.21$, Fig. 6C), suggesting that the variability seen across participants in their response to the centralization perturbation was not driven by differences in baseline production. Notably, similar increases in VSA/AVS across participants were driven by different patterns of adaptation at the individual vowel level (Fig. 2). Adaptation magnitude for a given vowel was not well-predicted by that vowel's baseline formant variability (/i/: $r = -0.21$, $p = 0.32$; /æ/: $r = -0.20$, $p = 0.34$; /ɑ/: $r = 0.37$, $p = 0.07$; /u/: $r = -0.09$, $p = 0.66$). Adaptation was also not well-predicted by the initial distance to the center of the vowel space across vowels (/i/: $r = 0.32$, $p = 0.12$; /æ/: $r = 0.21$, $p = 0.31$; /ɑ/: $r = 0.42$, $p = 0.04$; /u/: $r = 0.065$, $p = 0.76$; no vowel significant after correcting for multiple comparisons).

--- Figure 6 ---

Awareness of perturbation and strategy use

When participants were queried about their awareness of the perturbation, 16/25 responded that they thought they received a perturbation, but no participant correctly identified it as a change to their vowels (Table 1). Only 3/25 participants reported using a strategy, and no strategy addressed the applied perturbation. The reported strategies were: "Saying the words slower", "Kept mouth open between words", and "Looking away from the screen between words".

Number of participants	Perceived perturbation
9	Did not perceive a perturbation
6	Thought audio feedback had added noise (likely reflecting the 60 dB speech-shaped noise added to the signal)
5	Perceived a perturbation but unable to identify what it was
2	Pitch of voice altered
1	Speech delayed
1	Speech volume altered
1	Speech was "more nasal"

Table 1: Participant awareness of perturbation.

Discussion

In the current study, we used alterations of auditory feedback to drive participants to expand their working vowel space and increase the contrast between vowels. Speakers who were exposed to vowel centralization feedback learned to produce corner vowels farther from the center of their vowel space, partially overcoming the perceived centralization. This was reflected in global measures of vowel space (VSA) and vowel contrast (AVS), as well as vowel-specific measures. These changes were partially retained after the perturbation was removed, as well as in an assay of retention 10 minutes after the main experiment.

Overall, our findings show that speakers are capable of adapting to non-uniform transformations of vowel space feedback. Because the direction of the feedback shift was dependent on the produced formants, participants in the study had to learn vowel-dependent compensatory adjustments, each of which required unique changes to articulatory movements. The current results build on previous work (Rochet-Capellan and Ostry 2011) to show that multiple opposing transformations can be learned simultaneously across the extent of producible vowel space, establishing the ability of sensorimotor adaptation paradigms to enhance global contrast between vowels.

Increased vowel contrast is a commonly used metric for quantifying the effects of rehabilitative interventions for motor speech disorders (Lam and Tjaden 2016; Sapir et al. 2007; Tjaden et al. 2013; Tjaden and Wilding 2004; Whitfield and Goberman 2014), as it is associated with greater intelligibility (Ansel and Kent 1992; Bang et al. 2013; Kim et al. 2011; Neel 2008; Skodda et al. 2011; Tjaden et al. 2005; Weismer et al. 2001). In the current study, ~12 minutes of speaking resulted in a VSA increase of 15.8% (calculated using /u/, /i/, and /ɑ/), half the increase seen in individuals with Parkinson's disease after sixteen hour-long sessions of LSVT-LOUD (31.6% calculated with the same three vowels) (Sapir et al. 2007), the current standard of care for the hypokinetic dysarthria secondary to that disorder. Furthermore, participants adapted their

speech without conscious effort or awareness of the specific targets of the perturbation, which may be useful for patients who do not, or cannot, respond well to existing treatments due to issues with explicit strategy use (Sadagopan and Huber 2007).

The increase in vowel formant contrast in the current study is similar to the hyperarticulation observed when people are explicitly instructed to speak clearly (Ferguson and Kewley-Port 2002; Krause and Braida 2004; Lam and Tjaden 2016; Moon and Lindblom 1994; Picheny et al. 1986; Smiljanić and Bradlow 2008, 2009). Vowel hyperarticulation also occurs in many contexts that implicitly encourage clearer speech: when repeating a word after being misunderstood (Burnham et al. 2010a, 2010b; Oviatt et al. 1998a, 1998b), when speaking to infants (Kuhl et al. 1997; Lam and Kitamura 2012), when speaking to people who speak a foreign language (Scarborough et al. 2007), when speaking to individuals who are have hearing difficulties (Picheny et al. 1986; Scarborough and Zellou 2013), when speaking in "challenging" acoustic situations such as when a conversation partner is wearing headphones (Hazan and Baker 2011; Koster 2001), when using words for the first time in a discourse compared to subsequent repetitions of that word (Baker and Bradlow 2009), when words are less predictable from sentential context versus more predictable—e.g., "The next word is nine" vs. "A stitch in time saves nine" (Aylett and Turk 2006; Scarborough 2010), for words that are relatively uncommon vs. words that occur more frequently (Baker and Bradlow 2009; Scarborough 2010, 2013), for words that have a lower number of similar words vs. words with a higher number of similar words—i.e., hyperarticulation is associated with higher lexical density (Scarborough and Zellou 2013), and in words with prosodic stress or emphasis (Cho et al. 2011; de Jong 1995; de Jong et al. 1993). In every case, changes in formants are universally accompanied by relative increases in vowel duration, which allow more time for full articulatory movement (Lindblom 1990). Hyperarticulation is so closely tied to increased duration, in fact, that duration is often used as a proxy metric for hyperarticulation (Aylett and Turk 2004; Baker and Bradlow 2009; Freeman 2014). In contrast, we observed no differences between the adapt and control sessions in vowel duration, intensity, or pitch, and in fact saw an overall decrease in vowel duration in the adapt session relative to baseline. These results strongly suggest the increase in vowel contrast observed in the current study does not arise from general mechanisms of clarity-driven hyperarticulation but is, in fact, an adaptive response to counteract the auditory perturbation. Future work could examine adaptation to an outward-pushing perturbation that enhances vowel contrast (and which would require hypoarticulation as a compensatory mechanism), or to perturbations with unrelated vowel-specific effects, to address this issue directly.

