JASA EXPRESS LETTERS ARTICLE asa.scitation.org/journal/jel

Formant variability is related to vowel duration across speakers

Ding-lan Tang,[1] **Caroline A. Niziolek,**[2,3,a)] **and Benjamin Parrell**[2,3,a)]
[1]*Academic Unit of Human Communication, Learning, and Development, The University of Hong Kong, Hong Kong, Special Administrative Region, China*
[2]*Waisman Center, The University of Wisconsin—Madison, Madison, Wisconsin 53705, USA*
[3]*Department of Communication Sciences & Disorders, University of Wisconsin—Madison, Madison, Wisconsin 53706, USA*

Abstract: This study examined both the between-subject and within-subject relationships between vowel duration and formant variability during productions of both isolated words and connected speech by analyzing three existing datasets ($N = 132$). A positive between-subject correlation was observed in isolated words and, marginally, in connected speech. This finding is consistent with the idea that individuals who are more variable rely more on feedback-based control for vowel production, as longer durations allow more time for online corrections. Conversely, no such correlation was found within speakers at the trial level, suggesting that individuals do not modify their vowel duration online for each production.

[Editor: Peggy Mok] https://doi.org/10.1121/10.0039754

Received: 28 August 2025 **Accepted:** 17 October 2025 **Published Online:** 4 November 2025

1. Introduction

Duration has long been a crucial feature when describing and analyzing vowel productions and is known to vary both across and within speakers. For example, compared to healthy controls, slower speech rates (resulting in longer vowel durations) have been reported in individuals with Down and Williams syndromes (Bunton and Leddy, 2011; Setter *et al.*, 2007) and motor speech disorders, including dysarthria (Kent *et al.*, 1979; Liss *et al.*, 2009) and apraxia of speech (Collins *et al.*, 1983). Even when controlling for speech rate, individuals differ in their vowel durations. Vowels produced by women are typically longer than those produced by men during both isolated words (Hillenbrand *et al.*, 1995) and sentence productions (Jacewicz *et al.*, 2007). Developmentally, young children normally exhibit longer average vowel duration than older age groups (Lee *et al.*, 1999). Vowel duration also varies within speakers, which has been studied primarily through the lens of linguistic structure. Vowel duration is influenced by several contextual factors, including syllabic stress, pitch accent, the identities of adjacent segments, and the syllabic structure of a word (Van Santen, 1992). For example, stressed vowels are produced with a longer duration than unstressed vowels (Fry, 1958), and vowels preceding voiceless consonants are shorter than those preceding voiced consonants (Chen, 1970; Peterson and Lehiste, 1960).

While linguistic structure plays a large role in vowel duration, the duration of a single vowel produced by a single speaker can vary substantially even when repeated in the same context. One possible source of this non-linguistic variation is that vowel duration is related to the precision of vowel production. When we speak, we continuously monitor the resulting auditory feedback—the sound of our own voice. If there is a mismatch between the predicted and actual auditory feedback (i.e., prediction error), speakers will adjust their motor behavior to correct the online speech error (Guenther, 2016; Parrell *et al.*, 2019). Such an error-correction system operates in the presence of error induced by external perturbations (e.g., masking noise) or internal changes (e.g., muscle fatigue), serving to correct for the motor variability observed in our repetitive actions (Blustein *et al.*, 2021; Houde and Jordan, 2002; Tang *et al.*, 2022)

Previous studies have suggested that naturally occurring speech variability can elicit a prediction error signal: When speakers produce the same word repeatedly, neural responses evoked by auditory feedback in auditory cortex were less suppressed for productions that were farther away from the target in formant space (i.e., the center of the vowel distribution) than for those closer to the target (Beach *et al.*, 2024; Niziolek *et al.*, 2013; Tang *et al.*, 2025). This reduction in suppression is similar to the responses observed during speech errors induced by external perturbations (Behroozmand and Larson, 2011; Chang *et al.*, 2013; Kim *et al.*, 2025), suggesting that naturally occurring variability in vowel productions is treated like overt speech errors. If so, a vowel token produced on the periphery of a speaker's normal formant distribution may exhibit increased vowel duration in order to provide sufficient time to correct ongoing speech and better approximate vowel production targets. Such a relationship between vowel duration and formant variability could be a general property of individual speakers (e.g., speakers who are generally more variable may produce longer vowels) or could be due to differences in online control within speakers (e.g., productions that are initially farther from the vowel target may exhibit longer durations).

