

Attention plays a highly limited role in speech sensorimotor control

Jenna Krakauer¹, Chris Naber², Caroline Niziolek^{1,2*}, Benjamin Parrell^{1,2*}

¹ Dept. of Communication Sciences and Disorders, University of Wisconsin-Madison

² Waisman Center, University of Wisconsin-Madison

**equal contribution*

Abstract

Purpose: When vowel formants are externally perturbed, speakers change their production to oppose that perturbation both during the ongoing production (compensation) and in future productions (adaptation).. To date, attempts to explain the large variability across individuals in these responses have focused on trait-based characteristics such as auditory acuity, but evidence from other motor domains suggests that attention may modulate the motor response to sensory perturbations. Here, we test the extent to which attention impacts sensorimotor control for supralaryngeal articulation.

Method: Neurobiologically healthy speakers were exposed to random (Experiment 1) or consistent (Experiment 2) real-time auditory perturbation of vowel formants to measure online compensation and trial-to-trial adaptation, respectively. In both experiments, participants completed two conditions: one with a simultaneous visual distractor task to divide attention, and one without this secondary task.

Results: Divided attention slightly reduced online compensation, but only starting >300 ms after vowel onset, well beyond the typical duration of vowels in speech. Divided attention had no effect on adaptation.

Conclusions: The results from both experiments suggest that the use of sensory feedback in typical speech motor control is a largely automatic process unaffected by attention, suggesting that the source of cross-speakers variability in response to formant perturbations likely lie within the speech production system rather than in higher-level cognitive processes. Methodologically, these results suggest that compensation for formant perturbations should be measured prior to 300ms after vowel onset to avoid any potential impact of attention or other higher-order cognitive factors.

Introduction

Externally-applied perturbations to auditory feedback are frequently used to study mechanisms of sensorimotor control in speech production. For example, when the vowel formants (the resonant frequencies of the vocal tract) of the word *head* are altered to sound more like *hid*, speakers compensate by producing something closer to *had*, such that their perceived vowel is closer to the intended target. When auditory perturbations are applied in an unpredictable manner from repetition to repetition, participants make reactive, online adjustments which occur over the time course of a single vowel (Purcell & Munhall, 2006b; Tourville et al., 2008). This *compensation* reflects sensitivity to sensory information in online feedback control. Auditory perturbations also cause speakers to make adjustments to their predictive motor plans for speech, resulting in changes on future productions (Hantzsch et al., 2022; Houde & Jordan, 1998; Purcell & Munhall, 2006a). This *adaptation* measures the ability of speakers to update their feedforward control system in the presence of sensory errors.

While population-level responses to altered auditory feedback are consistently observed in both adaptation and compensation, there is a large degree of variability in the magnitude of both responses across individual participants (Cai et al., 2012; Munhall et al., 2009; Niziolek & Parrell, 2021). The source of this variability is poorly understood, and has typically been investigated at the trait level. A frequently suggested cause is differences in auditory acuity between speakers, with the idea that speakers who are more sensitive to small auditory differences will better detect and respond to auditory perturbations. This idea has some support in the literature, with several studies showing significant positive correlations between auditory acuity and adaptation (Martin et al., 2018; Nault & Munhall, 2020; Villacorta et al., 2007) or compensation (Lester-Smith et al., 2020). However, the strength of this relationship is relatively weak (R^2 of 0.15-0.3 for vowel formant formant acuity and adaptation), and is not always found (Cai et al., 2012; Feng et al., 2011), suggesting that auditory acuity is not the sole determinant of adaptation magnitude. Additionally, there is little evidence that responses to auditory perturbations are predicted by cognitive traits outside the speech motor control system proper, as measures of executive/cognitive control such as the flanker and numerical Stroop tasks are not correlated with adaptation magnitude in speech (Martin et al., 2018)

Although trait-level cognitive processes do not seem to affect speech motor control, there is evidence from other motor domains that attention may play a modulatory role in the use of sensory feedback in human motor control systems. In one study of upper limb control (Taylor & Thoroughman, 2007), participants held on to a robotic arm while reaching for a target, and a force perturbation pushed their arm off course at some point along the trajectory. As a way to divide their attention, they were asked to simultaneously discriminate between two tones at varying points during the reaching movement. Discriminating between the tones reduced cross-trial adaptation to the imposed force, suggesting that divided attention disrupts the feedforward control system's ability to update sensory errors. Conversely, divided attention did not affect the magnitude of the online compensatory response. On the other hand, evidence from studies of vocal pitch control implicates attention in both online and adaptive responses:

participants whose attention was divided by a secondary visual identification task while vocalizing at a steady pitch to match a target note (i.e., saying “ah”) show reduced responses to perturbations of vocal pitch, both in online compensation (Scheerer et al., 2016; Tumber et al., 2014) and in feedforward adaptation (Scheerer et al., 2016). However, there are substantial differences between vocal pitch control and control of the oral articulators used to produce consonants and vowels in speech, and there is strong evidence that pitch and oral articulation differ in their use of sensory feedback (for a review, see Parrell & Houde, 2019).

In oral articulation for speech, there is little data concerning the role of attention in online compensatory behavior. However, there is a growing body of evidence suggesting that cross-trial adaptation in speech is a highly automatic process unaffected by cognitive processes generally, unlike adaptation in reaching which has been shown to partially rely on explicit strategies (Taylor et al., 2014). Speakers who are fully informed of vowel formant perturbations and instructed to ignore them nonetheless exhibit adaptation of an equivalent magnitude to naïve participants (Munhall et al., 2009). Speakers show no signs that they use any cognitive strategies in response to formant perturbations (K. S. Kim & Max, 2021; Parrell & Niziolek, 2021). Lastly, when speakers are exposed to formant perturbations while producing manual movements controlling center-out trajectories of a visual cursor, speech adaptation magnitude is equivalent to a control condition without manual movements. Conversely, adaptation of these manual movements to a visual perturbation (visuomotor rotation) *is* impaired by concurrent speech production (Lametti et al., 2020). Although this last result suggests performing concurrent motor tasks impairs manual control, it is not clear to what extent co-producing speech with manual movements selectively impairs attention to speech production given the almost universal co-production of manual gesture with speech (e.g., de Ruiter, 1998; Levelt et al., 1985; McNeill, 2005).