It remains unclear from the current data whether participants learned a global vowel expansion pattern or local, vowel-specific changes. The variable learning by vowel, with low learning for /u/ in particular, suggests participants may have learned separate transformations. These results must be interpreted cautiously, however, as the capacity for adaptation may vary between vowels (Mitsuya et al. 2015). Additionally, F2 variability for /u/ in our data (36.2 mels) was substantially higher than other vowels (20.4 mels, all $p < 0.05$, see Table S2), consistent with previous reports (Clopper et al. 2019). Thus, perturbations for /u/ may have caused fewer

productions to fall outside acceptable category boundaries (Mitsuya et al. 2013; Niziolek and Guenther 2013). Future work examining generalization to other, untrained, vowels may help resolve whether the learning observed here is achieved through a combination of local transformations or a generalized, global pattern (Malfait et al. 2005; Rochet-Capellan et al. 2012).

Increases in vowel contrast persisted even after a washout period and 10-minute silent interval, suggesting sensorimotor adaptation may cause longer-term changes, consistent with previous results in whispered speech (Houde and Jordan 2002). Although some evidence of retention after a short break can be found in a figure in a previous study (Lametti et al. 2014), no statistical evidence for retention of sensorimotor learning in voiced speech has been reported previously. More research on retention of learning in sensorimotor adaptation will be vital to clinical translation of sensorimotor adaptation, including how retention is affected by different communicative settings.

In the control session, there was a trend for both VSA and AVS to decrease over the course of the experiment. This gradual decrease explains the pattern of results: consistent contrast between sessions, even when the individual phases did not always differ from the baseline. This can also be seen in the individual vowel data (Figures 5, S3). This pattern suggests that vowels have a tendency to become more centralized over the course of an hour-long study, highlighting the importance of conducting a control session to account for potential changes in speech unrelated to the auditory perturbation.

Similar to previous studies of sensorimotor adaptation in speech, we observed substantial variability across individuals in the magnitude of adaptation. This variability was seen in both global measures of vowel spacing as well as in changes in productions of individual vowels. Variability in adaptation was not correlated with speech behavior in the baseline phase. Previous work has suggested that variability in the magnitude of sensorimotor adaptation for perturbations of a single vowel may be caused by differences in the balance between auditory and somatosensory feedback use across individuals (Katseff et al. 2012; Lametti et al. 2012; Parrell et al. 2019). It is possible the variability we observed has a similar source.

Finally, we found a strong relationship between VSA and AVS. Previous studies in motor speech disorders have suggested that vowel contrast may be a more accurate predictor of intelligibility impairments in these populations (Neel 2008). The present results suggest that both measures capture similar effects, though AVS may be a slightly more sensitive measure. However, more research with a wider set of stimuli is needed to determine how this relationship is affected when non-corner vowels are included in the AVS measures, and it remains to be seen how intelligibility may be correlated with either measure.

In sum, we have shown that individuals can modify their speech to oppose auditory perturbations that target vowel distinctiveness. This demonstration of the ability to drive increased vowel contrast without the need for explicit strategies or conscious control moves research on sensorimotor adaptation closer to realizing the long-standing promise of this technique for clinical use.

Grants

This work was supported by NIH grants R01 DC017091 and R00 DC014520.

Competing Interests

The authors declare no competing interests.

References

Ansel BM, **Kent RD**. Acoustic-phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy. *J Speech Hear Res* 35: 296, 1992.

Aylett M, **Turk A**. The Smooth Signal Redundancy Hypothesis: A Functional Explanation for Relationships between Redundancy, Prosodic Prominence, and Duration in Spontaneous Speech. *Lang Speech* 47: 31–56, 2004.

Aylett M, **Turk A**. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *J Acoust Soc Am* 119: 3048–3058, 2006.

Baker RE, **Bradlow AR**. Variability in Word Duration as a Function of Probability, Speech Style, and Prosody. *Lang Speech* 52: 391–413, 2009.

Bang Y-I, **Min K**, **Sohn YH**, **Cho S-R**. Acoustic characteristics of vowel sounds in patients with Parkinson disease. *NeuroRehabilitation* 32: 649–654, 2013.

Boersma P, **Weenink D**. Praat: doing phonetics by computer. [Online]. http://www.praat.org/.

Burnham D, **Joeffry S**, **Rice L**. "D-o-e-s-Not-C-o-m-p-u-t-e": Vowel Hyperarticulation in Speech to an Auditory-Visual Avatar. In: *Proceedings of the 9th International Conference on Auditory-Visual Speech Processing (AVSP-2010)*. Japan: 2010a.

Burnham D, **Joeffry S**, **Rice L**. Computer- and human-directed speech before and after correction. In: *Speech Science and Technology*. Melbourne, Australia: 2010b, p. 13–17.

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 /iau/. In: *Proceedings of the 8th International Seminar on Speech Production*. Strasbourg, France: 2008, p. 65–68.

Cho T, **Lee Y**, **Kim S**. Communicatively driven versus prosodically driven hyper-articulation in Korean. *J Phon* 39: 344–361, 2011.

Clopper CG, **Burdin RS**, **Turnbull R**. Variation in /u/ fronting in the American Midwest. *J Acoust Soc Am* 146: 233–244, 2019.

Ferguson SH, **Kewley-Port D**. Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *J Acoust Soc Am* 112: 259–271, 2002.

Freeman V. Hyperarticulation as a signal of stance. *J Phon* 45: 1–11, 2014.

Hazan V, **Baker R**. Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *J Acoust Soc Am* 130: 2139–2152, 2011.

Houde JF, **Jordan MI**. Sensorimotor Adaptation in Speech Production. *Science* 279: 1213–1216, 1998.