[a)] Corresponding author: cniziolek@wisc.edu and bparrell@wisc.edu

Here, we tested these two non-exclusive possibilities. Specifically, we examined both the between-subject and within-subject relationships between vowel duration and variability by analyzing existing datasets ($N = 91$) in which subjects repeatedly produced monosyllabic words with different vowel sounds (Parrell and Niziolek, 2021; Tang *et al.*, 2022). We found a significant between-subject correlation between vowel duration and variability during the production of isolated words. That is, individuals who are more variable tend to produce longer vowels. While not conclusive, this positive correlation between vowel duration and formant variability suggests that individuals who produce vowels with higher initial variability (and, as such, less accurate feedforward/predictive control systems) may rely more on feedback control during vowel production, as longer durations provide longer time for online articulatory adjustments and corrections. While increased weighting of feedback control has been suggested as a compensatory mechanism for impairments in feedforward/predictive control in some speech motor disorders (Houde *et al.*, 2019; Parrell *et al.*, 2017; Parrell *et al.*, 2021), this result may suggest the same trade-off occurs in healthy speakers. Understanding this balance offers valuable insights into the mechanisms underlying speech motor control and individual differences in speech production.

We then examined whether this between-subject correlation also holds in connected speech (sentence-level) by analyzing another existing dataset ($N = 41$) in which subjects produced 40 different sentences that varied markedly in length, vocabulary, and grammar (Beach *et al.*, 2024). The results revealed only a marginally significant positive correlation between vowel duration and formant variability at the subject level during productions of connected speech. In contrast, we did not find evidence of a within-subject correlation between these measures during productions of either isolated or connected speech, suggesting the relative weighting of feedforward and feedback control remains relatively stable and is unlikely to change on a production-by-production basis.

2. Method

2.1 Participants

We analyzed data ($N = 91$; see Table 1) from two previous studies (Parrell and Niziolek, 2021; Tang *et al.*, 2022) to examine both the between- and within-subject relationships between duration and formant variability in isolated words. Data from a third study ($N = 41$; see Table 1) were analyzed to examine the between-subject relationship between duration and formant variability in connected speech (Beach *et al.*, 2024). All participants were native speakers of American English and reported no history of speech, hearing, or neurological disorders.

2.2 Speech production experiments

Data from study 1 ($N = 25$) and three experiments in study 2 ($N = 66$) were included in the analyses for isolated words. In each experiment, one of the three monosyllabic English words ("bead," "bad," and "bod," containing the vowels /i/, /æ/, and /ɑ/, respectively) was pseudorandomly selected and displayed on a computer screen for 1.5 s, one at time. The interstimulus interval (ISI) was randomly jittered between 0.75 and 1.5 s to reduce anticipatory responses. Participants were instructed to read each word aloud as it appeared. Speech was recorded at 16 kHz via either a head-mounted microphone (AKG C520; AKG, Hofgeismar, Germany) or a desktop microphone (Sennheiser MKE 600; Sennheiser, Wedemark, Germany) and played back to participants with a short delay (∼18 ms) via closed-back circumaural headphones (Beyerdynamic DT 770; Beyerdynamic, Heilbronn, Germany). The speech playback volume varied with the volume of participants' speech and was calibrated to be approximately 80 dB sound pressure level (SPL); this playback was mixed with speech-shaped noise at a volume of approximately 60 dB SPL. The aims of studies 1 and 2 were to investigate how speakers modify their vowel productions in response to novel auditory perturbations. Accordingly, each experiment consisted of several phases, including baseline, ramp, hold, washout, and retention [see Parrell and Niziolek (2021) and Tang *et al.* (2022) for method details], with formant perturbations applied to participants' auditory feedback during both the ramp and hold phases. The current study analyzes only data from the baseline phase (120 trials), during which participants received normal (unaltered) auditory feedback.

Table 1. Summary of the included studies.

	Study 1 (Parrell and Niziolek, 2021)	Study 2 (Tang *et al.*, 2022)	Study 3 (Beach *et al.*, 2024)
No. of participants included in analysis (gender)	25 (21 females, 4 males)	66 (47 females, 19 males): 24 in experiment 1 22 in experiment 2 20 in experiment 4	41 (31 females, 10 males): 14 in younger group 14 in middle-aged group 13 in older group
Participants' age	Mean 20.4 ± 2.9 years	Experiment 1: mean 28.8 ± 12.3 years Experiment 2: mean 25.9 ± 9.7 years Experiment 4: mean 28 ± 12.9 years	Younger group: mean 20.2 ± 2 years Middle-aged group: mean 42.1 ± 8.4 years Older group: mean 66.5.1 ± 5 years
No. of stimuli (vowels in a particular context) included in the final analysis	3	3	30 (see Table S1 in the supplementary material)

In study 3, instead of isolated words, the stimuli were 40 phonetically balanced sentences selected from the Harvard sentences (IEEE, 1969). Similar to the isolated words task, stimuli were displayed on a computer screen one at a time. Participants were instructed to read the item aloud as it appeared. Each sentence was presented for 4.5 s, with the ISI randomly jittered between 0.75 and 1.5 s. The apparatus and setup used for acoustic recording were similar to those used in studies 1 and 2 [see Beach *et al.* (2024) for method details]. Study 3 was a two-visit study in which participants completed one session with and one without altered auditory feedback. The current analyses only included data from the session without altered auditory feedback, which involves 440 total productions of 40 unique sentences (11 repetitions each), containing 14 vowels (/ɑ/, /æ/, /ʌ/, /ɔ/, /aʊ/, /aɪ/, /ɛ/, /ɝ/, /eɪ/, /ɪ/, /i/, /oʊ/, /ʊ/, and /u/) across 30 individual stimulus contexts (see Table S1 in the supplementary material).