In sum, while there is evidence from both limb and vocal pitch control that attention may modulate feedback-based online compensation and feedforward adaptation, these effects seem to differ across motor domains. It remains uncertain to what extent, if any, attention modulates the use of sensory feedback for error correction in speech motor control. In two experiments, we use a visual divided attention task to examine whether attention has a modulatory effect on compensation (Experiment 1) and adaptation (Experiment 2) in response to altered auditory feedback of vowel formants.

Methods

Experiment 1

Participants

Twenty-six participants (20 women, 6 men), ages 18-62 years ($M = 30.7$, $SD = 15.7$), completed Experiment 1. Participants were naïve to the feedback manipulation used in the experiment and were excluded if they had a self-reported history of speech or neurological disorders, did not

report having normal or corrected-to-normal vision, or failed a Hughson-Westlake hearing screening (threshold > 25 dB HL at any of 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz in either ear). Participants completed all Experiment 1 testing in one session. All participants were compensated for their participation either monetarily or through course credit. One additional participant began the study but was excluded due to software failure during data collection.

Equipment

Participants sat approximately 71 cm from a computer screen with a resolution of 1920 x 1080 pixels, which subtended a visual angle of approximately $40 \times 27^\circ$. A written word stimulus was displayed at eye level. On each trial, participants spoke a stimulus word out loud. Speech was recorded at 16 kHz with a head-mounted microphone (AKG C520) and split to two USB sound cards (Focusrite Scarlett 2i2). One sound card was used to process, alter (on some trials), and play back speech to participants using Audapter (Cai et al., 2008; Tourville et al., 2013). Speech feedback was played over closed-back circumaural headphones (Beyerdynamic DT 770) at a volume of ~ 80 dB SPL, and was mixed with speech-shaped noise at ~ 60 dB SPL in order to mask the participant's perception of their unaltered speech. Speech was processed in the same manner on every trial, regardless of whether a perturbation was applied. The delay induced by this processing is ~ 18 ms as measured following Kim et al. (2020). The second sound card was used to synchronize the presentation of the visual distractor task with speech production using Psychtoolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). The onset of vocalization was defined as the time when the speech signal first exceeded a minimal amplitude threshold. If no voice onset was detected, the word stimulus disappeared after 1.5 s and the trial was repeated.

Visual stimuli

On selected trials (see below), participants made judgments of the gestalt motion direction (right or left) of a random dot kinematogram (RDK, Boynton, 2009). The dot field stimuli consisted of 100 white dots in a circular aperture of $7 \times 7^\circ$ against a black background. The dots had a diameter of 10 pixels and moved at $3^\circ/\text{s}$, with a lifetime of 200 ms (12 frames at a 60 Hz refresh rate). At the end of a dot's lifetime, or if a dot moved out of the aperture, it was replaced in a random location, moving in its previously assigned direction. Similar procedures have been widely used in studies of visual motion perception (Edwards & Badcock, 1995; Milne et al., 2002; Thompson et al., 2017; von Lautz et al., 2019). The coherence of dot motion could range from 1 (all dots moving in the same direction) to 0 (all dots moving in random directions). To ensure the RDK task was equally challenging for all participants, the coherence of dot motion was set on a participant-specific basis to achieve 75% correct responses. The RDK was displayed at speech onset for 300 ms to ensure that participants had to make direction-of-motion judgements while producing speech. Both speech onset detection and visual field stimulus presentation was controlled through Psychtoolbox-3 (Kleiner et al., 2007; Pelli, 1997).