Houde JF, **Jordan MI**. Sensorimotor adaptation of speech I: Compensation and adaptation. *J Speech Lang Hear Res* 45: 295–310, 2002.

de Jong K. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *J Acoust Soc Am* 97: 491–504, 1995.

de Jong K, **Beckman M**, **Edwards J**. The interplay between prosodic structure and coarticulation. *Lang Speech* 36: 197–212, 1993.
Katseff S, **Houde JF**, **Johnson K**. Partial Compensation for Altered Auditory Feedback: A Tradeoff with Somatosensory Feedback? *Lang Speech* 55: 295–308, 2012.
Kim H, **Hasegawa-Johnson M**, **Perlman A**. Vowel Contrast and Speech Intelligibility in Dysarthria. *Folia Phoniatr Logop* 63: 187–194, 2011.
Koster S. Acoustic-phonetic characteristics of hyperarticulated speech for different speaking styles. In: *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings*. 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. IEEE, p. 873–876.
Krause JC, **Braida LD**. Acoustic properties of naturally produced clear speech at normal speaking rates. *J Acoust Soc Am* 115: 362–378, 2004.
Kuhl P, **Andruski EJ**, **Chistovich AI**, **Chistovich AL**, **Kozhevnikov VE**, **Ryskina VV**, **Stolyarova IE**, **Sundberg U**, **Lacerda F**. Cross-language analysis of phonetics units in language addressed to infants. *Science* 277: 684–686, 1997.
Lam C, **Kitamura C**. Mommy, speak clearly: induced hearing loss shapes vowel hyperarticulation: Hearing loss and vowel hyperarticulation. *Dev Sci* 15: 212–221, 2012.
Lam J, **Tjaden K**. Clear Speech Variants: An Acoustic Study in Parkinson's Disease. *J Speech Lang Hear Res* 59: 631–646, 2016.
Lametti DR, **Nasir SM**, **Ostry DJ**. Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *J Neurosci* 32: 9351–8, 2012.
Lametti DR, **Rochet-Capellan A**, **Neufeld E**, **Shiller DM**, **Ostry DJ**. Plasticity in the human speech motor system drives changes in speech perception. *J Neurosci* 34: 10339–46, 2014.
Lametti DR, **Smith HJ**, **Watkins KE**, **Shiller DM**. Robust Sensorimotor Learning during Variable Sentence-Level Speech. *Curr Biol* 28: 3106-3113.e2, 2018.
Lindblom B. Explaining phonetic variation: a sketch of the H&H theory. In: *Speech Production and Modelling*, edited by Hardcastle JW, Marchal A. Dordrecht: Kluwer Academic Publisher, 1990, p. 403–439.
Malfait N, **Gribble PL**, **Ostry DJ**. Generalization of Motor Learning Based on Multiple Field Exposures and Local Adaptation. *J Neurophysiol* 93: 3327–3338, 2005.
Martin CD, **Niziolek CA**, **Duñabeitia JA**, **Perez A**, **Hernandez D**, **Carreiras M**, **Houde JF**. Online Adaptation to Altered Auditory Feedback Is Predicted by Auditory Acuity and Not by Domain-General Executive Control Resources. *Front Hum Neurosci* 12, 2018.
Mitsuya T, **MacDonald EN**, **Munhall KG**, **Purcell DW**. Formant compensation for auditory feedback with English vowels. *J Acoust Soc Am* 138: 413–424, 2015.
Mitsuya T, **Samson F**, **Ménard L**, **Munhall KG**. Language dependent vowel representation in speech production. *J Acoust Soc Am* 133: 2993–3003, 2013.
Moon S-J-. J, **Lindblom B**. Interaction between duration, context, and speaking style in English stressed vowels. *J Acoust Soc Am* 96: 40–55, 1994.
Munhall GK, **MacDonald EN**, **Byrne SK**, **Johnsrude I**. Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *J Acoust Soc Am* 125: 384–390, 2009.
Neel AT. Vowel Space Characteristics and Vowel Identification Accuracy. *J Speech Lang Hear Res* 51: 574–585, 2008.
Niziolek CA, **Guenther FH**. Vowel Category Boundaries Enhance Cortical and Behavioral Responses to Speech Feedback Alterations. *J Neurosci* 33: 12090–12098, 2013.