2.3 Acoustic analysis

Isolated words. The values of the first two vowel formants (F1 and F2, resonances of the vocal tract that distinguish different vowels) were tracked offline using wave_viewer (Niziolek and Houde, 2015), a MATLAB GUI interface for formant tracking using PRAAT (Boersma and Weenink, 2019). Linear predictive coding (LPC) order (i.e., the number of coefficients used to estimate the formants) and preemphasis values (i.e., adjustments to spectral tilt to improve formant estimation) were set for each participant at values that generally provided smooth formant tracks and were adjusted if necessary for individual trials where large jumps in formant values were visible. The detection of vowel onset and offset was automatically performed using a participant-specific amplitude threshold, and any errors were manually corrected using the waveform and spectrogram. Vowel onset was identified as the point where the waveform showed periodicity and the formants were visible in the spectrogram. Vowel offset was identified as the point where formants, particularly F2 and higher, were no longer visible. In general, across the two studies, an average of 1.6% of trials were excluded due to production errors or unresolvable errors in formant tracking.

Connected speech. Sentence data were first automatically segmented into component phonemes using the Montreal Forced Aligner (McAuliffe *et al.*, 2017). Trained research assistants reviewed the segmentation and manually adjusted the vowel boundaries when needed, using the waveform and spectrogram [see Beach *et al.* (2024) for details]. After participants' productions had been accurately marked, F1 and F2 of the vowels in the sentences were tracked offline following the same procedure used in isolated words (Parrell and Niziolek, 2021; Tang *et al.*, 2022).

Variability for each trial was measured as the two-dimensional (2D) distance in F1/F2 space between production of a vowel and the center of the distribution for that vowel, measured over the first 50 ms of the vowel (Fig. 1). For connected speech, each instance of a vowel appearing in the same contextual environment (i.e., the same location within the same sentence) was analyzed separately. For example, the vowels in the words "holes" and "rows" were treated as different "vowels," despite being represented by the same phoneme, /oʊ/. Our goal was to capture movement variability within a context, rather than variability due to coarticulation. Only stressed vowel contexts in the sentence productions were included in these analyses, with the aim of having vowel durations of more than 50 ms. In order to obtain a reliable formant distribution for each vowel context, only those vowel contexts that were produced at least nine times by a given participant were included for further analyses (average of 29.05 ± 1.76 tokens). In total, data from 30 stimuli (vowels in a particular context; see Table S1 in the supplementary material) were included in the sentence-level analysis.

2.4 Statistical analysis

The primary goal was to test the relationship between vowel duration and formant variability (i.e., the distance from the center of a vowel's distribution) at both between- and within-subject levels. For isolated words, both the between- and within- subject correlations were quantified in a linear mixed-effects (LME) model: *duration* ∼ *distance* + *meanDistance*

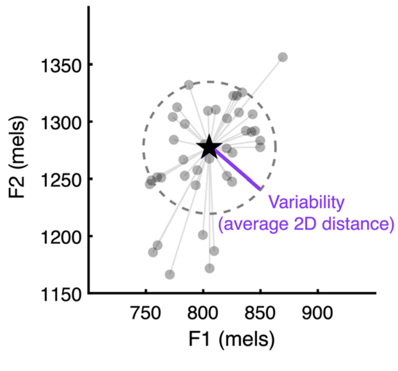

Fig. 1. Example from a representative participant showing how formant variability was measured. The initial distance for each trial was measured as the 2D distance in F1/F2 space between each production of a vowel (small gray dots) and the center of the distribution for that vowel (black star) during the first 50 ms of the vowel. Variability was then calculated as the average of these distances (purple line).

+ *vowel_æmeanDistance* + *vowel_imeanDistance* + *vowel_iDistance* + *vowel_æDistance* + *(1+vowel_æ|subject)* + *(1+vowel_i|subject)* + *(1+distance|subject)*. *Distance* is a subject centered distance variable defined by the difference between the trial level distance and the subject mean distance (*meanDistance*). Two dummy (binary) variables (*vowel_ae* and *vowel_i*) were created coding for two of the vowels against a reference vowel (/ɑ/). At the subject level (between-subject correlation), we additionally calculated Pearson's correlation coefficient (r) between vowel duration (i.e., the average duration of each vowel for a given participant) and variability (i.e., the average trial-wise distance of each vowel production from the center of that vowel's distribution for that participant) for each vowel. Moreover, to assess if individuals are consistent in the correlation across the three vowels, we first converted the average vowel durations and formant distances of each participant and vowel into z-scores. These z-scored points were then projected onto the correlation line, with positive projections indicating contributions toward the higher end (above the mean) and negative projections indicating contributions toward the lower end (below the mean). Subsequently, we computed the intraclass correlation coefficient (ICC) using a two-way mixed-effects model on these projections (Koo and Li, 2016). At the trial level (within-subject correlation), Pearson's r was calculated for each participant and vowel. A Fisher transformation was used to convert the correlation coefficients into z-scores, which were then compared against zero using one-sample t tests. We then tested whether any between-subject correlation occurred during productions of connected speech by conducting an LME model using data from study 3: *duration* ∼ *meanDistance* + *SenDuration* + *(1|vowel)* + *(1|subject)*, where *SenDuration* is the subject mean sentence duration. Pearson's r (between-subject correlation) was calculated for each vowel.

