

## Research Article

# A Computational Model for Estimating the Speech Motor System's Sensitivity to Auditory Prediction Errors

Ayoub Daliri<sup>a</sup> 

**Purpose:** The speech motor system uses feedforward and feedback control mechanisms that are both reliant on prediction errors. Here, we developed a state-space model to estimate the error sensitivity of the control systems. We examined (a) whether the model accounts for the error sensitivity of the control systems and (b) whether the two systems have similar error sensitivity.

**Method:** Participants ( $N = 50$ ) completed an adaptation paradigm, in which their first and second formants were perturbed such that a participant's / $\epsilon$ / would sound like her / $\ae$ /. We measured adaptive responses to the perturbations at early (0–80 ms) and late (220–300 ms) time points relative to the onset of the perturbations. As data-driven correlates of the error sensitivity of the feedforward and feedback systems, we used the average early responses and difference responses (i.e., late minus early responses), respectively. We fitted the state-space model to participants' adaptive

responses and used the model's parameters as model-based estimates of error sensitivity.

**Results:** We found that the late responses were larger than the early responses. Additionally, the model-based estimates of error sensitivity strongly correlated with the data-driven estimates. However, the data-driven and model-based estimates of error sensitivity of the feedforward system did not correlate with those of the feedback system.

**Conclusions:** Overall, our results suggested that the dynamics of adaptive responses as well as error sensitivity of the control systems can be accurately predicted by the model. Furthermore, our results suggested that the feedforward and feedback control systems function independently.

**Supplemental Material:** <https://doi.org/10.23641/asha.14669808>

Current theories of speech production (Guenther, 2016; Houde & Nagarajan, 2011) suggest that the speech motor system uses two control mechanisms to produce speech: feedforward and feedback control systems (for an elegant discussion of these control systems, see Parrell & Houde, 2019). The feedforward control system generates motor commands to achieve the desired sensory goals. The feedback control system monitors the sensory consequences of the motor commands to ensure the accuracy of produced speech movements. While preparing motor commands, the speech motor system also predicts sensory outcomes of the motor commands (e.g., Krakauer et al., 2019). The speech motor system compares its sensory prediction with incoming sensory feedback of the speech movements to estimate potential errors in its output (i.e., *sensory prediction*

*error*). The feedforward control system uses sensory prediction errors to adjust its motor commands so that future motor commands more accurately achieve the desired sensory goals (Daliri & Dittman, 2019; Daliri, Murray, et al., 2020; Daliri et al., 2014, 2013; Guenther, 2016). The feedback control system uses sensory prediction errors in the current movement to generate corrective motor responses to more accurately achieve the current movement's sensory goals (Daliri, Chao, & Fitzgerald, 2020; Guenther, 2016; Kearney et al., 2020; Parrell & Houde, 2019). The outputs of these control systems depend on their *error sensitivity*—the weight that they assign to prediction error (Berniker & Körding, 2008; Shadmehr & Mussa-Ivaldi, 2012). Overall, these control mechanisms strongly rely on sensory prediction errors (a) to generate corrective responses and (b) to calibrate motor commands for achieving the desired sensory goals.

Given the reliance of the feedforward and feedback control mechanisms on prediction errors (Parrell & Houde, 2019), one approach for studying these control mechanisms is to experimentally generate errors using somatosensory (e.g., Tremblay et al., 2003) or auditory feedback perturbations

<sup>a</sup>College of Health Solutions, Arizona State University, Tempe

Correspondence to Ayoub Daliri: [ayoub.daliri@asu.edu](mailto:ayoub.daliri@asu.edu)

Editor-in-Chief: Bharath Chandrasekaran

Editor: Kate Bunton

Received August 17, 2020

Revision received January 4, 2021

Accepted February 22, 2021

[https://doi.org/10.1044/2021\\_JSLHR-20-00484](https://doi.org/10.1044/2021_JSLHR-20-00484)

**Disclosure:** The author has declared that no competing interests existed at the time of publication.

(e.g., Houde & Jordan, 1998). In auditory perturbation paradigms (for a review, see Fuchs et al., 2019), a participant's speech is recorded via a microphone; the acoustic characteristics of the recorded speech (e.g., formants) are perturbed; then, the modified speech is played back to the participant via headphones (in near real-time). For example, as a participant plans to generate speech movements for the word "head," she also predicts to hear "head" through the headphones. However, when she produces the word, formant frequencies of her speech are experimentally perturbed to synthesize a word that sounds like "had," which is played back to her through the headphones. In this case, the participant experiences a mismatch between what she predicted to hear (i.e., "head") and what she hears (i.e., "had"). As mentioned above, participants use this prediction error in two ways (e.g., Guenther, 2016). First, participants may rely on the prediction error to generate a within-trial corrective motor response that starts within 100–200 ms from the onset of the perturbation. Corrective responses are typically interpreted as the contributions of the feedback control system (e.g., Daliri, Chao, & Fitzgerald, 2020). Second, participants may rely on the prediction error to modify their feedforward motor commands (i.e., adapt) to reduce prediction errors in future productions. The adaptive responses are typically interpreted as the contributions of the feedforward control system (e.g., Kearney et al., 2020). To measure online corrective responses, studies have applied perturbations on a set of randomly selected trials that are surrounded by several unperturbed trials—commonly known as the *compensation paradigm* (Daliri, Chao, & Fitzgerald, 2020; Niziolek & Guenther, 2013; Parrell et al., 2017; Tourville et al., 2008). The rationale for this approach is to minimize the adaptive changes in speech due to the exposure to perturbations (i.e., minimizing the contributions of the feedforward control system). To measure adaptive responses, studies have applied perturbations on several subsequent trials—commonly known as the *adaptation paradigm* (Abur et al., 2018; Ballard et al., 2018; Daliri et al., 2017; Daliri & Dittman, 2019; Daliri & Max, 2018; Houde & Jordan, 1998; Lester-Smith et al., 2020; Stepp et al., 2017). Together, the adaptation and compensation paradigms are powerful paradigms that can elucidate the contributions of the feedforward and feedback control mechanisms.

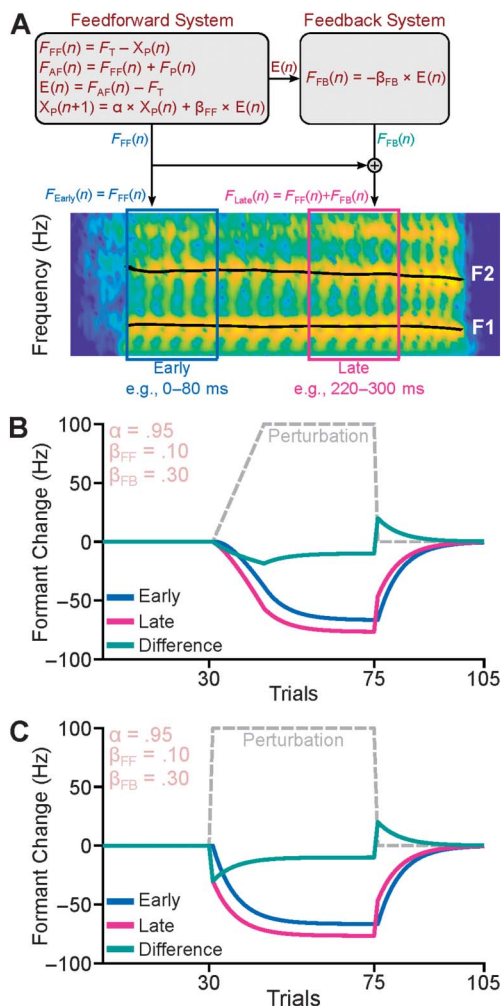
Although both the feedforward and feedback control systems rely on prediction errors, it is not clear (a) whether the two systems have similar error sensitivity and (b) whether the outputs of the two systems are related or function independently. Previous studies have used the adaptation and compensation paradigms to separately estimate the error sensitivity of feedforward and feedback systems (Franken et al., 2019; Hawco & Jones, 2009; Lester-Smith et al., 2020; Parrell et al., 2017; Scheerer & Jones, 2018). While many studies have used more typical adaptation and compensation paradigms, Franken et al. (2019) developed a new paradigm in which they systematically examined the effects of frequently versus infrequently applied perturbations. Overall, these studies have measured adaptive and corrective responses to perturbations as *data-driven correlates* of error

sensitivity of the feedforward and feedback control systems, respectively. Note that the magnitudes of adaptive and corrective responses depend on both the error sensitivity of the control systems and the magnitude of the error itself (i.e., response = error sensitivity  $\times$  error). However, several limitations of these paradigms make it difficult to determine the error sensitivity of the control mechanisms accurately. First, many adaptation studies (Fuchs et al., 2019), including our previous adaptation studies (Daliri & Dittman, 2019; Daliri & Max, 2018; Daliri et al., 2017), have examined formant changes measured at the midpoint of a vowel (e.g., the mid 20% of the vowel) as evidence for adaptive changes in the feedforward control system. However, if the midpoint is after  $\sim 150$  ms, it is most likely that the feedback control system has received the auditory feedback regarding the perturbations and has generated a corrective response (Daliri, Chao, & Fitzgerald, 2020; Kearney et al., 2020; Tourville et al., 2008). Thus, formant changes measured at the midpoint are not necessarily reflective of the contributions of the feedforward control system alone. Second, in the compensation paradigm, perturbed trials are randomly distributed to minimize potential effects of the feedforward control system; however, exposure to perturbations may lead to small—and likely unmeasurable (Daliri, Chao, & Fitzgerald, 2020; Parrell et al., 2017)—changes in the feedforward control system. Thus, the corrective responses measured in the compensation paradigm are not necessarily reflective of the contributions of the feedback control system alone. Third, measuring the feedforward and feedback control systems separately and in isolation discounts the potential bidirectional interactions between the two systems (Guenther, 2016). Finally, the adaptation and compensation paradigms are relatively time-consuming and require many trials; this limitation poses a practical challenge, especially for examining the control systems in patient populations and children. Therefore, the primary goal of this study is to develop a procedure to more efficiently and accurately examine the feedforward and feedback mechanisms and to estimate their sensitivity to prediction errors in auditory perturbation paradigms.