Niziolek CA, **Houde J**. Wave_Viewer: First Release. 2015.
Oviatt S, **Levow G-A**, **Moreton E**, **MacEachern M**. Modeling global and focal hyperarticulation during human–computer error resolution. *J Acoust Soc Am* 104: 3080–3098, 1998a.
Oviatt S, **MacEachern M**, **Levow G-A**. Predicting hyperarticulate speech during human-computer error resolution. *Speech Commun* 24: 87–110, 1998b.
Parrell B, **Agnew Z**, **Nagarajan S**, **Houde JF**, **Ivry RB**. Impaired Feedforward Control and Enhanced Feedback Control of Speech in Patients with Cerebellar Degeneration. *J Neurosci* 37: 9249–9258, 2017.
Parrell B, **Ramanarayanan V**, **Nagarajan S**, **Houde J**. The FACTS model of speech motor control: Fusing state estimation and task-based control. *PLOS Comput Biol* 15: e1007321, 2019.
Picheny MA, **Durlach NI**, **Braida LD**. Speaking Clearly for the Hard of Hearing II: Acoustic Characteristics of Clear and Conversational Speech. *J Speech Lang Hear Res* 29: 434–446, 1986.
Purcell DW, **Munhall KG**. Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *J Acoust Soc Am* 120: 966–77, 2006.
Rochet-Capellan A, **Ostry DJ**. Simultaneous acquisition of multiple auditory-motor transformations in speech. *J Neurosci* 31: 2657–62, 2011.
Rochet-Capellan A, **Richer L**, **Ostry DJ**. Nonhomogeneous transfer reveals specificity in speech motor learning. *J Neurophysiol* 107: 1711–7, 2012.
Roemmich RT, **Bastian AJ**. Closing the Loop: From Motor Neuroscience to Neurorehabilitation. *Annu Rev Neurosci* 41: 415–429, 2018.
Sadagopan N, **Huber JE**. Effects of loudness cues on respiration in individuals with Parkinson's disease. *Mov Disord* 22: 651–659, 2007.
Sapir S, **Spielman JL**, **Ramig LO**, **Story BH**, **Fox C**. Effects of Intensive Voice Treatment (the Lee Silverman Voice Treatment [LSVT]) on Vowel Articulation in Dysarthric Individuals With Idiopathic Parkinson Disease: Acoustic and Perceptual Findings. *J Speech Lang Hear Res* 50: 899–912, 2007.
Scarborough R. Lexical and contextual predictability: Confluent effects on the production of vowels. In: *Laboratory Phonology 10*, edited by Fougeron C, Kuehnert B, D'Imperio M, Vallee N. De Gruyter Mouton, 2010.
Scarborough R. Neighborhood-conditioned patterns in phonetic detail: Relating coarticulation and hyperarticulation. *J Phon* 41: 491–508, 2013.
Scarborough R, **Brenier J**, **Zhao Y**, **Hall-Lew L**, **Dmitrieva O**. An acoustic study of real and imagined foreigner-directed speech. In: *Proceedings of the 15th International Congress of Phonetic Sciences*. 2007, p. 2165–2168.
Scarborough R, **Zellou G**. Clarity in communication: "Clear" speech authenticity and lexical neighborhood density effects in speech production and perception. *J Acoust Soc Am* 134: 3793–3807, 2013.
Skodda S, **Visser W**, **Schlegel U**. Vowel Articulation in Parkinson's Disease. *J Voice* 25: 467–472, 2011.
Smiljanić R, **Bradlow AR**. Stability of Temporal Contrasts across Speaking Styles in English and Croatian. *J Phon* 36: 91–113, 2008.
Smiljanić R, **Bradlow AR**. Speaking and Hearing Clearly: Talker and Listener Factors in Speaking Style Changes. *Lang Linguist Compass* 3: 236–264, 2009.
Tjaden K, **Lam J**, **Wilding G**. Vowel Acoustics in Parkinson's Disease and Multiple Sclerosis: Comparison of Clear, Loud, and Slow Speaking Conditions. *J Speech Lang Hear Res* 56: 1485–

1502, 2013.
Tjaden K, **Rivera D**, **Wilding G**, **Turner GS**. Characteristics of the Lax Vowel Space in Dysarthria. *J Speech Lang Hear Res* 48: 554–566, 2005.
Tjaden K, **Wilding GE**. Rate and Loudness Manipulations in Dysarthria: Acoustic and Perceptual Findings. *J Speech Lang Hear Res* 47: 766–783, 2004.
Tourville JA, **Cai S**, **Guenther F**. Exploring auditory-motor interactions in normal and disordered speech. In: *Proceedings of Meetings on Acoustics*, p. 060180.
Villacorta VM, **Perkell JS**, **Guenther FH**. Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *J Acoust Soc Am* 122: 2306–19, 2007.
Weismer G, **Jeng J-Y**, **Laures JS**, **Kent RD**, **Kent JF**. Acoustic and Intelligibility Characteristics of Sentence Production in Neurogenic Speech Disorders. *Folia Phoniatr Logop* 53: 1–18, 2001.
Whitfield JA, **Goberman AM**. Articulatory–acoustic vowel space: Application to clear speech in individuals with Parkinson's disease. *J Commun Disord* 51: 19–28, 2014.

Figure Captions

Figure 1. Experiment design. A: Illustration of the perturbation field applied to speech. All perturbations point towards the center of the speaker's vowel space, effectively centralizing their vowels. B: Example spectrograms of the four target words showing the produced formants (blue), perturbed formants during the hold phase (red), and the vowel space center (yellow). C: Magnitude of the perturbation applied throughout the experiment. In the adapt session (red), the perturbation during the hold phase is 50% of the 2D distance (in F1/F2 space) between the current formant values and the vowel center. In a separate control session (blue), no perturbation is applied.

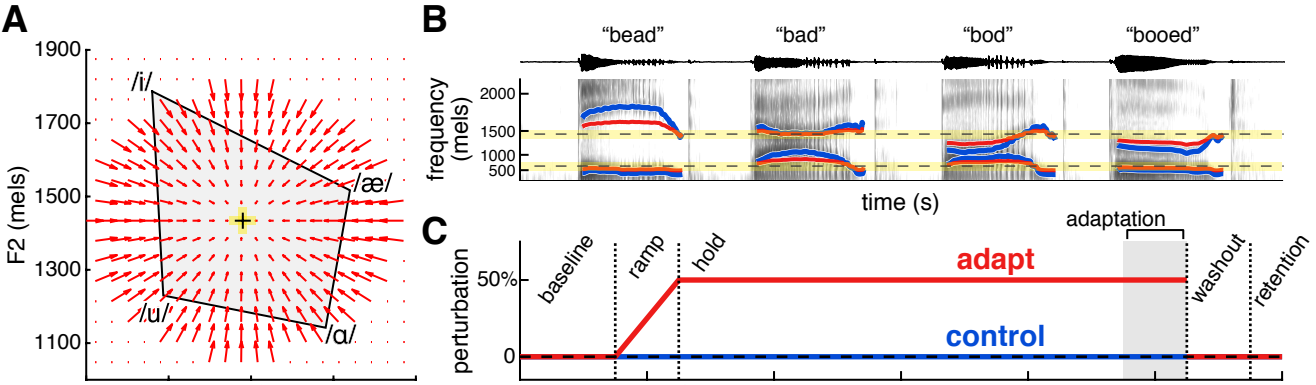
Figure 2: Illustration of vowel space increases. Data from three example participants showing the movement of the four corner vowels after exposure to perturbed feedback in the adapt session (red) or unperturbed feedback in the control session (blue) compared with their values in the baseline phase of each session (dashed black).