3. Results

3.1 Between-subject relationship in isolated words

As shown in Fig. 2, we found a significant positive correlation between vowel duration and formant distance (*meanDistance*, $\beta = 0.026$, $t = 3.42$, $p < 0.001$), such that speakers with larger formant variability produced longer vowels. Significant correlations were observed for each vowel individually (*bad* [/æ/], $r = 0.294$, $p = 0.005$; *bod* [/ɑ/], $r = 0.289$, $p = 0.005$; *bead* [/i/], $r = 0.293$, $p = 0.005$), with no difference across vowels. To further evaluate individual consistency across the three vowels, we calculated the ICC (based on a two-way mixed-effects model) of the individual participants' duration/variability locations within each vowel's distribution, using the projections onto each vowel-specific trend line. The across-vowel ICC was 0.83, representing a very good individual consistency across vowels; essentially, participants who tended to produce longer durations and greater formant variability in one vowel were also higher along the trend line in the other vowels.

3.2 Within-subject relationship in isolated words

While vowel duration increased with average variability across subjects, we found no evidence for a similar relationship between duration and formant distance within subjects at the trial level (distance, $\beta = 0.000$, $t = -0.84$, $p = 0.4$). Figure 3(A) shows the within-subject correlations from four representative participants as well as the between-subject correlation across all participants. Pearson's r was calculated for individual participants and vowels. Pearson's r (after Fisher z-transformation) did not differ from zero for any of the three vowels (*bad* [/æ/], mean $z = -0.026$, $p = 0.174$; *bod* [/ɑ/], mean $z = -0.028$, $p = 0.144$; *bead* [/i/], mean $z = 0.012$, $p = 0.522$), consistent with no within-subject relationship between vowel duration and variability [Fig. 3(B)].

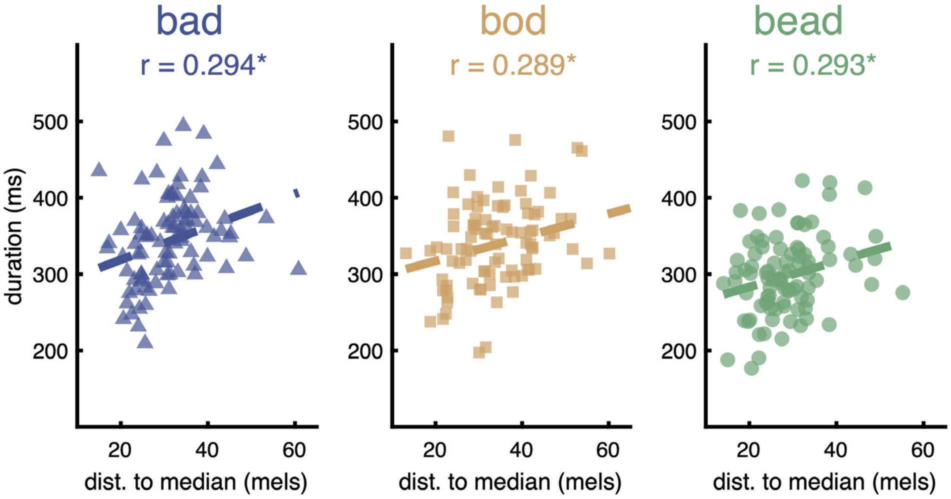

Fig. 2. Between-subject correlations between vowel duration and formant variability [distance (dist.)] during productions of isolated words. Each data point represents the average of one speaker's production of a single stimulus word: *bad* (blue triangles, left), *bod* (yellow squares, middle), and *bead* (green circles, right). Significant (*) correlations are indicated by dashed lines.

3.3 Between-subject relationship in connected speech

We further examined whether the relationship between vowel duration and average formant variability across speakers that we observed in isolated word production also held in connected speech. The results revealed a similar positive correlation between vowel duration and formant distance in sentence productions, although the statistical significance was marginal [$\beta = 0.56$, $t = 1.82$, $p = 0.077$) [Fig. 4(B)]. At the level of individual vowel contexts, significant between-subject correlations between vowel duration and formant distance were only observed in a subset of six vowel contexts (shown in color in Fig. 4), all showing the expected positive relationship (*lead* [/i/], $r = 0.347$, $p = 0.03$; *plead* [/i/], $r = 0.31$, $p = 0.049$; *tend* [/ɛ/], $r = 0.34$, $p = 0.03$; *wrote* [/oʊ/], $r = 0.38$, $p = 0.015$; *up* [/ʌ/], $r = 0.53$, $p < 0.001$; *crawled* [/ɔ/], $r = 0.33$, $p = 0.034$).