Although adaptive and corrective responses are commonly used as data-driven correlates of error sensitivity of the control systems, these measures are limited in that they are strongly dependent on the analysis choice. For example, adaptive responses are often defined as the average responses over several trials after participants have experienced the perturbations, and their responses have become stable (e.g., Daliri & Dittman, 2019). However, the averaged adaptive response depends on the number of trials and the location of the trials relative to the beginning of the introduction of the perturbations. Thus, there are large differences in adaptive responses across studies (for a review, see Fuchs et al., 2019). One approach to minimize this limitation is to use computational models that calculate error sensitivity based on the participant's adaptive responses in all trials (and not just a few trials). In a recent study (Daliri & Dittman, 2019), we adopted a computational model (state-space model), commonly used in the

limb motor control studies (Shadmehr & Mussa-Ivaldi, 2012), to estimate error sensitivity of the feedforward control system in an adaptation paradigm. In this study, we further developed the state-space model to be able to estimate the error sensitivity of both the feedforward and feedback control systems in an adaptation paradigm (see Figure 1A). Given that the speech output includes the contributions of both control systems, the model includes feedforward and feedback components that use error sensitivity (model parameters) and prediction errors

**Figure 1.** We developed a new state-space model to estimate the error sensitivity of the feedforward and feedback control systems (A). Formants at early time points ( $F_{\text{Early}}$ ) were used to estimate the feedforward control system's contributions. Formants at late time points ( $F_{\text{Late}}$ ) were used to estimate the contributions of the combined feedforward and feedback control systems. The difference responses (the difference between late and early responses) were used to estimate the contributions of the feedback control system alone. We used the model to predict (based on a set of arbitrary parameters) change in the first formant at early and late time points in response to a perturbation of the first formant (100 Hz) in two adaptation paradigms, in which the perturbation is gradually introduced (B) or suddenly introduced (C). F1 = first formant; F2 = second formant.



to generate their outputs. The model's parameters of error sensitivity (feedforward and feedback error sensitivity;  $\beta_{FF}$  and  $\beta_{FB}$ ) can be estimated by fitting the model to participants' responses. Although both feedforward and feedback control systems contribute to speech production, the relative contributions of the two systems are not equal throughout a production (Guenther, 2016). The contributions of the feedback control system during the early time points of a given speech token is minimal that is because the feedback control system needs to wait until the auditory feedback of the produced speech becomes available to calculate prediction error and generate corrective responses accordingly (Guenther, 2016; Kearney et al., 2020; Parrell & Houde, 2019). Therefore, responses at early time points ( $F_{\text{Early}}$  in Figure 1A) can be used to estimate the contributions of the feedforward control system, and responses at late time points (time points later than  $\sim 150$  ms) can be used to estimate the contributions of the combined feedforward and feedback control systems ( $F_{\text{Late}}$  in Figure 1A). In other words, the difference between late and early responses can be used as an estimate of the feedback control system. Note that the difference between late and early responses can also be influenced by the articulatory constraints of the next consonant. However, we can minimize this effect by adjusting the early and late responses based on their values during unperturbed trials—when there are no perturbation-induced errors and only articulatory constraints exist. Another important advantage of computational models—in addition to providing *model-based error sensitivity*—is that the model can predict early and late responses in auditory perturbation experiments before conducting the experiments. For example, Figure 1 shows the simulated (based on an arbitrary set of model parameters) early and late formant responses (e.g., the first formant) in an adaptation paradigm, in which the formant perturbation is gradually introduced (see Figure 1B; this is the most common adaptation paradigm) or suddenly introduced (see Figure 1C; this adaptation paradigm is less common). These two simulated responses suggest that the difference between late and early responses has a sudden change when the perturbation is introduced suddenly, whereas this change gradually develops when the perturbation is introduced gradually. Overall, by using this model and examining early and late responses in one auditory perturbation task (rather than conducting separate adaptation and compensation paradigms), we can (a) more efficiently and accurately determine (model-based) error sensitivity of both the feedforward and feedback control systems, and (b) predict early and late responses to auditory feedback perturbations and generate model-driven hypotheses.

In this study, we conducted an auditory perturbation experiment to examine whether the state-space model's parameters can be used as estimates of the error sensitivity of the feedforward and feedback control systems. Participants completed an adaptation paradigm in which formants of their /e/ were shifted toward their /æ/ (an increase in the first formant, and a decrease in the second formant). Our preliminary simulations (see Figures 1B and 1C) suggested

that the difference between late and early responses would be more evident and perhaps more easily measured if the perturbations are introduced suddenly. Thus, similar to our previous studies (Daliri & Max, 2018; Kim et al., 2020), we used an adaptation paradigm in which the perturbations were introduced suddenly. We measured participants' early adaptive responses (within 0–80 ms) and late adaptive responses (within 220–300 ms). We hypothesized that if the feedforward and feedback mechanisms contribute to the late responses, and only the feedforward mechanism contributes to early responses, then the late responses would be different from early responses. Additionally, to test whether the model accounts for the error sensitivity of the control systems, we examined the relationship between the model-based and data-driven estimates of error sensitivity of the control systems. We used feedforward and feedback error sensitivity ( $\beta_{FF}$  and  $\beta_{FB}$ ) as model-based estimates of error sensitivity. As data-driven estimates of error sensitivity of the feedforward and feedback systems, we used average early adaptive responses and average difference response (the difference between late and early responses), respectively. Because both feedforward and feedback control systems rely on prediction errors, we also examined the relationship between error sensitivity in the feedforward and feedback control systems.

## Method

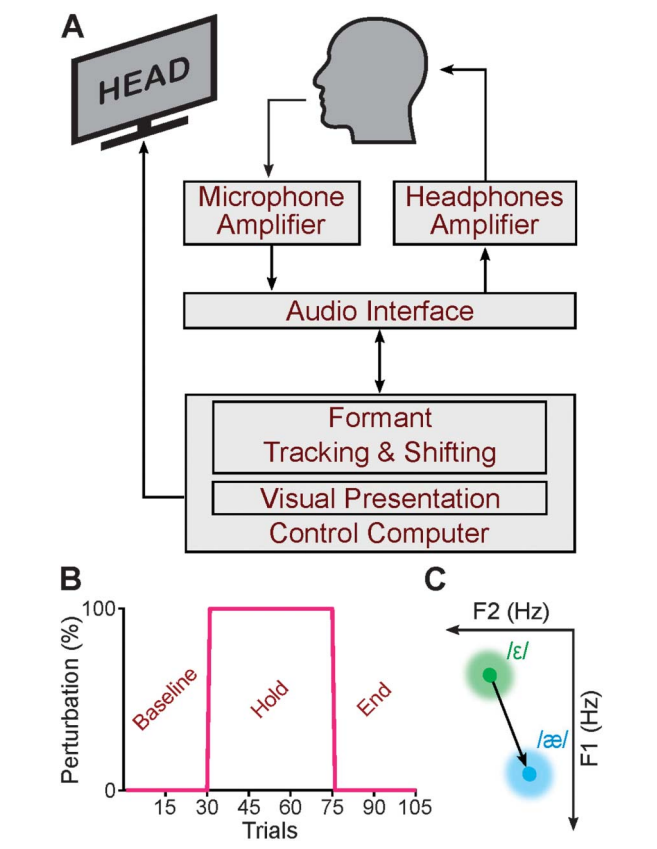
### Participants

Fifty right-handed adults were recruited to participate in this study (10 men; age range: 18–48 years,  $M = 23.51$  years,  $SD = 5.01$  years). We used the following inclusion criteria to recruit participants: (a) self-reported absence of neurological, psychological, or speech-language disorders; (b) being a native speaker of American English; and (c) having a binaural pure-tone hearing threshold less than 20 dB HL at all octave frequencies of 250–8000 Hz (American Speech-Language-Hearing Association, 1997). Before the experimental session, participants signed a written consent form. The institutional review board of Arizona State University approved all study protocols.