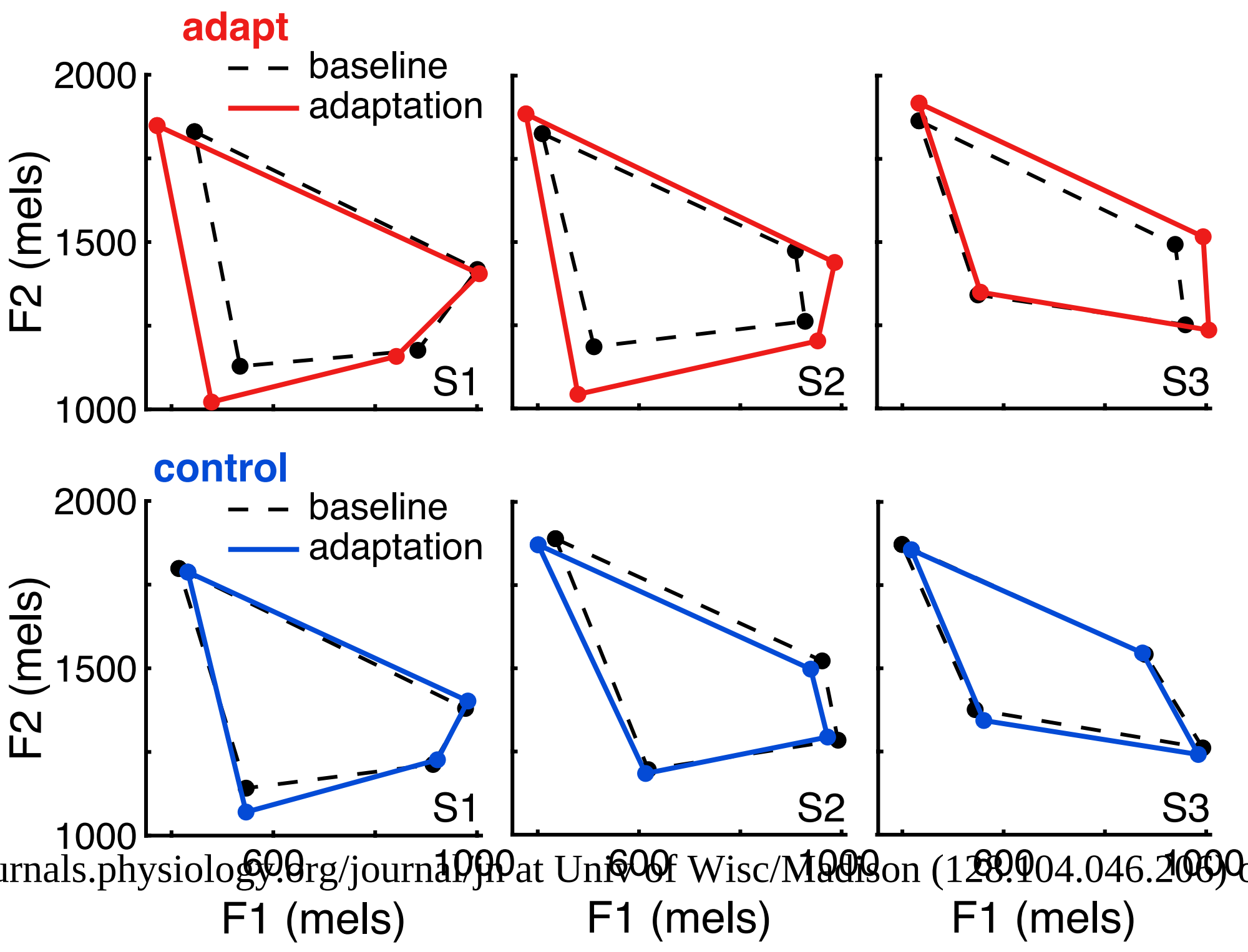
Figure 3: Vowel contrast adaptation. A: Baseline-normalized vowel space area (VSA) increases in the adapt session (red) but not the control session (blue). C: Individual and group means for the adaptation, washout, and retention phases. Each pair of points connected by a gray line represents data from a single participant. B, D: Same as (A) and (C), showing average vowel spacing (AVS). Error bars show standard error. See also Figure 6A,B.

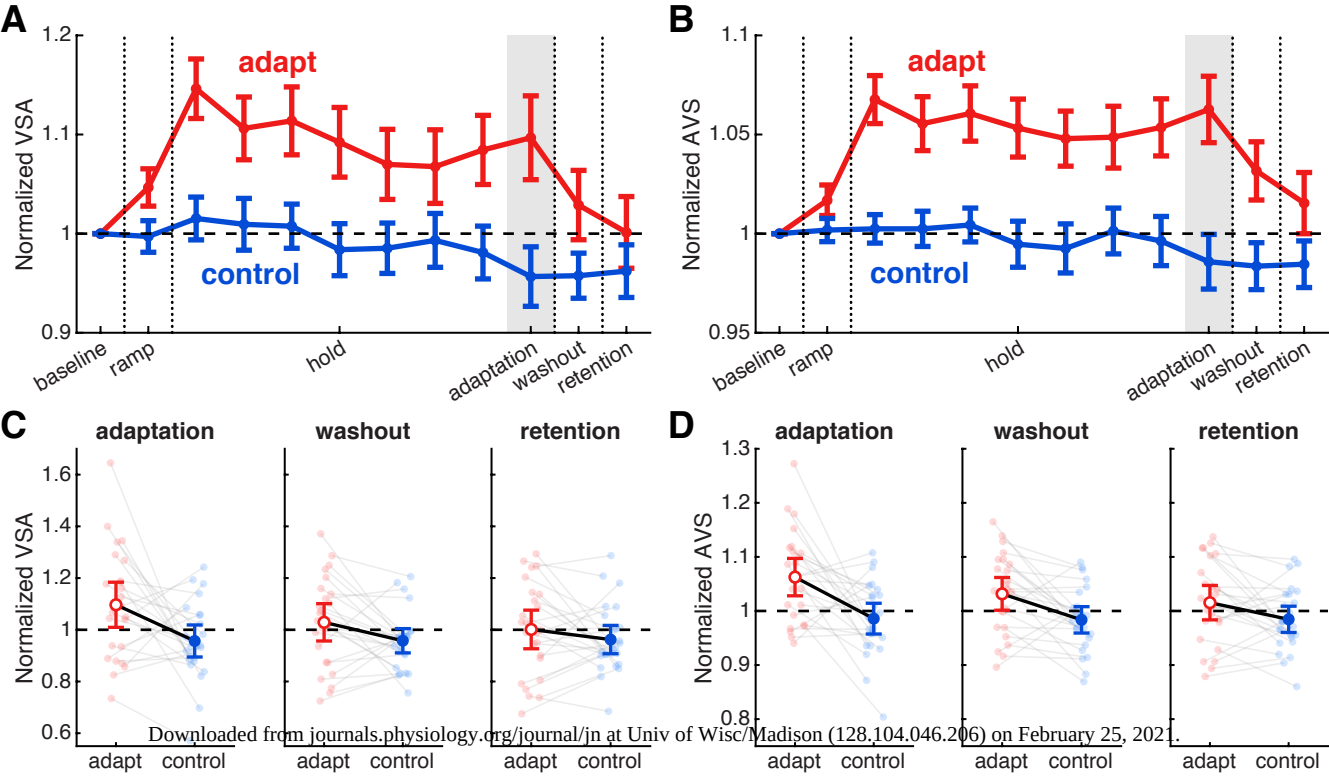
Figure 4: Changes in vowel duration. A: Baseline-normalized changes in vowel duration in the adapt session (red) and the control session (blue). B: Individual and group means for normalized vowel duration in the adaptation, washout, and retention phases. Each pair of points connected by a gray line represents data from a single participant. See also Figure S2.