4. Discussion

In the current study, we tested both the between-subject and within-subject relationships between vowel duration and variability during productions of both isolated words and connected speech. We found a positive between-subject correlation in isolated words, suggesting individuals who are more variable tend to produce longer vowels.

Speech production, one of the most complicated motor behaviors, requires joint efforts of both feedforward and feedback control (Guenther, 2016; Houde and Nagarajan, 2011; Parrell *et al.*, 2019). Feedforward control relies on prior experience and memory to make predictions and guide speech movements, enabling rapidity in speech but lacking the ability to monitor and correct errors in speech output. In contrast, feedback control uses sensory information from ongoing speech to make real-time adjustments to the motor commands, playing a critical role in maintaining stability in the presence of noise, although it tends to be slower and less accurate when sensory feedback is delayed or unreliable. Both feedforward and feedback control are

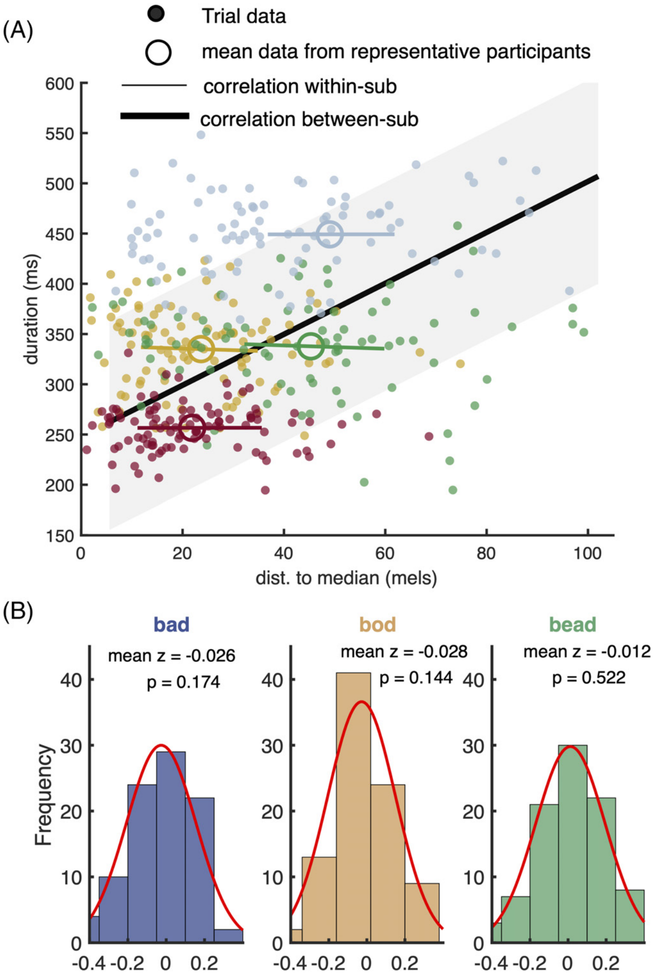

Fig. 3. Between-subject and within-subject correlations in isolated words. (A) Within-subject correlations for four representative participants and between-subject correlation across all participants. Small dots represent trial data, and colored solid lines indicate within-subject correlations across those trials. Different colors indicate different participants. The black solid line represents the significant between-subject correlation across all participants, with the shading representing the 95% confidence intervals. (B) Histogram of Fisher z-transformed within-subject correlation coefficients in different vowels during productions of isolated words.