### Apparatus

Figure 2A shows the apparatus of the experiment. Participants were seated in front of a computer monitor inside a double-walled, sound-attenuating booth. To record speech signals, a microphone (SM58, Shure) was placed ~15 cm away from the corner of the participant's mouth at a ~45° angle. The speech signal was amplified (TubeOpto 8, ART) and passed through an external audio interface (8pre, MOTU). The audio interface digitized the signal, transmitted the signal to a computer, and transmitted it back to the audio interface. The output of the audio interface was amplified (S.phone, Samson Technologies Corp.) and binaurally played back to the participants via insert earphones (ER-1, Etymotic Research Inc.). Before each experimental session, we calibrated the amplification levels of the microphone and

**Figure 2.** Schematic of the apparatus for applying formant perturbations (A). Participants completed an adaptation paradigm consisting of three phases: baseline, hold, and end (B). In the baseline and end phases, participants received normal (unperturbed) auditory feedback. In the hold phase, participants received perturbed auditory feedback. The perturbation's magnitude and direction were designed using each participant's  $\epsilon$ - $\text{æ}$  distance and angle (C). The perturbation shifted a participant's / $\epsilon$ / toward her / $\text{æ}$ / (increase in F1 and decrease in F2). F1 = first formant; F2 = second formant.



the earphones amplifiers to ensure that the intensity of the signal at the insert earphones was 5 dB higher than the intensity of the microphone signal (Abur et al., 2018; Daliri & Max, 2015a; Max & Daliri, 2019; Merrikhi et al., 2018).

To apply auditory perturbation, we used Audapter (Cai, 2015), publicly available software for near real-time formant tracking and shifting. The exact input–output delay depends on several factors, such as the audio interface hardware and signal processing routines involved in the formant tracking and shifting (Kim et al., 2019). We used a digital audio recorder (Tascam DR-680MKII) to simultaneously record the input of the audio interface (microphone) and the output of the audio interface (manipulated auditory feedback; insert earphones) on two separate channels. Our analysis showed that the input–output delay was ~16.4 ms. We also used a 2-cc coupler (Type 4946, Bruel & Kjaer Inc.) connected to a sound level meter (Type 2250A, Bruel & Kjaer Inc.) to measure the input–output delay of the insert earphones (Daliri & Max, 2015a, 2015b, 2016). Our measurements

showed that the insert earphones introduced less than 1-ms delay. Overall, the total input–output delay was ~17.4 ms. Audapter uses linear predictive coding (LPC) analysis in combination with dynamic programming and a set of heuristic rules to track formants. We used an LPC order of 15 for female participants and an LPC order of 17 for male participants. We also used a sampling frequency of 48 kHz, a downsampling factor of 3, and a buffer size of 96 samples.

### Procedure

This study was conducted in one session that lasted less than 1 hr. All 50 participants completed a training task, a pretest task, and an adaptation task.

### Training Task

To ensure that participants are accustomed to the experimental setup, they completed a training task that consisted of 30 trials. In each trial, a consonant–vowel–consonant word containing / $\epsilon$ / (“hep,” “head,” and “heck”) appeared in black font on a gray background. The target word remained on the monitor for 2.5 s, and there was a short break (1–1.5 s) between two consecutive trials. The training task consisted of 10 repetitions of each of the words in random order. In this task, participants were trained to produce words with the duration and intensity within the desired ranges (400–600 ms; 70–80 dB SPL). For this purpose, after the completion of each trial, they were given visual feedback regarding their duration and intensity. Eight participants struggled with maintaining their productions within the desired ranges, and therefore, we repeated the training task for these participants.

### Pretest Task

The goal of this task was to find the centroids of / $\epsilon$ / and / $\text{\ae}$ / for each participant. For this purpose, participants completed 50 trials of a word reading task—25 repetitions of each vowel in the context of /hVp/. The design of each trial was similar to the design of the training trials, except that visual feedback regarding the intensity and duration of the productions was presented to participants only if their intensity or duration was outside the desired ranges. We used a custom-written MATLAB script to automatically extract the average of the first formant (F1) and the second formant (F2) for each production. Note that the formants were calculated by Audapter. We used the calculated formants to estimate the vowel centroids—the center of a vowel distribution in the F1–F2 coordinates. The vowel distributions and centroids were visually inspected to ensure that the centroids had been accurately calculated. We used the centroids to calculate (a)  $\epsilon$ – $\text{\ae}$  Euclidean distance (in Hz) in the F1–F2 coordinates, and (b) the angle between centroids. The  $\epsilon$ – $\text{\ae}$  distance and angle were used to define participant-specific formant perturbations in the adaptation task. Additionally, to increase the accuracy of the Audapter’s formant tracking algorithm, we used the F1 and F2 of the

centroid of / $\epsilon$ / as participant-specific initial values for formant tracking in the adaptation task.

### Adaptation Task

The adaptation task consisted of 105 trials. In each trial, a consonant–vowel–consonant word containing / $\epsilon$ / (“hep,” “head,” and “heck”) was presented on the monitor. The order of the words was randomized. As shown in Figure 2B, this task consisted of a baseline phase (30 trials), a hold phase (45 trials), and an end phase (30 trials). In the baseline phase, participants received normal (unperturbed) auditory feedback. In the hold phase, participants received perturbed auditory feedback. The perturbation was designed using each participant’s  $\epsilon$ – $\text{\ae}$  distance and angle (calculated in the pretest task). The perturbation was designed to shift a given participant’s / $\epsilon$ / toward her / $\text{\ae}$ / by increasing F1 and decreasing F2 (see Figure 2C). In other words, the perturbation magnitude was equal to the Euclidean distance between the participant’s centroids of / $\epsilon$ / and / $\text{\ae}$ / in the F1–F2 coordinates. The average perturbation magnitude was 269.5 Hz ( $SD = 116.9$  Hz) across all participants. Note that both formant tracking and formant perturbation were designed to initiate at the onset of the vowel. In the end phase, participants received normal, unperturbed auditory feedback.

### Data Analysis

We used a MATLAB script to inspect the accuracy of the Audapter’s formant tracking by displaying formants on the spectrogram of each production. Additionally, all productions were inspected to exclude trials with production errors (e.g., mispronunciations). Using the spectrogram and time-domain waveform, we manually selected the onset and offset of the vowel for each production. We then extracted F1 and F2 trajectories between the onset time and the offset time. As mentioned earlier, the responses at time points immediately after the onset of a perturbation are mostly influenced by the feedforward control mechanisms, and responses at time points later than ~150 ms from the onset of the perturbation are influenced by both the feedforward and feedback control mechanisms. For each of the F1 and F2, we averaged formant values within three time windows: early, mid, and late. The early time window was placed on the first 80 ms of the vowel, the mid time window was placed at 110–190 ms, and the late time window was placed at 220–300 ms. The early and late time windows were most suited to test our hypotheses; however, for consistency purposes, we included the results for the mid time window in Supplemental Material S1. The rationale for leaving 30 ms between two subsequent time windows was to ensure that the data in the time windows do not overlap. Although participants were instructed to produce vowels with a duration longer than 400 ms (see the Training Task), several participants had vowel durations shorter than 400 ms. Therefore, we focused our analysis on the first 300 ms and excluded all trials that had a vowel duration of less than 300 ms. Approximately 7.83% ( $SD = 4.67\%$ ) of all

trials of the adaptation task were excluded. In the adaptation task, perturbations were parallel to a line connecting each participant's /ε/ to /æ/ (i.e., perturbation line). Similar to our previous studies (Daliri, Chao, & Fitzgerald, 2020; Daliri & Dittman, 2019), we projected the extracted F1 and F2 in each trial to the participant-specific perturbation line (adaptation response) and to a line orthogonal to the perturbation line (deviation response). Both adaptation response and deviation response were calculated relative to the participant-specific centroid of /ε/ (reference point). Positive adaptation responses were in the direction toward /æ/, and positive deviation responses were in the direction toward the outside of the vowel space.

Because the target words were in the context of /hεC/ (“hep,” “head,” and “heck”), the three target words had different formant transitions at the end of the vowel (due to articulatory constraints of the consonant). Given the primary hypotheses of the study, we were interested in the difference between formant values in the late window and early window. Thus, to minimize the potential effects of formant transition on the adaptation and deviation responses, we corrected both the early and late responses based on their average responses in the baseline phase on a word-specific basis. Lastly, because perturbation magnitudes were participant specific, to be able to compare responses across participants, we normalized responses by dividing them by the participant-specific perturbation magnitude (i.e., participant-specific ε-æ distance). We used the adaptation response as the dependent variable for statistical analysis. Although adaptation responses were more appropriate for testing our hypotheses, we also examined deviation responses for consistency purposes (see Supplemental Material S2).