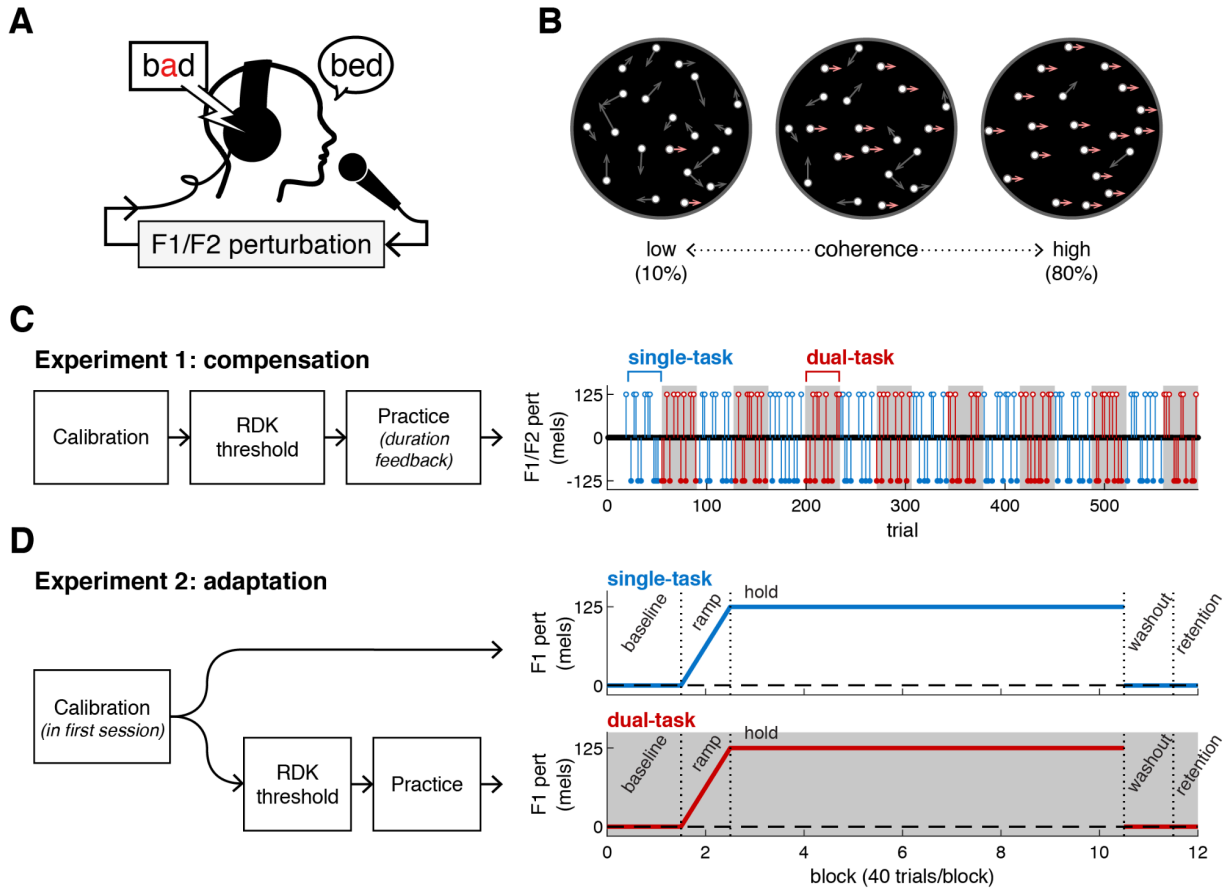


Figure 1. Experiment design. *A*: Schematic of the altered auditory feedback apparatus. *B*: Schematic of the random dot kinematogram (RDK) at 10, 50, and 80% coherence levels (Robertson et al., 2012). The task difficulty is lower as coherence increases. *C*: Timeline for Experiment 1 (compensation). Blocks of single-task (blue) and dual-task (red; gray shaded background) conditions alternated over the course of the experiment. Vowel formants were shifted towards /ae/ (upward; open circles) or towards /i/ (downward, filled circles), or were unaltered (black line). *D*: Timeline for Experiment 2 (adaptation). The single- and dual-task conditions were run in separate sessions, counterbalanced in order.

Pre-experiment calibration and training

Prior to the main experiment, three pre-experiment phases were completed: 1) a speech calibration phase to set the linear predictive coding (LPC) order used for real-time tracking of vowel formants and to define the direction of the auditory perturbation, 2) a phase to determine the threshold of RDK coherence, and 3) a practice phase.

Participants first completed a brief speech production task in order to set a participant-specific LPC order in Audapter. Participants read the words “bid,” “bed,” and “bat” as they appeared on the computer screen. These words have vowels that cover the range of vowel formants participants could be expected to produce during the main phase of the experiment. Participants produced each word 10 times in a randomized order. From these tokens, the best

LPC order was picked via visual inspection, starting with the gender-specific LPC defaults in Audapter. Based on these tokens, participant-specific vowel formant shifts were designed. The direction of the “upward” shift was defined as the angle, in the space defined by the first two vowel formants (F1 and F2), between the mean F1 and F2 values for the /ε/ vowel in *bed* and the /æ/ vowel in *bad*. The direction of the “downward” shift was similarly defined as the angle between the mean formant values the vowel in *bed* and the /ɪ/ vowel in *bid*. Formant values for each token were taken from the middle 50% of the vowel. The magnitude of both shifts was fixed at 125 mels (a perceptually-calibrated unit of frequency). Generally, the upward shift resulted in a raising of F1 and, to a lesser extent, a lowering of F2. Conversely, the downward shift generally resulted in a lowering of F1 and, to a lesser extent, a raising of F2. Similar participant-specific formant shifts have been shown to maximize the amount of observed compensation relative to shifts of only F1 (Daliri et al., 2020).

Second, the threshold of coherence for the RDK task (the percentage of dots that are moving in one direction) was determined for each participant using a three-down one-up staircase method, leading to a target threshold of 79.4% correct (Leek, 2001). The initial coherence level was 1 (all dots moving in the same direction); the initial step size was 50% (up or down) of the current coherence level, which reduced to 25% after the fourth reversal (c.f., Edwards & Badcock, 1995; Johnson et al., 1992; Mason et al., 2003). The coherence threshold was determined for each participant so the distractor task could be equally as challenging for all participants. The coherence of the last six reversals was averaged, and this value was used for the main phase of the experiment. Coherence thresholds ranged from 4-100% across participants (mean of 59%). Overall accuracy on the RDK during the main experiment was $64.4 \pm 17.7\%$ (23.3-93.3% across participants), suggesting that the thresholding made the task sufficiently difficult to divide attention.

Third, participants completed a practice phase in which they were trained to produce vowels with a sufficient duration to measure compensatory behavior, which has a latency of ~150 ms after vowel onset (Cai et al., 2012; Parrell et al., 2017; Tourville et al., 2008). Participants first completed a 10-trial block of the speech production task only, followed by a 10-trial block of speech trials with the concurrent RDK task. After each trial, participants were cued with visual feedback to help achieve a target vowel duration between 400-650 ms. Vowel duration was estimated as the duration for which the amplitude of the speech signal was above a minimum threshold. If duration fell within the acceptable range, a green circle was presented. If the vowel duration was too long, a yellow circle was presented with the text “speak a little faster,” and if the vowel duration was too short, a blue circle was presented with the text “speak a little more slowly.” Each practice block was redone, if necessary, until at least 8/10 trials were in the correct duration range.

Main experiment procedure

The experiment consisted of two conditions. In the dual-task condition, an RDK was presented on each trial at the onset of participants’ vocalization. Participants were instructed to indicate the

global direction of dot movement by pressing the left or right arrow key. The single-task condition was identical, except no RDK was presented.

At the start of each trial, a single monosyllabic stimulus word (*head*, *bed*, or *dead*) was presented on the screen. Participants were instructed to read the words on the screen as they appeared. Participants were given visual feedback about their vowel duration on every trial, as described above. Trial duration was 1.5 s, with an inter-trial interval of 750 ms plus an additional jitter of 0-750 ms.