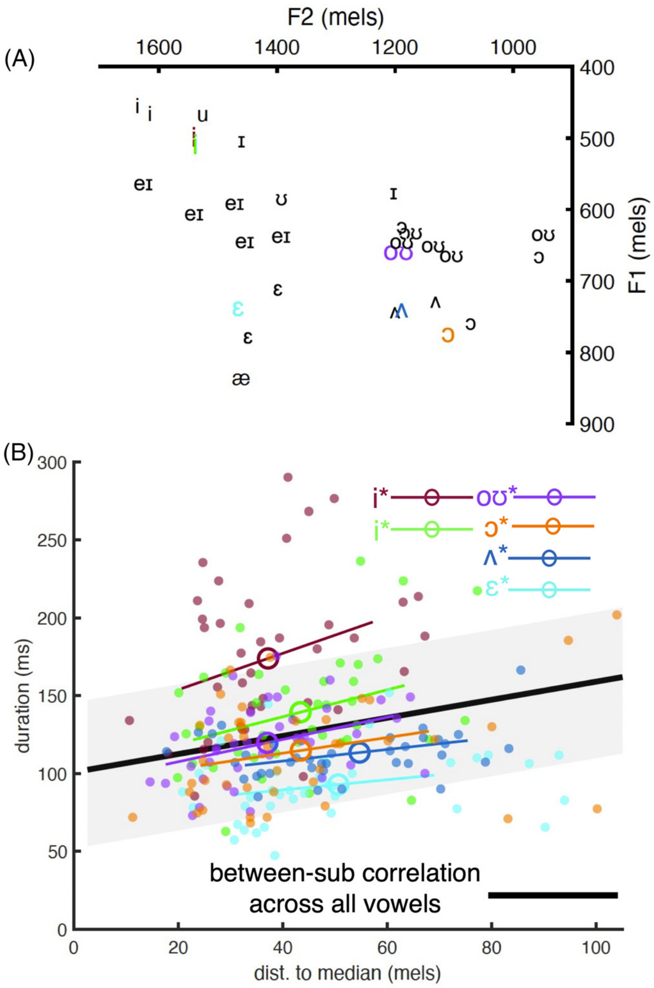

Fig. 4. Between-subject correlations in connected speech. (A) Distribution of the 30 vowel contexts included in the connected speech analysis. Colored vowels indicate vowel contexts with a significant positive between-subject correlation between vowel duration and formant variability. No significant negative correlations were found. (B) Significant between-subject correlations in different vowel contexts. Small dots represent trial data. Different colors indicate different vowel contexts. The black solid line represents the significant between-subject correlation across all participants and all vowels, with the shading representing the 95% confidence intervals.

essential for speech production, but the relative weighting of feedforward and feedback control can vary, depending on a range of factors, such as age and language proficiency (Guenther *et al.*, 2006; Liu *et al.*, 2010; Cai *et al.*, 2020). For example, as children's speech production skills develop and their speech becomes faster, it is thought that they gradually shift from being mainly reliant on feedback control to being primarily reliant on feedforward control (Guenther *et al.*, 2006; Liu *et al.*, 2010), a pattern also found in upper limb control (Malone *et al.*, 2025). A recent study demonstrated the influence of language proficiency on the relative weighting of feedforward and feedback control, showing that compared to L1 (the first language), L2 (the second language) speech production may rely more on feedback control (Cai *et al.*, 2020). The significant positive correlation between vowel duration and formant variability observed in the current study indicates that individuals who produce vowels with higher initial variability may rely more on feedback control for vowel production, with longer vowel durations allowing for more time to make online articulatory adjustments and corrections. This is consistent with findings in people with motor speech disorders: Individuals with apraxia of speech and individuals who stutter both exhibit longer vowel durations compared to fluent controls (Collins *et al.*, 1983; Riley and Ingham, 2000), potentially due to compromised feedforward control and a subsequent increased weighting of feedback control (Maas *et al.*, 2015; Max *et al.*, 2004).

However, it is important to note that the observed relationship between vowel variability and duration is correlational and does not imply causation or a specific directional effect. In other words, the current analyses cannot determine whether increased variability leads to longer duration, or vice versa. Further research is necessary to examine these potential mechanisms. Moreover, while our results suggest that individuals who produce vowels with higher initial variability may rely more on feedback-based control for vowel production, feedback control was not directly assessed in the current study. Further research is needed to directly examine this hypothesis, such as testing whether individuals with longer vowel duration exhibit increased compensatory responses to unpredictable auditory formant perturbations. Another limitation to consider is the relatively atypical speaking environment. Participants received auditory feedback—specifically, playback of their own speech mixed with speech-shaped noise—through closed-back circumaural headphones. This unusual speaking setting may have influenced the results, as speaking with noise has been shown to lead to increased vowel durations (Rostolland and Parant, 1974; Summers *et al.*, 1988); while this is unlikely to have led directly to the correlation found between variability and vowel duration, if durations were increased, this may have amplified the strength of the observed correlation.

In the current study, we further examined whether the positive between-subject relationship observed during production of isolated words also held in more ecologically valid connected speech. Compared to isolated words, connected speech imposes greater processing demands due to the need for simultaneous planning, monitoring, and correction of numerous speech sounds. Connected speech is also associated with increased variability of oral motor patterns and related acoustic signals. These factors might weaken the relationship between vowel duration and variability observed in isolated words. Our results showed that the positive relationship between vowel duration and variability held only in a subset of the vowels analyzed in our connected speech data (see Fig. 4 and see Table S1 in the supplementary material). In general, findings from study 3 should be interpreted with caution. First, the overall correlation between vowel duration and variability was weaker and only reached marginal significance during connected speech production compared to isolated words. Second, several factors that can influence vowel duration were not well-controlled. In study 3, the included words (see Table S1) differed in word length (monosyllabic vs disyllabic words) and sentence position (sentence initial vs medial). These differences could potentially affect vowel duration and help explain why significant correlations were observed only in certain vowels within specific contextual environments. For example, the expected positive correlation was observed when participants produced the vowel /i/ in the word "lead," but not when they produced the same vowel in the word "beach." Further research is needed to clarify how the relationship between vowel duration and variability may vary across vowels and context during productions of real-world connected speech.