### Computational Modeling

In our previous study (Daliri & Dittman, 2019), we developed a state-space model to estimate the contributions of the feedforward control system in an adaptation paradigm. Our model was based on state-space models that have been used in studies of limb motor learning (Shadmehr & Mussa-Ivaldi, 2012). The model assumes that, in a given trial ( $n$ ), we produce a feedforward motor command ( $F_{FF}$ ) to achieve a specific auditory target ( $F_T$ ) based on our estimate of potential perturbations ( $X_P$ ) in the current trial. In other words,  $X_P$  is our current prediction regarding the magnitude of the perturbation before we generate a feedforward motor command.

$$F_{FF}(n) = F_T - X_P(n). \quad (1)$$

$$F_{AF}(n) = F_{FF}(n) + F_P(n). \quad (2)$$

After the execution of the feedforward motor command, we receive the auditory feedback ( $F_{AF}$ ) that consists of the auditory consequences of the motor commands and formant perturbations ( $F_P$ ). We then calculate the prediction error in the current trial by comparing the auditory

feedback with the auditory target, which can be further simplified using Equations 1 and 2.

$$E(n) = F_{AF}(n) - F_T = F_P(n) - X_P(n). \quad (3)$$

Based on this simplification, the prediction error is the difference between the perturbation in the current trial and our current estimate of the perturbation. After calculating the prediction error, we update our estimate of the perturbation using a weighted sum of our current estimate of the perturbation (i.e., current prediction) and the prediction error.

$$X_P(n+1) = \alpha \times X_P(n) + \beta_{FF} \times E(n); \quad (4) \\ 0 \leq \alpha \leq 1, 0 \leq \beta_{FF} \leq 1.$$

The weight ( $\alpha$ ) that we assign to our current prediction determines the similarity between our prediction of the perturbation in the next trial and the current perturbation. This parameter is often called the “decay factor” or “forgetting factor” (Daliri & Dittman, 2019; Shadmehr & Mussa-Ivaldi, 2012). Here, we use the term *prediction sensitivity*, as this term more accurately reflects the function of the parameter. The weight that we assign to the prediction errors ( $\beta_{FF}$ ) corresponds to our sensitivity to prediction error—a large  $\beta_{FF}$  means a high error sensitivity. Stated differently, we use the weighted prediction error and the weighted estimate of the perturbation to update our estimate of the perturbation that will be used to develop a feedforward motor command in the next trial; thus,  $\beta_{FF}$  corresponds to the error sensitivity of the feedforward control system. Because prediction errors are also used in the feedback control system (Guenther, 2016; Kearney et al., 2020; Parrell & Houde, 2019), we developed the state-space model to include a feedback component ( $F_{FB}$ ).

$$F_{FB}(n) = -\beta_{FB} \times E(n); 0 \leq \beta_{FB} \leq 1. \quad (5)$$

We developed our online corrective responses to perturbations based on the prediction error and the sensitivity of the feedback control system to prediction errors (i.e., feedback error sensitivity;  $\beta_{FB}$ ). Because the contribution of the feedback control system is minimal during the early time points of a given speech token, we used responses at early time points ( $F_{Early}$ ) as an estimate of the contribution of the feedforward control system. However, responses at time points later than ~150 ms ( $F_{Late}$ ) are influenced by both the feedforward and feedback control systems (Guenther, 2016; Kearney et al., 2020; Parrell & Houde, 2019), and thus, the difference between the late and early responses can be used as an estimate of the feedback control system.

$$F_{Early}(n) = F_{FF}(n). \quad (6)$$

$$F_{Late}(n) = F_{FF}(n) + F_{FB}(n). \quad (7)$$

We used the “fmincon” function of MATLAB to fit the model to each participant’s adaptive responses. The “fmincon” function is based on a gradient-based optimization

method that finds the minimum value of a constrained nonlinear multivariable cost function. The computational model included three parameters ( $\alpha$ ,  $\beta_{FF}$ , and  $\beta_{FB}$ ) with a range of 0–1. The optimization started with a set of random values for the parameters; we calculated simulated early and late responses based on the parameter values; then, we calculated the root-mean-square of the difference between the participant's early and late adaptive responses and the simulated early and late responses. The “fmincon” function used this error to iteratively find a set of parameters that minimizes the difference between measured and simulated responses. It should be noted that we supplied the function with the lower bound of 0.001 and the upper bound of 0.999. This was because we needed to logit-transform the parameters for statistical analyses, and the logit transform cannot be calculated for 0 and 1. Because we assumed the model's parameters ( $\alpha$ ,  $\beta_{FF}$ , and  $\beta_{FB}$ ) are not influenced by the consonantal environment, we used adaptive responses for all trials (words) to fit the model. We used a procedure similar to our previous study (Daliri & Dittman, 2019) for the optimization (10,000 iterations; optimality tolerance of  $10^{-6}$ ). For each participant, we (a) extracted the optimized parameters ( $\alpha$ ,  $\beta_{FF}$ , and  $\beta_{FB}$ ), (b) calculated the simulated early and late responses based on the optimized parameters, and (c) calculated the goodness of fit for the optimized model (using r-squared).

### Statistical Analysis

We used R Version 3.5.1 (R Core Team, 2018) to conduct the statistical analyses. We entered the adaptation responses that were calculated based on the early and late time windows into the statistical analyses. For each of the early and late responses and for each word, we calculated the average responses in the *baseline* phase (30 trials; 10 per word), the last 30 trials of the *hold* phase (10 per word), and the *end* phase (30 trials; 10 per word), as dependent variables (see gray-shaded areas in Figure 3A). To examine the adaptation responses, we used a linear mixed-effect model implemented in the lme4 package. We used phase (baseline, hold, and end), time window (early and late), and word (/hep/, /hed/, and /hek/) as fixed factors and participant as a random factor (random intercept). To evaluate the statistical significance of the main effects and interactions, we used the lmerTest package (Kuznetsova et al., 2017) with Satterthwaite's method for approximating the degrees of freedom. We also conducted post hoc analyses (i.e., pairwise comparisons) using the emmeans package (Lenth, 2019). We used Tukey's method to correct for multiple comparisons with Kenward-Roger for approximating the degrees of freedom. We used the Psych package (Revelle, 2018) to examine the relationship between the model-based estimates ( $\beta_{FF}$  and  $\beta_{FB}$ ) and data-driven estimates (average early and difference responses in the hold phase) of error sensitivity of the control systems. We also examined the relationship between the error sensitivity of the feedforward and feedback control systems using Pearson correlation coefficients.

## Results

### Data-Driven Estimates of Error Sensitivity

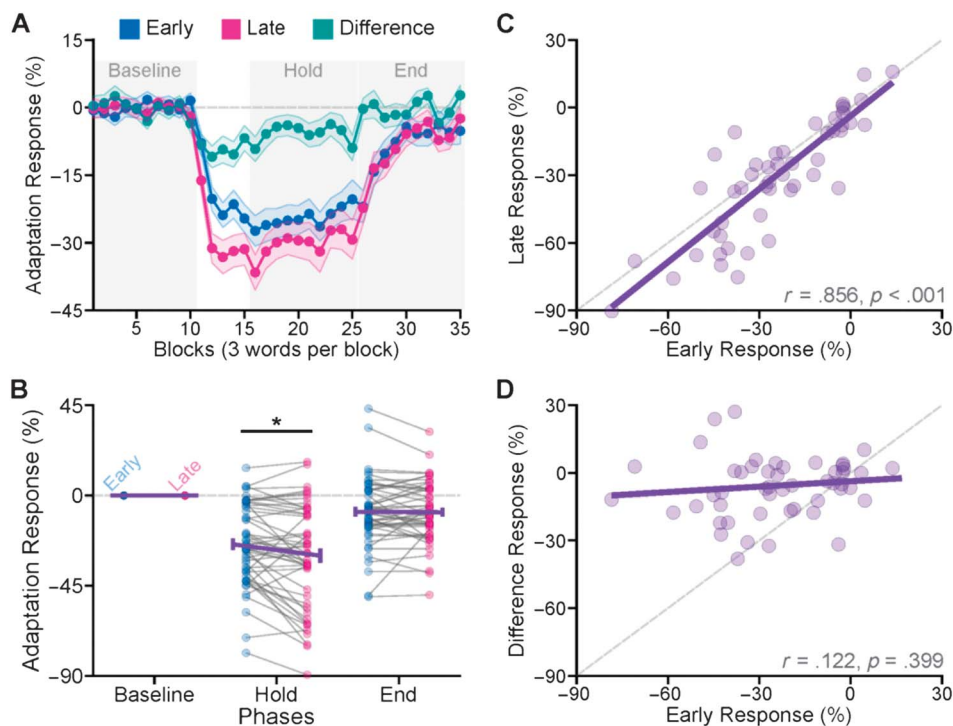
Figure 3A shows the group-average trajectories of the adaptation responses calculated based on the early (0–80 ms) and late (220–300 ms) time windows. The results for the mid-time window (110–190 ms) are shown in the Supplemental Material S1. As mentioned in the introduction, the early responses were influenced by the feedforward control system alone, late responses were influenced by both the feedforward and feedback control systems, and, thus, the difference responses (the difference between late and early responses) were influenced by the feedback control system alone. For each participant's early and late adaptation responses, we calculated average responses in three phases (baseline, hold, and end) indicated by the gray-shaded areas in Figure 3A. Examining adaptation responses, we found statistically significant main effects of phase,  $F(2,833) = 316.629$ ,  $p < .001$ , and time window,  $F(1,833) = 4.698$ ,  $p = .030$ . We did not find statistically significant main effect of word,  $F(2,833) = 0.642$ ,  $p = .526$ , Phase  $\times$  Word interaction,  $F(4,833) = 0.338$ ,  $p = .853$ , Time Window  $\times$  Word interaction,  $F(2,833) = 0.911$ ,  $p = .403$ , and Phase  $\times$  Time Window  $\times$  Word interaction,  $F(4,833) = 0.586$ ,  $p = .673$ . The Phase  $\times$  Time Window interaction was statistically significant,  $F(2,833) = 4.207$ ,  $p = .015$ . As shown in Figure 3B, this interaction indicated that late responses were larger than early responses in the hold phase ( $p = .004$ ) but not in the baseline phase ( $p = .999$ ) and the end phase ( $p = .999$ ). It should be noted that, out of all 50 participants, 30 participants showed this pattern in the hold phase (i.e., larger late responses than early responses).