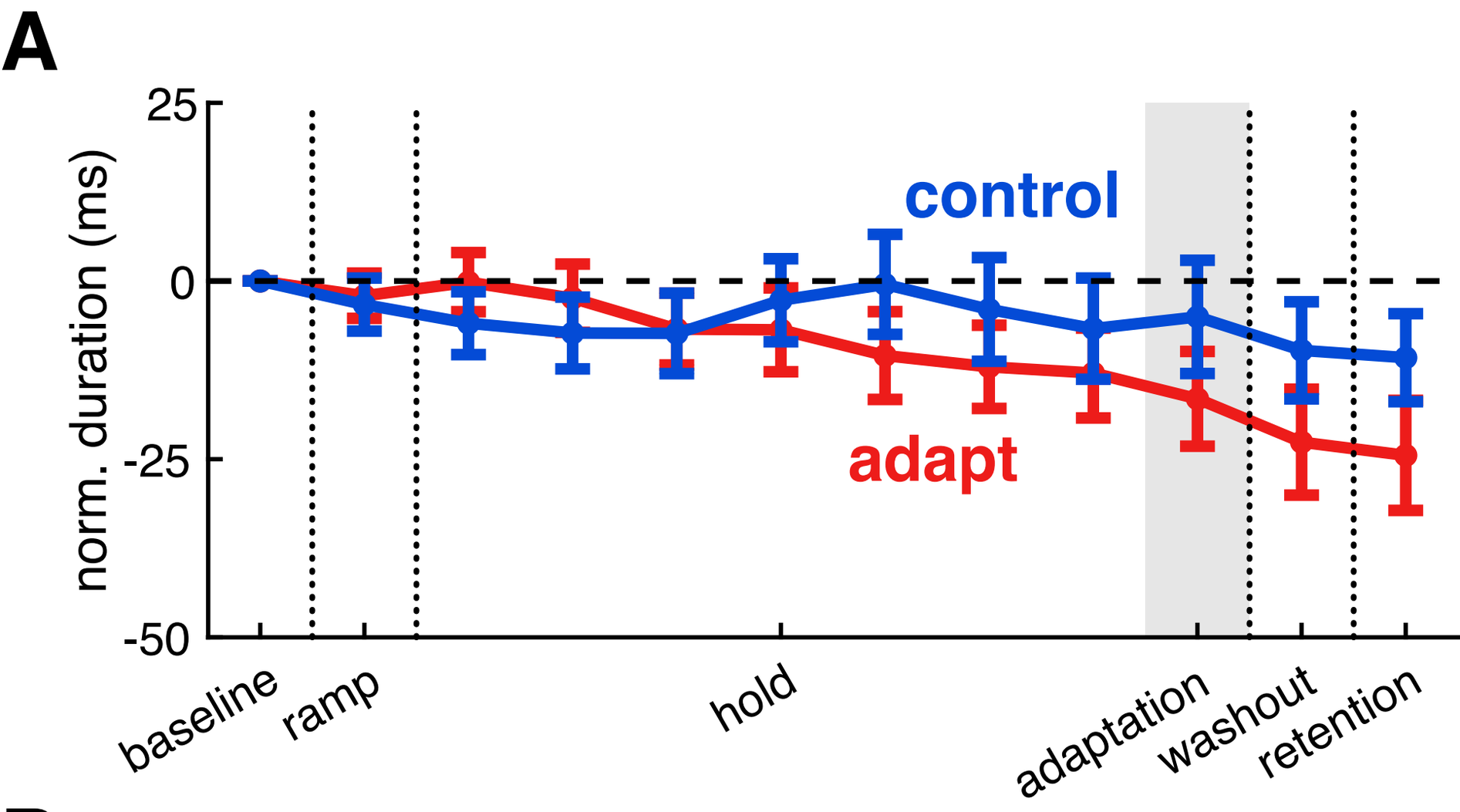
Figure 5: Vowel-specific compensatory changes. A-C: Mean change (± standard error) in formants for each vowel in the adaptation, washout, and retention phases relative to baseline values (normalized to (0,0)). Bright colors (open circles) show data from the adapt session; dull colors (filled circles) show data from the control session. D-F: Group means (± standard error) and individual participant adaptation for each vowel. Colors as in A-C. See also Figures 6, S2.