The experiment consisted of 33 blocks of 18 trials each (594 total trials). Within each block, each stimulus word was presented six times. For two repetitions of each stimulus word, vowel formants were altered, with one “upward” shift (towards /æ/ as in *bad*) and one “downward” shift (towards /ɪ/ as in *bid*) for each stimulus word. Words and perturbations within each block were pseudorandomized, with the constraint that no adjacent trials received identical perturbations either within or across blocks.

The first block of the experiment was always an acclimatization block with no auditory feedback alteration and no RDK; this data was not included in the final analysis but was included to make sure the participant was used to hearing their speech processed through our system. Starting on block 2, the control and dual-task conditions alternated every two blocks. Half of the participants started with the control condition in block 2, and the other half started with the dual-task condition. Due to an error in the counter-balancing code, the first ten participants had an uneven number of trials distributed to each condition (31 total blocks, leading to 14 blocks of one condition and 16 blocks of another). However, we observed no difference in behavior in this subgroup; as such, this group was included in the final analysis. A self-timed break was given each time the conditions switched, and instructions were presented on the screen at each break so participants knew which task to complete next.

At the end of the experiment, participants completed a questionnaire assessing their awareness of the perturbation and any potential strategies they used.

Data analysis

Vowel formants were tracked using *wave_viewer* (Niziolek & Houde, 2015), a tool that provides a MATLAB GUI interface to Praat (Boersma & Weenink, 2019). Vowel onset and offset were first estimated based on a participant-specific amplitude threshold, then hand-checked. Errors in vowel onset and offset were corrected by hand-marking the location of the vowel using the spectrogram and waveform of the speech sample. Vowel onset was identified by the presence of F1 and F2 on the spectrogram and periodicity on the waveform. Vowel offset was identified when F1 and F2 were no longer visible on the spectrogram. Within this marked time range, formant values were tracked based on specific parameters set for each participant (LPC order and pre-emphasis); these parameters were adjusted on a per-trial basis if there were formant tracking errors. Trials consisting of errors in production or disfluencies were removed. Across participants, 1.2% of trials were excluded (0-9% per participant).

Word-specific baseline formant trajectories were calculated as the mean of the F1 and F2 trajectories for all trials without an auditory perturbation. These baseline trajectories were then

subtracted, on a word-specific basis, from the trajectories of each perturbed trial, giving a measure of the deviation from the expected behavior caused by the auditory perturbations. Normalized trajectories in F1/F2 space were then projected onto the angular direction of the auditory perturbation for each participant to measure compensation (Daliri et al., 2020; Kothare et al., 2020; Niziolek & Guenther, 2013; Niziolek & Parrell, 2021). The magnitude of the compensatory response was measured as the mean value of this compensation trajectory from 200-300 ms after the onset of the vowel, measured separately for the single- and dual-task conditions. Given our strong hypothesis that divided attention should reduce compensation in the dual-task condition, a one-tailed t-test ($\alpha = 0.05$) was used to test for differences between the conditions. After inspecting the resulting data, an additional post-hoc analysis was done assessing differences between the conditions in a later window from 300-400 ms after vowel onset, with a correction for multiple comparison (one-tailed t-test, $\alpha = 0.025$). This window represents a later phase of the response while still being under the minimum duration participants were required to produce. Summary statistics report means and standard errors across participants.

Experiment 2

Participants

Twenty participants (all women), ages 18-39 years ($M = 21.35$, $SD = 4.63$), completed Experiment 2. Exclusionary criteria were the same as those for Experiment 1. Participants completed two sessions, separated by at least one week.

Equipment

The same equipment for visual stimulus presentation, RDK presentation, and auditory feedback alterations was used in Experiment 2 as in Experiment 1.

Pre-experiment phases

All participants completed the speech calibration as in Experiment 1 during the first session. The same LPC order was used for both sessions. In the dual-task session, participants completed both the RDK threshold setting and practice phases as in Experiment 1. Because there was no requirement for speech of a certain length, there was no duration feedback given during this phase. The practice phase was repeated only if the participant requested extra practice. For participants assigned to the dual-task condition in session 1, the RDK coherence thresholding and practice phases followed the speech calibration phase. RDK coherence thresholds ranged from 4-100% across participants (mean of 59%). Overall accuracy on the RDK during the main experiment was $66.4 \pm 23.9\%$ (1.0-84.8% across participants).

Procedure

Experiment 2 consisted of the same single- and dual-task conditions as Experiment 1. The single- and dual-task conditions were completed in separate sessions, with the order of the sessions counterbalanced across participants. Each session consisted of a 30-trial baseline phase

with unaltered feedback, a 120-trial hold phase with F1 shifted by +125 mels, a 30-trial washout phase with unaltered feedback, and a 30-trial retention phase with unaltered feedback. There was a ten-minute pause between the end of the washout phase and the start of the retention phase in which the participant was instructed to not speak to avoid potential de-adaptation induced by normal speech (Parrell & Niziolek, 2021). A break was given every 30 trials. Words were distributed equally in each phase, but the order of these words was randomized. Trial duration was 1.5 s, with an inter-trial interval of 750 ms plus an additional jitter of 0-750 ms. After the end of session 2, participants completed a questionnaire assessing their awareness of the perturbation, and any potential strategies they used.