Using responses in the hold phase, we examined the relationship between the adaptation responses. As shown in Figure 3C, early responses statistically significantly correlated with late responses,  $r = .856$ ,  $p < .001$ ; however, as shown in Figure 3D, there was not a significant relationship between early and difference responses,  $r = .122$ ,  $p = .399$ .

### Model-Based Estimates of Error Sensitivity

To estimate the error sensitivity of the feedforward and feedback control systems, we fitted the state-space model (see Computational Modeling section) to the early and late responses of each participant. Figure 4A shows the group-average modeled trajectories for early, late, and difference responses. Our analyses of the goodness of fit of the models showed that, on average, the model explained 58.9% ( $SD = 21.5\%$ ;  $Mdn = 62.1\%$ ) of the variance of the adaptation data ( $r$ -squared ranged from .125 to .917). Based on the fitted model for each participant, we extracted three parameters: prediction sensitivity ( $\alpha$ ), feedforward error sensitivity ( $\beta_{FF}$ ), and feedback error sensitivity ( $\beta_{FB}$ ). Because the model's parameters were bounded between 0 and 1, we transformed the parameters to normalize their distributions for statistical analyses, using a logit transform ( $\text{logit}(\alpha) =$

**Figure 3.** (A) The group-average trajectories for early, late, and difference responses (the difference between early and late adaptation responses). For each participant's early and late responses, we calculated average responses in three phases (baseline, hold, and end) indicated by the gray-shaded areas in A. (B) Average-group and individual data for early versus late responses in all three phases. The late responses were larger than the early responses in the hold phase. The difference trajectories in A showed a sudden negative change after the start of the perturbation (Block 11) and a sudden positive change after the perturbation was removed (Block 26). We used average responses in the hold phase to examine the relationship between the adaptation responses. (C) Early responses statistically significantly correlated with late responses; however, there was not a significant relationship between early and difference responses (D). Error bars correspond to  $\pm 1$  standard error, and asterisks correspond to  $p < .01$ .



$\log(\alpha / (1-\alpha))$ ; e.g.,  $\text{logit}(0.5) = 0$ ). The average of nontransformed prediction sensitivity was 0.733, and it ranged from 0.028 to 0.999. The average nontransformed  $\beta_{FF}$  was 0.058 (range: 0.002–0.372) and the average nontransformed  $\beta_{FB}$  was 0.141 (range: 0.001–0.550). The transformed  $\beta_{FF}$  was statistically smaller than the transformed  $\beta_{FB}$ ,  $t(49) = -2.187$ ,  $p = .033$ . As shown in Figures 4B–4D, we examined the relationship between the three parameters of the model. Prediction sensitivity statistically significantly correlated with the feedforward error sensitivity ( $r = -.434$ ,  $p = .002$ ) and feedback error sensitivity ( $r = .429$ ,  $p = .002$ ). However, feedforward error sensitivity did not correlate with feedback error sensitivity ( $r = -.065$ ,  $p = .656$ ).

### The Relationship Between Model-Based and Data-Driven Estimates of Error Sensitivity

We used the average early adaptive responses (in the last 30 trials of the hold phase) as a data-driven estimate of the error sensitivity of the feedforward control system, and the average difference responses (in the last 30 trials of the hold phase) as an estimate of the error sensitivity of the feedback control system. We used the

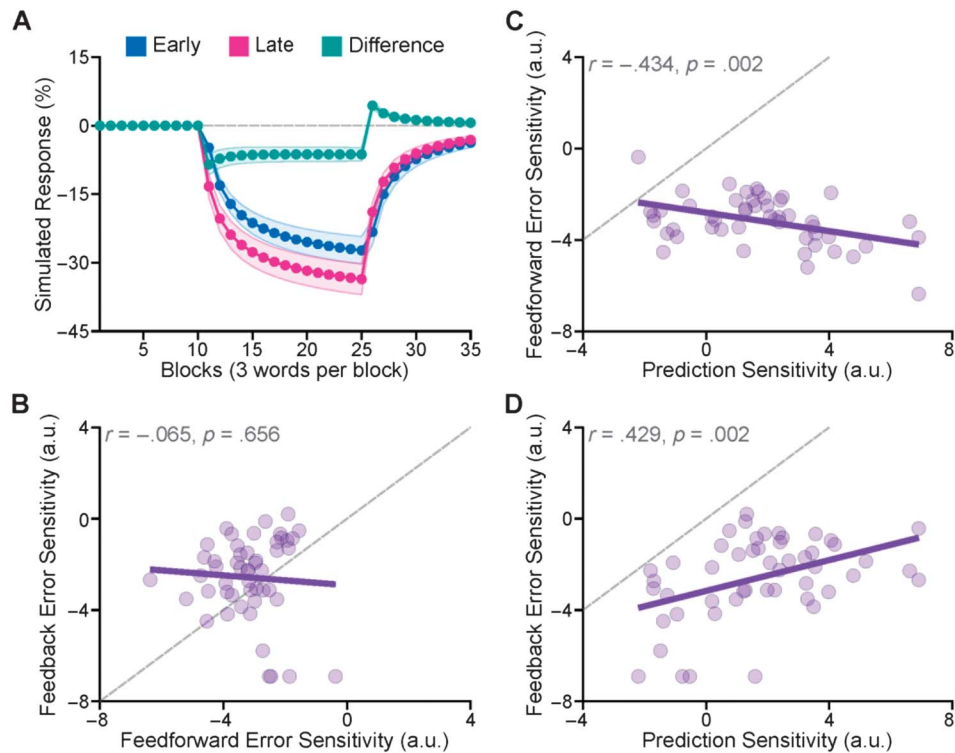
feedforward error sensitivity ( $\beta_{FF}$ ) and feedback error sensitivity ( $\beta_{FB}$ ) as model-driven estimates of error sensitivity for the participant. Our correlational analyses indicated that (a) transformed feedforward error sensitivity significantly correlated with early responses,  $r = -.437$ ,  $p = .002$  (see Figure 5A), (b) transformed feedback error sensitivity significantly correlated with the late responses,  $r = -.587$ ,  $p < .001$  (see Figure 5B), and with the difference responses,  $r = -.479$ ,  $p < .001$  (see Figure 5C).

## Discussion

The speech motor system uses feedforward and feedback control mechanisms that are both reliant on prediction errors (i.e., discrepancies between sensory prediction and sensory feedback). In this study, we developed a state-space model to estimate the error sensitivity of the feedforward and feedback control systems. We also conducted an auditory perturbation experiment to examine whether the state-space model's parameters can be used as estimates of the error sensitivity of the feedforward and feedback control systems. Participants completed an adaptation paradigm in which formants of their / $\epsilon$ / were shifted toward their / $\text{æ}$ /



**Figure 4.** We fitted the state-space model to each participant's adaptation responses. A shows the group-average trajectories of the simulated responses (early, late, and difference). Shaded areas in A correspond to  $\pm 1$  standard error. On average, the model explained  $\sim 60\%$  of the variance of the individual adaptation data ( $r$ -squared ranged from .125 to .917). Using the fitted models, we extracted three parameters for each participant: prediction sensitivity ( $\alpha$ ), feedforward error sensitivity ( $\beta_{FF}$ ), and feedback error sensitivity ( $\beta_{FB}$ ). Note that we logit-transformed them to normalize their distributions because model parameters were bounded between 0 and 1. Feedforward error sensitivity did not correlate with feedback error sensitivity (B). Prediction sensitivity negatively correlated with feedforward error sensitivity (C) but positively correlated with feedback error sensitivity (D). The gray dashed line in B, C, and D is the identity line (a line with the slope of 1). a.u. = arbitrary unit.