Conversely, we did not find a significant within-subject correlation between vowel duration and formant variability. That is, speakers did not adjust their vowel duration on a production-by-production basis to account for changes in initial formant values. It is worth mentioning that our results do not suggest that the relative weighting of feedforward and feedback control is stable and cannot be changed. Previous work has demonstrated the dynamic nature of the weighting of feedforward and feedback control (Scheerer and Jones, 2014). Rather, our findings suggest that the relative weighting of feedforward and feedback control is unlikely to change on a production-by-production basis.

Supplementary Material

See the supplementary material for Table S1.

Acknowledgments

This work was supported by NIH Grant R01 DC019134, a grant awarded through the University of Wisconsin—Madison Fall Research Competition, and a core grant to the Waisman Center from the National Institute of Child Health and Human Development (P50HD105353).

Author Declarations

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

The Institutional Review Board of the University of Wisconsin—Madison approved the experimental protocols of all studies mentioned above and approved our procedures for analyzing the previously collected data (2017-1128).

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Beach, S. D., Johnson, S. A., Parrell, B., and Niziolek, C. A. (**2024**). "Increased vowel contrast and intelligibility in connected speech induced by sensorimotor adaptation," bioRxiv:2024.08.04.606537.

Beach, S. D., Tang, D., Kiran, S., and Niziolek, C. A. (**2024**). "Pars opercularis underlies efferent predictions and successful auditory feedback processing in speech: Evidence from left-hemisphere stroke," Neurobiol. Lang. **5**(2), 454–483.

Behroozmand, R., and Larson, C. R. (**2011**). "Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback," BMC Neurosci. **12**(1), 54.

Blustein, D. H., Shehata, A. W., Kuylenstierna, E. S., Englehart, K. B., and Sensinger, J. W. (**2021**). "An analytical method reduces noise bias in motor adaptation analysis," Sci. Rep. **11**(1), 9245.

Boersma, P., and Weenink, D. (**2019**). "Praat: Doing phonetics by computer (version 6.0.47) [computer program]," http://www.praat.org/ (Last viewed 2025).

Bunton, K., and Leddy, M. (**2011**). "An evaluation of articulatory working space area in vowel production of adults with Down syndrome," Clin. Ling. Phon. **25**(4), 321–334.

Cai, X., Yin, Y., and Zhang, Q. (**2020**). "A cross-language study on feedforward and feedback control of voice intensity in Chinese–English bilinguals," Appl. Psycholing. **41**(4), 771–795.

Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., and Houde, J. F. (**2013**). "Human cortical sensorimotor network underlying feedback control of vocal pitch," Proc. Natl. Acad. Sci. U.S.A. **110**(7), 2653–2658.

Chen, M. (**1970**). "Vowel length variation as a function of the voicing of the consonant environment," Phonetica **22**(3), 129–159.

Collins, M., Rosenbek, J. C., and Wertz, R. T. (**1983**). "Spectrographic analysis of vowel and word duration in apraxia of speech," J. Speech Lang. Hear. Res. **26**(2), 224–230.

Fry, D. B. (**1958**). "Experiments in the perception of stress," Lang. Speech **1**(2), 126–152.

Guenther, F. H. (**2016**). *Neural Control of Speech* (MIT Press, Cambridge, MA).

Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (**2006**). "Neural modeling and imaging of the cortical interactions underlying syllable production," Brain Lang. **96**(3), 280–301.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**(5), 3099–3111.

Houde, J. F., Gill, J. S., Agnew, Z., Kothare, H., Hickok, G., Parrell, B., Ivry, R. B., and Nagarajan, S. S. (**2019**). "Abnormally increased vocal responses to pitch feedback perturbations in patients with cerebellar degeneration," J. Acoust. Soc. Am. **145**(5), EL372–EL378.

Houde, J. F., and Jordan, M. I. (**2002**). "Sensorimotor adaptation of speech I: Compensation and adaptation," J. Speech Lang. Hear. Res. **45**(2), 295–310.

Houde, J. F., and Nagarajan, S. S. (**2011**). "Speech production as state feedback control," Front. Hum. Neurosci. **5**, 82.

IEEE (**1969**). "IEEE Recommended Practice for Speech Quality Measurements," IEEE Trans. Audio Electroacoust. Au **17**(3), 225.

Jacewicz, E., Fox, R. A., and Salmons, J. (**2007**). "Vowel duration in three American English dialects," Am. Speech **82**(4), 367–385.

Kent, R. D., Netsell, R., and Abbs, J. H. (**1979**). "Acoustic characteristics of dysarthria associated with cerebellar disease," J. Speech Lang. Hear. Res. **22**(3), 627–648.