Data analysis

Vowel boundaries and formants were analyzed as in Experiment 1. Across participants, 1.9% of trials were excluded (0-10%) for production errors or disfluencies. F1 values were measured 50-100 ms after vowel onset to avoid 1) the effects of coarticulation from the onset consonant during the first 50 ms of the vowel and 2) any influence of online compensatory responses, which typically begin ~150 ms after vowel onset (Cai et al., 2012; Parrell et al., 2017; Tourville et al., 2008). All trials within a session were normalized to the mean F1 produced on the last 10 trials of the baseline phase. Adaptation was measured as the mean of this normalized F1 in the last 10 trials of the hold phase. Aftereffects were measured as the mean F1 during the first 10 trials of the washout phase. All responses were sign-flipped such that positive values reflect changes in speech that opposed the (upward) F1 perturbation, as in Experiment 1. Statistical analysis was conducted using a repeated-measures ANOVA with main effects of phase (adaptation vs. aftereffects) and session (single- vs dual-task), and their interaction. Two-tailed t-tests against a mean of 0 were used to assess whether learning occurred in each phase of each session. Though it was not our a priori focus, we additionally conducted a post-hoc analysis of persistent changes in F1 during the retention phase, comparing the single- and dual-task conditions against each other using a two-sided t-test. Effect sizes are reported as Cohen's *d*.

Results

Experiment 1

Experiment 1 assessed compensation to unpredictable F1 perturbations that were randomly applied to speech feedback. Participants compensated for the perturbations in both the single- and dual-task conditions (Fig 1A). Although the magnitude of compensation in the 200-300 ms window after vowel onset was slightly higher in the single-task (6.5 ± 1.3 mels) than in the dual-task condition (5.2 ± 1.0 mels), this difference was not significant ($t(25) = 1.3$, $p = 0.1$, $d = 0.25$). In the later 300-400 ms window, not only was compensation larger in both the single-task (9.4 ± 2.1 mels) and dual-task (6.6 ± 1.7 mels) than in the earlier window, but the compensation in the single-task condition was significantly larger than in the dual-task condition (2.9 ± 1.1 mels, $t(25) = 2.6$, $p = 0.007$, $d = 0.51$, Fig 1B,C). We ensured that this difference in the later time

window was not due to differences in vowel duration between the two conditions: average vowel durations extended well beyond 400 ms in both the single- (449 ± 12 ms) and dual-task (439 ± 13 ms) conditions. Although durations in the single-task condition were slightly longer than in the dual-task condition, both were substantially longer than the 400 ms analysis window, and the difference between the conditions was not significant ($t(25) = 2, p = 0.06, d = 0.39$). This suggests that the difference in the late window is not due to increased coarticulatory effects of the final consonant in the dual-task condition.

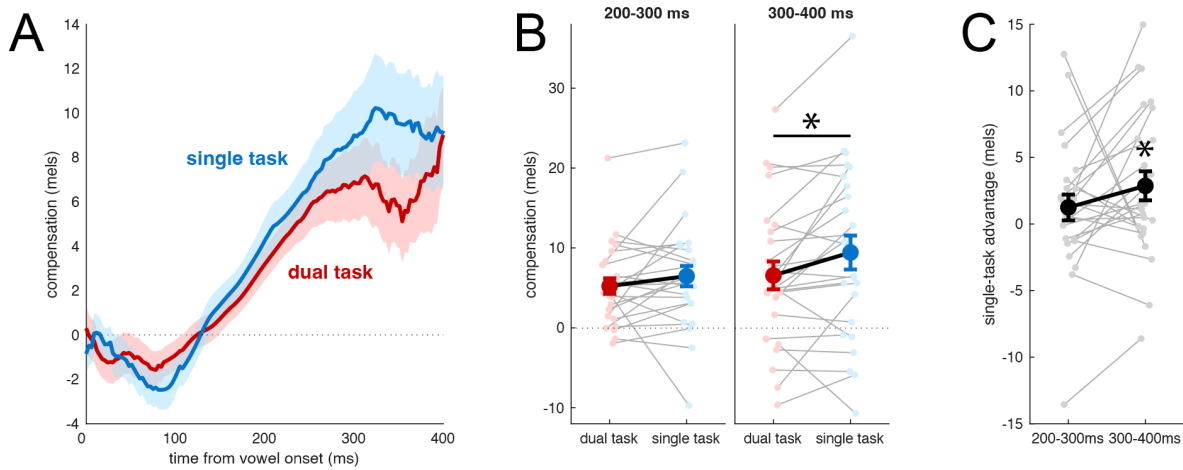
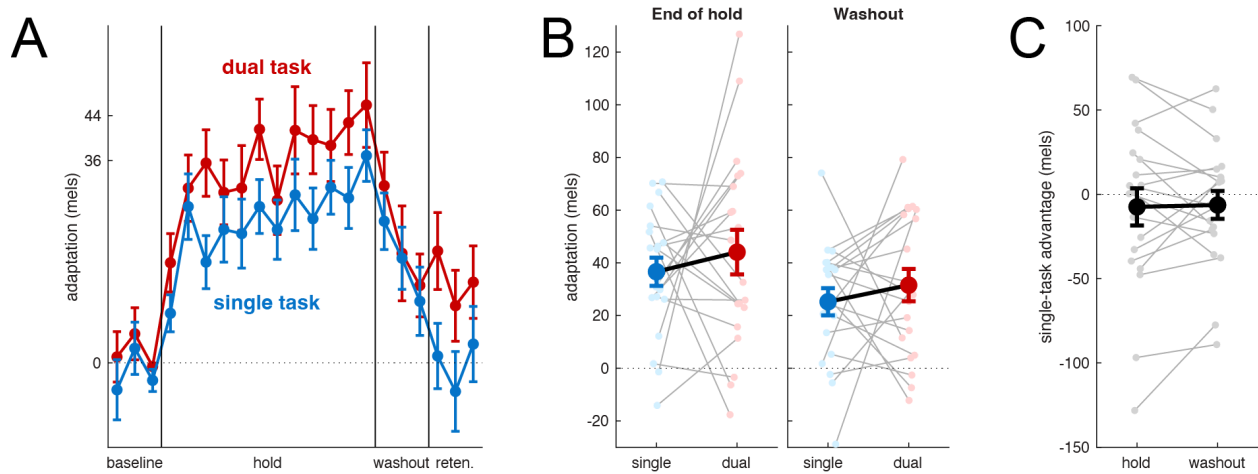


Figure 2: Compensation results (Experiment 1). **A:** Compensatory changes produced in response to the vowel formant perturbation, measured from vowel onset (blue: single-task; red: dual-task). Solid lines show mean responses and shaded areas represent standard error across participants. **B:** Average response magnitude in the planned, early window (200-300 ms after vowel onset, left) and the unplanned, late window (300-400 ms after vowel onset, right). Large solid circles show means and standard error across participants. Individual participants are shown in light circles. Light gray lines connect responses from the same participant. **C:** Difference in magnitude between the single- and dual-task conditions in the early (left) and late (right) windows. Positive values indicate larger adaptation in the single-task condition (a single-task advantage); negative values indicate larger adaptation in the dual-task condition. * represents a significant difference between the conditions, corrected for multiple comparisons.