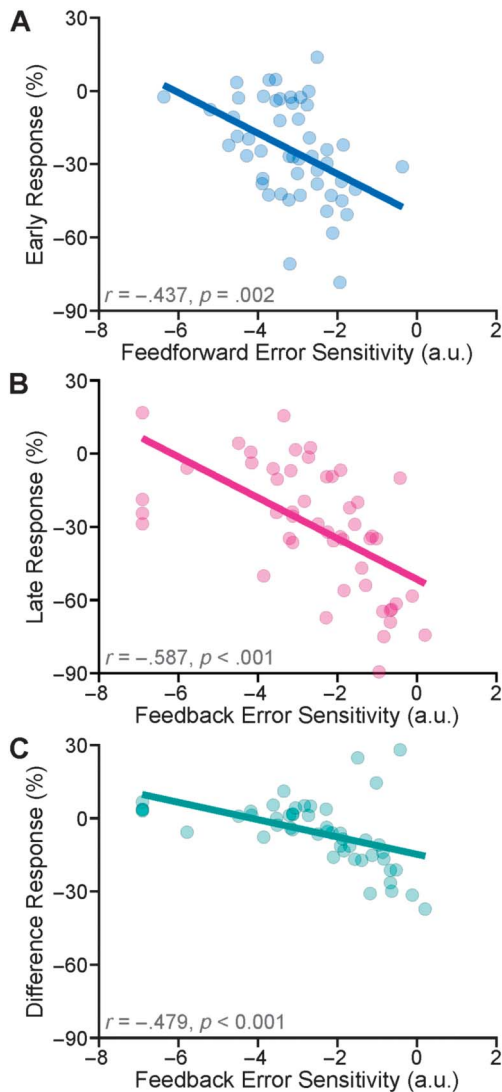


(the formant perturbation was applied suddenly without a ramp phase; see Figure 2B). We measured early adaptive responses (within 0–80 ms) to estimate the contributions of the feedforward control system. We also measured late adaptive responses (within 220–300 ms) to estimate the contributions of the combined feedforward and feedback control systems. We hypothesized that if the feedforward and feedback mechanisms contribute to the late responses, and only the feedforward mechanism contributes to early responses, then the late responses would be different from early responses. To determine model-based estimates of error sensitivity for each participant, we fitted the state-space model to the participant's adaptive responses and extracted the feedforward ( $\beta_{FF}$ ) and feedback ( $\beta_{FB}$ ) error sensitivity. To determine data-driven estimates of error sensitivity, we used the average early adaptive responses as an estimate of the error sensitivity of the feedforward control system and the average difference responses (the difference between late and early responses) as an estimate of the error sensitivity of the feedback control system. To test whether the model accounts for the error sensitivity of the control systems, we examined the relationship between the model-based and data-driven estimates of error sensitivity of the control systems. Additionally, because

both feedforward and feedback control systems rely on prediction errors, we examined the relationship between error sensitivity in the feedforward and feedback control systems.

Consistent with our hypothesis, we found that the late adaptation responses were larger than early adaptation responses in the hold phase. These results suggested that early responses were influenced by the feedforward system, whereas the late responses were influenced by both the feedforward and feedback systems. Thus, the difference responses included the contributions of the feedback control system. As shown in Figure 3A, the difference responses changed immediately after the start of the perturbation in the hold phase and again immediately after the removal of the perturbation in the end phase. These results are convincing evidence that the difference response (i.e., the output of the feedback control system) is driven by auditory prediction error. In the context of our state-space model, the auditory prediction error is zero during the baseline phase; immediately after the perturbation starts in the hold phase, participants experience a large auditory prediction error (i.e., the discrepancy between auditory target and auditory feedback; see Equation 3). Based on this large prediction error, the feedback control system generates a large corrective

**Figure 5.** As data-driven estimates of error sensitivity of the feedforward and feedback control systems, we used the average early adaptation responses during the last 30 trials of the hold phase and the average difference responses (the difference between late and early responses) in the last 30 trials of the hold phase, respectively. As model-based estimates of error sensitivity, we used (logit-transformed) feedforward and feedback error sensitivity ( $\beta_{FF}$  and  $\beta_{FB}$ ) of the fitted state-space models. Our correlational analyses indicated that transformed feedforward error sensitivity significantly correlated with early responses (A), and transformed feedback error sensitivity significantly correlated with the late responses (B) and with the difference responses (C). a.u. = arbitrary unit.



response (see Equation 5), which explains the large change in difference responses at the beginning of the hold phase. However, the feedforward control system uses the error to update its estimate of the perturbation (see Equation 4) and to modify the feedforward motor commands in subsequent trials (see Equation 1). Over the course of trials in the hold phase, the feedforward gradually updates its estimate of the perturbation and adjusts its motor commands to decrease

auditory prediction errors. Immediately after the perturbation is removed in the end phase, participants again experience a large prediction error—participants expect to receive auditory perturbations, but they receive no perturbations. This large prediction error again suggests a large corrective response, which explains the large change in the difference responses at the beginning of the end phase. During the end phase, the feedforward system continues adjusting its motor commands to reduce the prediction error gradually. Overall, these results suggest that (a) contributions of feedforward and feedback control systems can be measured in one task using early responses and late responses and (b) the state-space model can be used to explain how feedforward and feedback control systems use auditory prediction errors to generate adaptive and corrective responses in auditory perturbation paradigms.

Although we interpreted the difference between early and late responses in the context of auditory prediction error, one could argue that the difference between late and early responses is due to the formant transitions (coarticulatory effects) at the end of the vowel. Late responses were closer to the end of the vowel and are more likely to be influenced by the formant transitions imposed by the articulatory constraints of the next consonant. Both early and late responses were baseline-corrected on a word-specific basis such that, for each word, we subtracted the average responses during the baseline trials from all trials. Thus, changes in the difference between early and late responses throughout trials are unlikely to be primarily driven by formant transitions. However, the magnitude of the corrective responses implemented by the feedback control system (i.e., the difference responses) can be modulated by the articulatory constraints of the next consonant (in addition to prediction errors). That is because to implement the corrective responses, participants would need to change their articulatory configurations during the vowel, but the extent of this change may be constrained by the articulatory configurations of the next consonant. In fact, the results of our previous compensation study (Daliri, Chao, & Fitzgerald, 2020) indicated that the magnitude of corrective responses might be influenced by articulatory constraints. However, in the current study, the main effect of word and interactions involving the word factor were not statistically significant. This discrepancy may be explained by the fact that we measured corrective responses based on the time window of 220–300 ms, whereas in the previous study, we measured corrective responses based on the time window of 300–400 ms (i.e., when corrective responses are larger). Overall, while the difference responses are primarily driven by auditory prediction errors (induced by the auditory perturbations), we cannot completely rule out the potential effects of the articulatory constraints of consonantal environments on the magnitude of the difference responses.

One of the goals of this study was to test whether the state-space model accounts for the error sensitivity of the control systems. Toward this goal, we examined the relationship between the model-based estimates of error sensitivity and data-driven estimates of error sensitivity of the

control systems. We found that feedforward error sensitivity ( $\beta_{FF}$ ) correlated with the average early responses, and feedback error sensitivity ( $\beta_{FB}$ ) correlated with the average difference responses. Close examination of the fitted models showed that, on average, the model explained  $\sim 60\%$  of the variance of the individual adaptive responses— $r$ -squared, as a measure of goodness of fit, ranged from .125 to .917. These results suggested that the state-space model did, in fact, capture the dynamics of adaptive responses, and the model's parameters can be used to estimate each participant's error sensitivity. However, given the large variability of the goodness of fit and the large individual variability in the adaptive responses (see Figure 3B), there may be other contributing factors to adaptive responses that were not included in the model (e.g., characteristics of participants or auditory perturbations). We also examined the relationship between error sensitivity parameters ( $\beta_{FF}$  and  $\beta_{FB}$ ) and prediction sensitivity parameter ( $\alpha$ ). In our computational model (see Equation 4),  $\alpha$  is the weight assigned to the participant's prediction of the magnitude of the perturbation. The model uses the weighted prediction error and the weighted estimate of the perturbation in the current trial to update the estimate of perturbation, which is used to modify the feedforward motor command in the next trial. In other words, prediction sensitivity determines the similarity of our prediction of the perturbation in the next trial with the current prediction of the perturbation (Daliri & Dittman, 2019; Shadmehr & Mussa-Ivaldi, 2012). Our analyses showed that prediction sensitivity correlated negatively with feedforward error sensitivity (see Figure 4C;  $r = -.434$ ): participants who relied more on their prediction (higher prediction sensitivity) tended to rely less on prediction error (lower feedforward error sensitivity). Interestingly, we found a positive correlation between prediction sensitivity and feedback error sensitivity (see Figure 4D;  $r = .429$ ): Participants who relied more on their prediction tended to generate larger corrective responses. These results may indicate that the feedforward and feedback control systems are distinct and use prediction errors differently. Finally, as shown in Figure 4, there were larger individual variabilities for prediction sensitivity and feedback error sensitivity, whereas there was a relatively small individual variability for feedforward error sensitivity. These results may suggest that at least in adaptation paradigms without a ramp phase, the large individual variabilities in early adaptive responses may be less related to individual variability in feedforward error sensitivity and more related to individual variability in prediction sensitivity (in addition to other factors such as sensitivity to motor-sensory errors). Overall, our computational model accounts for the error sensitivity of the feedforward and feedback control systems and can provide insights about individual variability in responding to auditory perturbation errors.