Kim, K. S., Hinkley, L. B., Brent, K., Gaines, J. L., Pongos, A. L., Gupta, S., Dale, C. L., Nagarajan, S., and Houde, J. F. (**2025**). "Neurophysiological evidence of sensory prediction errors driving speech sensorimotor adaptation," J. Neuroscience **45**(27), e2084242025.

Koo, T. K., and Li, M. Y. (**2016**). "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," J. Chiropractic Med. **15**(2), 155–163.

Lee, S., Potamianos, A., and Narayanan, S. (**1999**). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," J. Acoust. Soc. Am. **105**(3), 1455–1468.

Liss, J. M., White, L., Mattys, S. L., Lansford, K., Lotto, A. J., Spitzer, S. M., and Caviness, J. N. (**2009**). "Quantifying speech rhythm abnormalities in the dysarthrias," J. Speech Lang. Hear. Res. **52**(5), 1334–1352.

Liu, P., Chen, Z., Larson, C. R., Huang, D., and Liu, H. (**2010**). "Auditory feedback control of voice fundamental frequency in school children," J. Acoust. Soc. Am. **128**(3), 1306–1312.

Maas, E., Mailend, M.-L., and Guenther, F. H. (**2015**). "Feedforward and feedback control in apraxia of speech: Effects of noise masking on vowel production," J. Speech Lang. Hear. Res. **58**(2), 185–200.

Malone, L. A., Hill, N. M., Tripp, H., Zipunnikov, V., Wolpert, D. M., and Bastian, A. J. (**2025**). "The control of movement gradually transitions from feedback control to feedforward adaptation throughout childhood," npj Sci. Learn. **10**(1), 13.

Max, L., Guenther, F. H., Gracco, V. L., Ghosh, S. S., and Wallace, M. E. (**2004**). "Unstable or insufficiently activated internal models and feedback-biased motor control as sources of dysfluency: A theoretical model of stuttering," Contemp. Issues Commun. Sci. Disord. **31**(Spring), 105–122.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (**2017**). "Montreal Forced Aligner: Trainable text-speech alignment using Kaldi," in *Interspeech 2017*, Stockholm, Sweden (ISCA, Stockholm, Sweden), pp. 498–502.

Niziolek, C. A., and Houde, J. (**2015**). "Wave_Viewer: First release [computer software]," Zenodo. https://doi.org/10.5281/ZENODO.13839.

Niziolek, C. A., Nagarajan, S. S., and Houde, J. F. (**2013**). "What does motor efference copy represent? Evidence from speech production," J. Neurosci. **33**(41), 16110–16116.

Parrell, B., Agnew, Z., Nagarajan, S., Houde, J., and Ivry, R. B. (**2017**). "Impaired feedforward control and enhanced feedback control of speech in patients with cerebellar degeneration," J. Neurosci. **37**(38), 9249–9258.

Parrell, B., Kim, H. E., Breska, A., Saxena, A., and Ivry, R. (**2021**). "Differential effects of cerebellar degeneration on feedforward versus feedback control across speech and reaching movements," J. Neurosci. **41**(42), 8779–8789.

Parrell, B., and Niziolek, C. A. (**2021**). "Increased speech contrast induced by sensorimotor adaptation to a nonuniform auditory perturbation," J. Neurophysiol. **125**(2), 638–647.

Parrell, B., Ramanarayanan, V., Nagarajan, S., and Houde, J. (**2019**). "The FACTS model of speech motor control: Fusing state estimation and task-based control," PLoS Comput. Biol. **15**(9), e1007321.

Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**(6), 693–703.

Riley, G. D., and Ingham, J. C. (**2000**). "Acoustic duration changes associated with two types of treatment for children who stutter," J. Speech Lang. Hear. Res. **43**(4), 965–978.

Rostolland, D., and Parant, C. (**1974**). "Physical analysis of shouted voice," in *Eighth International Congress on Acoustics*, London, UK (ICA, Madrid, Spain).

Scheerer, N. E., and Jones, J. A. (**2014**). "The predictability of frequency-altered auditory feedback changes the weighting of feedback and feedforward input for speech motor control," Eur. J. Neurosci. **40**(12), 3793–3806.

Setter, J., Stojanovik, V., van Ewijk, L., and Moreland, M. (**2007**). "The production of speech affect in children with Williams syndrome," Clin. Ling. Phon. **21**, 659–672.

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). "Effects of noise on speech production: Acoustic and perceptual analyses," J. Acoust. Soc. Am. **84**(3), 917–928.

Tang, D.-L., Parrell, B., Beach, S. D., and Niziolek, C. A. (**2025**). "The brain's sensitivity to sensory error can be modulated by altering perceived variability," J. Neurosci. **45**(5), e0024242024.

Tang, D.-L., Parrell, B., and Niziolek, C. A. (**2022**). "Movement variability can be modulated in speech production," J. Neurophysiol. **128**(6), 1469–1482.

Van Santen, J. P. H. (**1992**). "Contextual effects on vowel duration," Speech Commun. **11**(6), 513–546.