Experiment 2

At the end of the hold phase, participants adapted to oppose the alteration to F1 in both the single- (36.6 ± 5.4 mels, $t(19) = 6.82, p < 0.0001, d = 1.46$) and dual-task (44.0 ± 8.5 mels, $t(19) = 5.2, p < 0.0001, d = 1.12$) conditions. Adaptation was slightly reduced in the washout phase (single-task: 25.2 ± 6.12 mels, $t(19) = 4.9, p = 0.0001, d = 1.05$; dual-task: 31.5 ± 6.1 mels, $t(19) = 5.16, p < 0.0001, d = 1.11$), but the difference between the two phases was not significant ($F(1,57) = 3.77, p = 0.06$). Critically, there was no difference between learning in the single- and dual-task conditions ($F(1,57) = 1.25, p = 0.27$) and no interaction between condition and experiment phase ($F(1,57) = 0.01, p = 0.93$). In short, divided attention had no effect on

adaptation. Post-hoc tests showed that although retention was numerically higher in the dual-task condition (14.8 ± 6.1 mels) than in the single-task condition (0.2 ± 5.6 mels), this difference was not significant ($t(38) = 1.74$, $p = 0.10$, $d = 0.56$).



wiscPaswis Figure 2: Adaptation results (Experiment 2). Changes in F1 that opposed the applied F1 shift are shown as positive values in all panels. **A:** F1 changes over time produced in response to the vowel formant perturbation. Dots represent the mean of 10 trials. Blue trace shows single-task condition; red trace, dual-task condition. Filled circles and error bars show mean responses and standard error across participants. **B:** Adaptation magnitude at the end of the hold phase (left) and in the washout phase (right). Large solid circles show means and standard error across participants. Individual participants are shown in light circles. Light gray lines connect responses from the same participant. **C:** Difference in magnitude between the single and dual-task conditions in the hold (left) and washout (right) phases. Positive values indicate larger adaptation in the single-task condition (a single-task advantage); negative values indicate larger adaptation in the dual-task condition.

Discussion

Experiments altering participants' auditory feedback have shown a consistent average opposing response, both within an utterance (compensation) and across repeated utterances (adaptation). However, responses are variable both within and across participants, and the sources driving this variability are currently unknown. Here, we used a divided attention task to assess whether attention modulates the ability of speakers to compensate and adapt. Divided attention decreased the magnitude of within-trial compensation toward the end of the syllable, but did not significantly modulate the magnitude of cross-trial adaptation.

The results from Experiment 1 showed no significant difference between the single- and dual-task conditions across participants in the planned analysis window from 200-300 ms after vowel onset. However, we did find a significant single-trial advantage in a later window, with smaller compensation in the dual-task condition 300-400 ms after vowel onset. One potential reason why the effects of divided attention emerge only later in the vowel could be because the initial component of the response relies on automatic control using the sensorimotor system, whereas the later component of the response relies at least somewhat on volitional control. A similar effect has been observed in vocal pitch control, where participants who were instructed to volitionally change their pitch in response to an auditory pitch perturbation made initial

short-latency (~150 ms) compensatory responses that opposed the perturbation regardless of the instructed direction of change before making longer-latency responses (~400 ms) in the instructed direction (Hain et al., 2000). Although this previous study examined the control of pitch, rather than vowel formants, the current results suggest a similar division in the response to sensory errors in supralaryngeal articulatory control for speech: an early, automatic response to auditory errors that is unaffected by cognitive factors (e.g., strategy and attention) as well as later responses affected by cognitive factors.

The fact that compensatory responses are modified by divided attention only starting 300 ms after vowel onset suggest that, under normal conditions, online feedback control in speech is not substantially impacted by higher-level cognitive factors, given that vowel durations in connected speech are substantially shorter than this (Umeda, 1975). From a practical perspective, however, it suggests that measurements of compensatory responses to auditory perturbations in speech should be sure to examine an early time window before this 300 ms mark. Especially when attempting to estimate the integrity of the feedback control system in patients with neurological disorders (e.g., Abur et al., 2021; Mollaei et al., 2016; Parrell et al., 2017, 2021), isolating the normal behavior of the speech motor system as it typically operates and in the absence of higher-level cognitive factors is critical.

In Experiment 2, participants adapted their speech to oppose the auditory perturbation equivalently in the single and dual task conditions. These results are consistent with previous work that has suggested speech adaptation is an implicit, automatic process that is unaffected by instructions (Munhall et al., 2009), strategy (K. S. Kim & Max, 2021), or concurrent manual motor activity (Lametti et al., 2020). The current results expand on these previous studies to show that attention, specifically, does not play a modulatory role in speech motor adaptation.

Broadly, the results from both experiments suggest that the use of sensory feedback in typical speech motor control is a mostly automatic process unaffected by attention. Along with previous research in this area, this suggests that the sources of variability frequently observed across participants in studies of both compensation and adaptation in speech likely lie within the speech production system, rather than in higher-level cognitive processes.