Another goal of this study was to examine the relationship between error sensitivity in the feedforward and feedback control systems, given their reliance on prediction errors. Examining data-driven measures of error sensitivity, we found that the average early responses did not correlate

with the average of difference responses. Similarly, there was no significant relationship between model-based measures of error sensitivity (feedforward and feedback error sensitivity). These results suggested that, despite the reliance on auditory prediction error, the feedforward and feedback systems operate independently and use the prediction error differently. In other words, participants with higher feedforward error sensitivity do not necessarily have higher feedback error sensitivity. These results are also consistent with the results of several previous studies that used adaptation and compensation tasks to separately measure feedforward and feedback control systems' responses to auditory errors (Franken et al., 2019; Hawco & Jones, 2009; Lester-Smith et al., 2020; Parrell et al., 2017; Scheerer & Jones, 2018). Our results are especially consistent with the results of a recent study by Franken et al. (2019). They examined early responses (time window of 50–150 ms) and late responses (time window of 1000–1500 ms) when participants experienced frequent and infrequent auditory perturbations (in two different paradigms). They reported a lack of correlation between within-trial corrective responses and adaptive responses, suggesting that the processes underlying these responses are distinct and independent. Collectively, our results indicated that the feedforward and feedback control systems use prediction errors but independently transform the error to adaptive and corrective responses and have uncorrelated error sensitivity.

To estimate corrective responses in a given trial, we used the difference between late and early responses. Early and late responses depend on the specific time windows that are used to extract the responses. In this study, we used the early time window of 0–80 ms and the late time window of 220–300 ms (see the Supplemental Material S1 for results of the time window of 110–190 ms). The rationale for using these time windows was based on previous studies that have reported that corrective responses start within 100–200 ms (typically around 150 ms) from the onset of perturbations and gradually increase up to at least 400 ms after the onset of perturbations (Daliri, Chao, & Fitzgerald, 2020; Niziolek & Guenther, 2013; Parrell et al., 2017). Therefore, responses at time points before 100 ms are primarily influenced by the feedforward control system, whereas responses at time points after 200 ms are influenced by the feedforward and feedback control systems. In support of this argument, we found a strong correlation between early and late responses (see Figure 3C). Although participants were trained to produce vowels longer than 400 ms, several participants produced vowels that were shorter than 400 ms; thus, we selected 220–300 ms as the late time window to limit the number of excluded trials (i.e., trials with duration less than 300 ms). This choice of time window could influence the magnitude of late responses and, therefore, the difference responses. Additionally, because the models were fitted to the early and late responses, the model's parameters—especially the feedback error sensitivity—are influenced by choice of the time window. We speculate if the late time window was at later time points (e.g., 300–400 ms), the difference responses and feedback error

sensitivity may have been larger (as the corrective responses gradually increase). It should be noted that our current computational model is based on the assumption that the early and late time windows are distinct and located before and after a time point where the feedback control system generates a response (e.g., ~150 ms). The model can be further developed in future studies to be able to use all data points in a trial (rather than data from early and late time-windows). Based on such models, one could examine whether or not the onset of the responses generated by the feedback control system changes over the course of an auditory perturbation paradigm.

Although a comprehensive comparison between the state-space model and previous models is beyond the scope of this study, it is important to acknowledge the limitations of the state-space model in relation to previous models. The structure of the state-space model that we developed in this study is similar to those of the previous models (Guenther, 2016; Houde & Nagarajan, 2011; Kearney et al., 2020; Parrell & Houde, 2019). The state-space model is particularly similar to the “simpleDIVA” model that was recently developed by Guenther and colleagues (Kearney et al., 2020). Both the state-space model and the simpleDIVA model include the feedforward and feedback control systems and use prediction errors. To qualitatively compare the two models, we fitted the simpleDIVA to the adaptive responses of each participant. The average simulated responses of all participants are shown in Supplemental Material S3. These results showed that the two models performed similarly with one noticeable difference: The difference responses at the end of the hold phase appeared to be smaller (or nonexistent) for the simpleDIVA. This qualitative difference may be related to a key difference in how the feedforward motor commands are calculated and updated across trials in the two models. In the simpleDIVA, the output of the feedback component (i.e., corrective response) is used to modify the feedforward commands (i.e., adaptive response). In the state-space model, the prediction error is used to update the estimate of the perturbations, and this estimate is used to calculate the feedforward motor commands. The simpleDIVA’s approach implies that the feedback and feedforward control systems are linked; however, our results along with the results of previous studies (Franken et al., 2019; Hawco & Jones, 2009; Lester-Smith et al., 2020; Parrell et al., 2017; Scheerer & Jones, 2018) suggest that the feedforward and feedback control systems may function independently. Another key difference between the two models is that whereas the simpleDIVA model calculates both auditory and somatosensory prediction errors, the state-space model includes the auditory prediction errors only. Given the importance of somatosensory feedback (Daliri et al., 2013; Guenther, 2016; McGuffin et al., 2020), this is a major limitation of the current state-space model that needs to be addressed in future versions of the model.

One aspect of our study design that deserves comment is the magnitude of the auditory perturbation. We used a participant-specific perturbation that shifted a participant’s

/e/ toward her /æ/: the magnitude of shift was equal to the participant-specific distance between the centroids of /e/ and /æ/. This perturbation would generate a categorical error, and the brain may integrate such errors differently in comparison with within-category errors (Chao et al., 2019; Niziolek & Guenther, 2013). Our current state-space model does not differentiate between within-category and cross-category errors and respond to these two types of errors in the same way. Additionally, the model assumes the error sensitivity is constant and is not dependent on the magnitude or relevance of the prediction error. Previous studies have suggested that the brain evaluates the prediction errors and adapts less to large or irrelevant errors (Berniker & Körding, 2008; Daliri & Dittman, 2019; Shadmehr & Mussa-Ivaldi, 2012). Additionally, in a recent study (Daliri, Chao, & Fitzgerald, 2020), we showed that corrective responses to large perturbations are proportionally smaller than responses to small perturbations. To address these limitations, the structure of the state-space model needs to be modified so that the model can evaluate the errors (e.g., based on the magnitude of the error) and generate responses based on its error evaluation.

In summary, we examined the contributions of the feedforward and feedback control systems in an adaptation paradigm. Toward this goal, we examined early adaptive responses and late adaptive responses. Given that early responses are influenced by the feedforward system, and late responses are influenced by both the feedforward and feedback systems, we examined the difference between late and early responses (difference responses) to isolate the contributions of the feedback control system. Additionally, we developed a state-space model to estimate the feedforward and feedback control systems’ error sensitivity. We found that early responses (estimate of the feedforward system) did not correlate with difference responses (estimate of the feedback system), suggesting that the feedforward and feedback control systems function independently despite their reliance on prediction error. We also found that model-based estimates of error sensitivity of the feedforward ( $\beta_{FF}$ ) and feedback ( $\beta_{FB}$ ) systems correlated with early responses and difference responses (data-driven estimates of the feedforward and feedback systems). These results suggested that the state-space model was able to capture the dynamic of adaptive responses, and the model’s parameters can be used to estimate each participant’s error sensitivity. Together, these results showed that it is possible to efficiently and accurately determine the feedforward and feedback control systems’ sensitivity to prediction errors using early and late adaptive responses in combination with the state-space model.

## Acknowledgments

This work was supported by National Institutes of Health Grant R21 DC017563 (awarded to A. Daliri). We thank Damaris Ochoa and Sara Chao for their contributions to participant recruitment for this project.