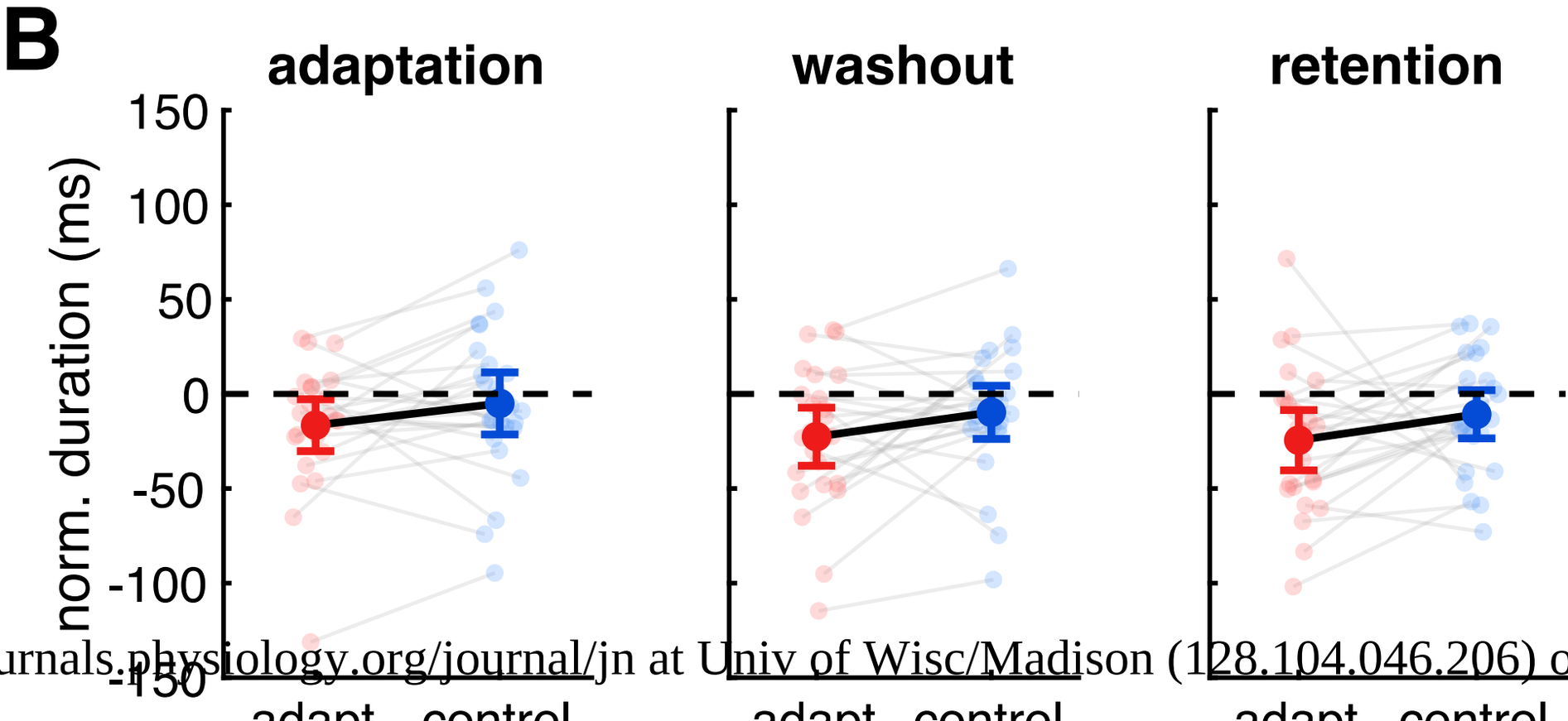
Figure 6: Individual variability in vowel space adaptation. A-B: Individual participant differences between adapt and control sessions in normalized VSA (A) and normalized AVS (B), as measured in the adaptation phase. Each bar represents a participant, ordered in both panels by descending (positive) difference in VSA between the two sessions. **C**: Baseline VSA and AVS values do not predict adaptation magnitude. **D**: As in A, individual participant differences between the adapt and control sessions in the normalized distance to center` for each vowel, as measured in the adaptation phase.