References

- Abur, D., Subaciute, A., Daliri, A., Lester, -Smith Rosemary A., Lupiani, A. A., Cilento, D., Enos, N. M., Weerathunge, H. R., Tardif, M. C., & Stepp, C. E. (2021). Feedback and Feedforward Auditory-Motor Processes for Voice and Articulation in Parkinson's Disease. *Journal of Speech, Language, and Hearing Research*, 64(12), 4682–4694. https://doi.org/10.1044/2021_JSLHR-21-00153
- Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer (Version 6.0.47)*. Praat. <http://www.praat.org/>
- Boynton, G. M. (2009). *movingDots* [Computer software]. <http://courses.washington.edu/matlab1/matlab/movingDots.m>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436. <https://doi.org/10.1163/156856897X00357>
- Cai, S., Beal, D., Ghosh, S., Tiede, M., Guenther, F., & Perkell, J. (2012). Weak responses to auditory feedback perturbation during articulation in persons who stutter: Evidence for abnormal auditory-motor transformation. *Plos One*, 7(7). <https://doi.org/10.1371/journal.pone.0041830>

- Cai, S., Boucek, M., Ghosh, S., Guenther, F. H., & Perkell, J. (2008). A System for Online Dynamic Perturbation of Formant Trajectories and Results from Perturbations of the Mandarin Triphthong /iau/. *Proceedings of the 8th International Seminar on Speech Production*, 65–68.
- Daliri, A., Chao, S.-C., & Fitzgerald, L. C. (2020). Compensatory Responses to Formant Perturbations Proportionally Decrease as Perturbations Increase. *Journal of Speech, Language, and Hearing Research*, 63(10), 3392–3407. https://doi.org/10.1044/2020_JSLHR-19-00422
- de Ruiter, J.-P. P. (1998). *Gesture and speech production* [Doctoral dissertation]. Radboud University.
- Edwards, M., & Badcock, D. (1995). *Global motion perception: No interaction between the first and second-order motion pathways*. 35(18), 2589–2602. [https://doi.org/10.1016/0042-6989\(95\)00003-1](https://doi.org/10.1016/0042-6989(95)00003-1)
- Feng, Y., Gracco, V. L., & Max, L. (2011). Integration of auditory and somatosensory error signals in the neural control of speech movements. *Journal of Neurophysiology*, 106(2), 667–679. <https://doi.org/10.1152/jn.00638.2010>
- Hain, T., Burnett, T., Kiran, S., Larson, C., Singh, S., & Kenney, M. (2000). Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Exp Brain Res*, 130, 133–141. <https://doi.org/10.1007/s002219900237>
- Hantzsch, L., Parrell, B., & Niziolek, C. A. (2022). A single exposure to altered auditory feedback causes observable sensorimotor adaptation in speech. *eLife*, 11, e73694. <https://doi.org/10.7554/eLife.73694>
- Houde, J., & Jordan, M. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354). <https://doi.org/10.1126/science.279.5354.1213>
- Johnson, C. A., Chauhan, B. C., & Shapiro, L. R. (1992). Properties of staircase procedures for estimating thresholds in automated perimetry. *Investigative Ophthalmology & Visual Science*, 33(10), 2966–2974.
- Kim, K. S., & Max, L. (2021). Speech auditory-motor adaptation to formant-shifted feedback lacks an explicit component: Reduced adaptation in adults who stutter reflects limitations in implicit sensorimotor learning. *European Journal of Neuroscience*, 53(9), 3093–3108. <https://doi.org/10.1111/ejn.15175>
- Kim, K., Wang, H., & Max, L. (2020). It's About Time: Minimizing Hardware and Software Latencies in Speech Research with Real-Time Auditory Feedback. *Journal of Speech, Language, and Hearing Research*, 63(8), 2522–2534. https://doi.org/10.1044/2020_JSLHR-19-00419
- Kleiner, M., Brainard, D. H., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, 36, 1–16. <https://doi.org/10.1068/v070821>
- Kothare, H., Raharjo, I., Ramanarayanan, V., Ranasinghe, K., Parrell, B., Johnson, K., Houde, J. F., & Nagarajan, S. S. (2020). Sensorimotor adaptation of speech depends on the direction of auditory feedback alteration. *The Journal of the Acoustical Society of America*, 148(6), 3682–3697. <https://doi.org/10.1121/10.0002876>
- Lametti, D. R., Quek, M. Y. M., Prescott, C. B., Brittain, J.-S., & Watkins, K. E. (2020). The perils of learning to move while speaking: One-sided interference between speech and visuomotor adaptation. *Psychonomic Bulletin & Review*, 27(3), 544–552. <https://doi.org/10.3758/s13423-020-01725-8>
- Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception &*