## References

- Abur, D., Lester-Smith, R. A., Daliri, A., Lupiani, A. A., Guenther, F. H., & Stepp, C. E. (2018). Sensorimotor adaptation of voice fundamental frequency in Parkinson's disease. *PLOS ONE*, 13(1), Article e0191839. <https://doi.org/10.1371/journal.pone.0191839>
- American Speech-Language-Hearing Association. (1997). *Guidelines for audiologic screening: Panel on audiologic assessment*.
- Ballard, K. J., Halaki, M., Sowman, P., Kha, A., Daliri, A., Robin, D. A., Tourville, J. A., & Guenther, F. H. (2018). An investigation of compensation and adaptation to auditory perturbations in individuals with acquired Apraxia of speech. *Frontiers in Human Neuroscience*, 12, 510. <https://doi.org/10.3389/fnhum.2018.00510>
- Berniker, M., & Körding, K. P. (2008). Estimating the sources of motor errors for adaptation and generalization. *Nature Neuroscience*, 11(12), 1454–1461. <https://doi.org/10.1038/nn.2229>
- Cai, S. (2015). *Audapter*. [https://github.com/shanqing-cai/audapter\\_matlab](https://github.com/shanqing-cai/audapter_matlab)
- Chao, S.-C., Ochoa, D., & Daliri, A. (2019). Production variability and categorical perception of vowels are strongly linked. *Frontiers in Human Neuroscience*, 13, 96. <https://doi.org/10.3389/FNHUM.2019.00096>
- 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](https://doi.org/10.1044/2020_JSLHR-19-00422)
- Daliri, A., & Dittman, J. (2019). Successful auditory motor adaptation requires task-relevant auditory errors. *Journal of Neurophysiology*, 122(2), 552–562. <https://doi.org/10.1152/jn.00662.2018>
- Daliri, A., & Max, L. (2015a). Electrophysiological evidence for a general auditory prediction deficit in adults who stutter. *Brain and Language*, 150, 37–44. <https://doi.org/10.1016/j.bandl.2015.08.008>
- Daliri, A., & Max, L. (2015b). Modulation of auditory processing during speech movement planning is limited in adults who stutter. *Brain and Language*, 143, 59–68. <https://doi.org/10.1016/j.bandl.2015.03.002>
- Daliri, A., & Max, L. (2016). Modulation of auditory responses to speech vs. nonspeech stimuli during speech movement planning. *Frontiers in Human Neuroscience*, 10, 234. <https://doi.org/10.3389/fnhum.2016.00234>
- Daliri, A., & Max, L. (2018). Stuttering adults' lack of pre-speech auditory modulation normalizes when speaking with delayed auditory feedback. *Cortex*, 99, 55–68. <https://doi.org/10.1016/j.cortex.2017.10.019>
- Daliri, A., Murray, E. S. H., Blood, A. J., Burns, J., Noordzij, J. P., Nieto-Castañón, A., Tourville, J. A., & Guenther, F. H. (2020). Auditory feedback control mechanisms do not contribute to cortical hyperactivity within the voice production network in adductor Spasmodic Dysphonia. *Journal of Speech, Language, and Hearing Research*, 63(2), 421–432. [https://doi.org/10.1044/2019\\_JSLHR-19-00325](https://doi.org/10.1044/2019_JSLHR-19-00325)
- Daliri, A., Prokopenko, R. A., Flanagan, J. R., & Max, L. (2014). Control and prediction components of movement planning in stuttering versus nonstuttering adults. *Journal of Speech, Language, and Hearing Research*, 57(6), 2131–2141. [https://doi.org/10.1044/2014\\_JSLHR-S-13-0333](https://doi.org/10.1044/2014_JSLHR-S-13-0333)
- Daliri, A., Prokopenko, R. A., & Max, L. (2013). Afferent and efferent aspects of mandibular sensorimotor control in adults who stutter. *Journal of Speech, Language, and Hearing Research*, 56(6), 1774–1788. [https://doi.org/10.1044/1092-4388\(2013\)12-0134](https://doi.org/10.1044/1092-4388(2013)12-0134)
- Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., & Chang, S.-E. (2017). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. *Developmental Science*, 21(2), Article e12521. <https://doi.org/10.1111/desc.12521>
- Franken, M. K., Acheson, D. J., McQueen, J. M., Hagoort, P., & Eisner, F. (2019). Consistency influences altered auditory feedback processing. *Quarterly Journal of Experimental Psychology*, 72(10), 2371–2379. <https://doi.org/10.1177/1747021819838939>
- Fuchs, S., Cleland, J., & Rochet-Capellan, A. (Eds.). (2019). *Speech production and perception: Learning and memory*. Peter Lang. <https://doi.org/10.3726/b15982>
- Guenther, F. H. (2016). *Neural control of speech*. MIT Press. <https://doi.org/10.7551/mitpress/10471.001.0001>
- Hawco, C. S., & Jones, J. A. (2009). Control of vocalization at utterance onset and mid-utterance: Different mechanisms for different goals. *Brain Research*, 1276, 131–139. <https://doi.org/10.1016/j.brainres.2009.04.033>
- Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279(5354), 1213–1216. <https://doi.org/10.1126/science.279.5354.1213>
- Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, 5, 82. <https://doi.org/10.3389/fnhum.2011.00082>
- Kearney, E., Nieto-Castañón, A., Weerathunge, H. R., Falsini, R., Daliri, A., Abur, D., Ballard, K. J., Chang, S.-E., Chao, S.-C., Murray, E. S. H., Scott, T. L., & Guenther, F. H. (2020). A simple 3-parameter model for examining adaptation in speech and voice production. *Frontiers in Psychology*, 10, 2995. <https://doi.org/10.3389/fpsyg.2019.02995>
- Kim, K. S., Daliri, A., Flanagan, J. R., & Max, L. (2020). Dissociated development of speech and limb sensorimotor learning in stuttering: Speech auditory-motor learning is impaired in both children and adults who stutter. *Neuroscience*, 451, 1–21. <https://doi.org/10.1016/j.neuroscience.2020.10.014>
- Kim, K. S., Wang, H., & Max, L. (2019). 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](https://doi.org/10.1044/2020_JSLHR-19-00419)
- Krakauer, J. W., Hadjiosif, A. M., Xu, J., Wong, A. L., & Haith, A. M. (2019). Motor learning. *Comprehensive Physiology*, 9(2), 613–663. <https://doi.org/10.1002/cphy.c170043>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). Imertest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lenth, R. (2019). *emmeans: Estimated marginal means, aka least-squares means* (Volume 1.3.3). R package. <https://cran.r-project.org/package=emmeans>
- 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](https://doi.org/10.1044/2020_JSLHR-20-00192)
- Max, L., & Daliri, A. (2019). Limited pre-speech auditory modulation in individuals who stutter: Data and hypotheses. *Journal of Speech, Language, and Hearing Research*, 62(8S), 3071–3084. [https://doi.org/10.1044/2019\\_JSLHR-S-CSMC7-18-0358](https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0358)
- McGuffin, B. J., Liss, J. M., & Daliri, A. (2020). The orofacial somatosensory system is modulated during speech planning and production. *Journal of Speech, Language, and Hearing Research*, 63(11), 3628–3642. [https://doi.org/10.1044/2020\\_JSLHR-20-00192](https://doi.org/10.1044/2020_JSLHR-20-00192)

- Research*, 63(8), 2637–2648. [https://doi.org/10.1044/2020\\_JSLHR-19-00318](https://doi.org/10.1044/2020_JSLHR-19-00318)
- Merrikhi, Y., Ebrahimpour, R., & Daliri, A.** (2018). Perceptual manifestations of auditory modulation during speech planning. *Experimental Brain Research*, 236(7), 1963–1969. <https://doi.org/10.1007/s00221-018-5278-3>
- 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>
- Parrell, B., Agnew, Z., Nagarajan, S., Houde, J., & 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](https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0127)
- R Core Team.** (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Revelle, W.** (2018). *psych: Procedures for psychological, psychometric, and personality research*. Northwestern University. <https://cran.r-project.org/package=psych>
- Scheerer, N. E., & Jones, J. A.** (2018). The role of auditory feedback at vocalization onset and mid-utterance. *Frontiers in Psychology*, 9, 2019. <https://doi.org/10.3389/FPSYG.2018.02019>
- Shadmehr, R., & Mussa-Ivaldi, S.** (2012). *Biological learning and control: How the brain builds representations, predicts events, and makes decisions*. MIT Press. <https://mitpress.mit.edu/books/biological-learning-and-control>
- Stepp, C. E., Lester-Smith, R. A., Abur, D., Daliri, A., Noordzij, P. J., & Lupiani, A. A.** (2017). Evidence for auditory-motor impairment in individuals with hyperfunctional voice disorders. *Journal of Speech, Language, and Hearing Research*, 60(6), 1545–1550. [https://doi.org/10.1044/2017\\_JSLHR-S-16-0282](https://doi.org/10.1044/2017_JSLHR-S-16-0282)
- 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>
- Tremblay, S., Shiller, D. M., & Ostry, D. J.** (2003). Somatosensory basis of speech production. *Nature*, 423(6942), 866–869. <https://doi.org/10.1038/nature01710>