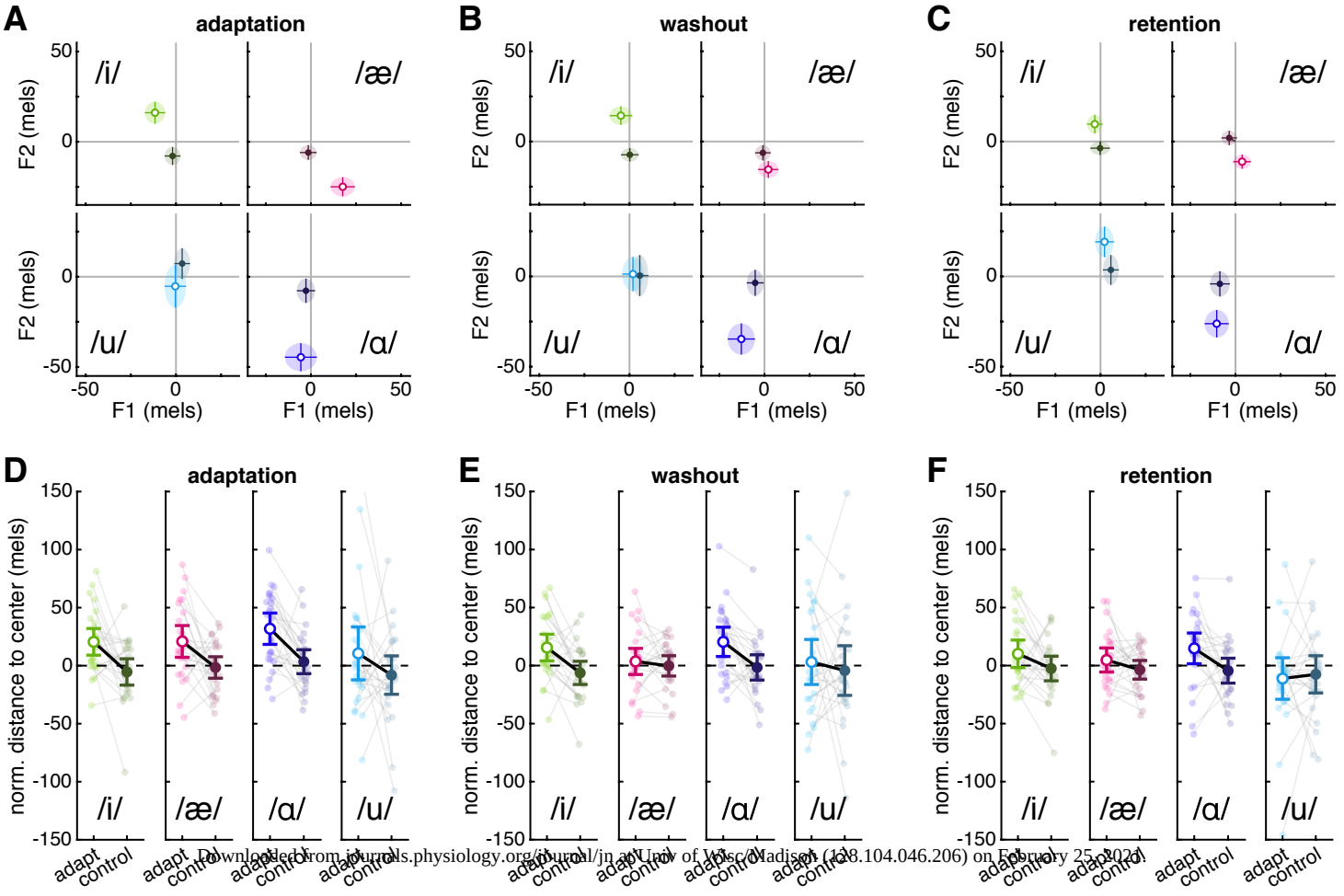


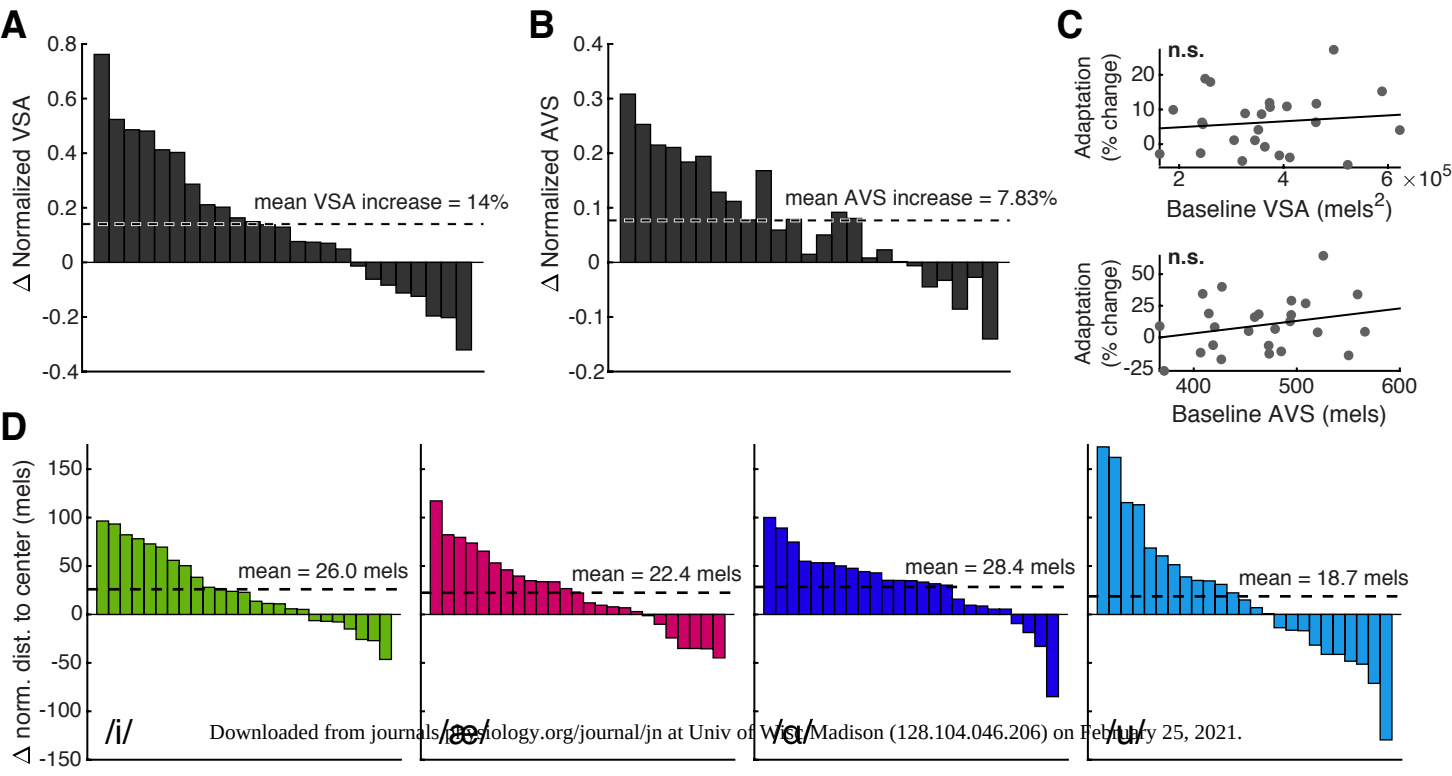












Number of participants	Perceived perturbation
9	Did not perceive a perturbation
6	Thought audio feedback had added noise (likely reflecting the 60 dB speech-shaped noise added to the signal)
5	Perceived a perturbation but unable to identify what it was
2	Pitch of voice altered
1	Speech delayed
1	Speech volume altered
1	Speech was “more nasal”

Table 1: Participant awareness of perturbation.