- Psychophysics*, 63(8), 1279–1292. <https://doi.org/10.3758/BF03194543>
- Lester-Smith, R. A., Daliri, A., Enos, N., Abur, D., Lupiani, A. A., Letcher, S., & Stepp, C. E. (2020). The Relation of Articulatory and Vocal Auditory–Motor Control in Typical Speakers. *Journal of Speech, Language, and Hearing Research*, 63(11), 3628–3642. https://doi.org/10.1044/2020_JSLHR-20-00192
- Levelt, W. J. M., Richardson, G., & La Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language*, 24(2), 133–164. [https://doi.org/10.1016/0749-596X\(85\)90021-X](https://doi.org/10.1016/0749-596X(85)90021-X)
- Martin, C., Niziolek, C., Dunabeitia, J., Perez, A., Hernandez, D., Carreiras, M., & Houde, J. (2018). Online Adaptation to Altered Auditory Feedback is Predicted by Auditory Acuity and Not by Domain-General Executive Control Resources. *Frontiers in Human Neuroscience*, 12. <https://doi.org/10.3389/fnhum.2018.00091>
- Mason, A. J. S., Braddick, O. J., & Wattam-Bell, J. (2003). Motion coherence thresholds in infants—different tasks identify at least two distinct motion systems. *Vision Research*, 43(10), 1149–1157. [https://doi.org/10.1016/S0042-6989\(03\)00077-4](https://doi.org/10.1016/S0042-6989(03)00077-4)
- McNeill, D. (2005). *Gesture and thought*. University of Chicago Press.
- Milne, E., Swettenham, J., Hansen, P., Campbell, R., Jeffries, H., & Plaisted, K. (2002). High motion coherence thresholds in children with autism. *Journal of Child Psychology and Psychiatry*, 43(2), 255–263. <https://doi.org/10.1111/1469-7610.00018>
- Mollaei, F., Shiller, D. M., Baum, S. R., & Gracco, V. L. (2016). Sensorimotor control of vocal pitch and formant frequencies in Parkinson’s disease. *Brain Res*, 1646, 269–277. <https://doi.org/10.1016/j.brainres.2016.06.013>
- Munhall, K. G., MacDonald, E. N., Byrne, S. K., & Johnsrude, I. (2009). Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *The Journal of the Acoustical Society of America*, 125(1), 384–390. <https://doi.org/10.1121/1.3035829>
- Nault, D. R., & Munhall, K. G. (2020). Individual variability in auditory feedback processing: Responses to real-time formant perturbations and their relation to perceptual acuity. *The Journal of the Acoustical Society of America*, 148(6), 3709–3721. <https://doi.org/10.1121/10.0002923>
- Niziolek, C. A., & Guenther, F. H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *Journal of Neuroscience*, 33(29), 12090–12098. <https://doi.org/10.1523/JNEUROSCI.1008-13.2013>
- Niziolek, C. A., & Houde, J. (2015). *Wave Viewer: First Release* [Computer software]. <https://doi.org/10.5281/ZENODO.13839>
- Niziolek, C. A., & Parrell, B. (2021). Responses to Auditory Feedback Manipulations in Speech May Be Affected by Previous Exposure to Auditory Errors. *Journal of Speech, Language, and Hearing Research*, 64(6S), 2169–2181. https://doi.org/10.1044/2020_JSLHR-20-00263
- Parrell, B., Agnew, Z., Nagarajan, S., Houde, J. F., & Ivry, R. B. (2017). Impaired Feedforward Control and Enhanced Feedback Control of Speech in Patients with Cerebellar Degeneration. *Journal of Neuroscience*, 37(38), 9249–9258. <https://doi.org/10.1523/JNEUROSCI.3363-16.2017>
- Parrell, B., & Houde, J. (2019). Modeling the Role of Sensory Feedback in Speech Motor Control and Learning. *Journal of Speech, Language, and Hearing Research*, 62(8S), 2963–2985. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0127

- Parrell, B., Kim, H. E., Breska, A., Saxena, A., & Ivry, R. B. (2021). Differential effects of cerebellar degeneration on feedforward versus feedback control across speech and reaching movements. *Journal of Neuroscience*, 41(42), 8779–8789. <https://doi.org/10.1523/JNEUROSCI.0739-21.2021>
- Parrell, B., & Niziolek, C. A. (2021). Increased speech contrast induced by sensorimotor adaptation to a nonuniform auditory perturbation. *Journal of Neurophysiology*, 125(2), 638–647. <https://doi.org/10.1152/jn.00466.2020>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442. <https://doi.org/10.1163/156856897X00366>
- Purcell, D. W., & Munhall, K. G. (2006a). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *Journal of Acoustical Society of America*, 120(2), 966–977.
- Purcell, D. W., & Munhall, K. G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *Journal of Acoustical Society of America*, 119(4), 2288–2297.
- Robertson, C. E., Martin, A., Baker, C. I., & Baron-Cohen, S. (2012). Atypical Integration of Motion Signals in Autism Spectrum Conditions. *PLOS ONE*, 7(11), e48173. <https://doi.org/10.1371/journal.pone.0048173>
- Scheerer, N., Tumber, A., & Jones, J. (2016). Attentional demands modulate sensorimotor learning induced by persistent exposure to changes in auditory feedback. *Journal of Neurophysiology*, 115(2), 826–832. <https://doi.org/10.1152/jn.00799.2015>
- Taylor, J. A., Krakauer, J. W., & Ivry, R. B. (2014). Explicit and implicit contributions to learning in a sensorimotor adaptation task. *J Neurosci*, 34(8), 3023–3032. <https://doi.org/10.1523/JNEUROSCI.3619-13.2014>
- Taylor, J. A., & Thoroughman, K. A. (2007). Divided Attention Impairs Human Motor Adaptation But Not Feedback Control. *Journal of Neurophysiology*, 98(1), 317–326. <https://doi.org/10.1152/jn.01070.2006>
- Thompson, B., McKinlay, C., Chakraborty, A., Anstice, N., Jacobs, R., Paudel, N., Yu, T.-Y., Ansell, J., Woules, T., & Harding, J. (2017). Global motion perception is associated with motor function in 2-year-old children. *Neuroscience Letters*, 658, 177–181. <https://doi.org/10.1016/j.neulet.2017.08.062>
- Tourville, J. A., Cai, S., & Guenther, F. (2013). Exploring auditory-motor interactions in normal and disordered speech. *Proceedings of Meetings on Acoustics*, 19, 060180. <https://doi.org/10.1121/1.4800684>
- Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*, 39(3), 1429–1443. <https://doi.org/10.1016/j.neuroimage.2007.09.054>
- Tumber, A., Scheerer, N., & Jones, J. (2014). Attentional demands influence vocal compensations to pitch errors heard in auditory feedback. *Plos One*. <https://doi.org/10.1371/journal.pone.0109968>
- Umeda, N. (1975). Vowel duration in American English. *The Journal of the Acoustical Society of America*, 58(2), 434–445. <https://doi.org/10.1121/1.380688>
- Villacorta, V., Perkell, J., & Guenther, F. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America*, 122. <https://doi.org/10.1121/1.2773966>
- von Lautz, A., Herding, J., & Blankenburg, F. (2019). Neuronal signatures of a random-dot

motion comparison task. *NeuroImage*, 193, 57–66.
<https://doi.org/10.1016/j.neuroimage.2019.02.071>