Determining the Relevance of Prediction Errors to Auditory-Motor Adaptation

by

Kimiya Kasraeian

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved July 2021 by the
Graduate Supervisory Committee:

Ayoub Daliri, Chair
Xin Luo
Corianne Rogalsky

ARIZONA STATE UNIVERSITY

August 2021

ABSTRACT

In this study, I used two computational models (state-space model and simple DIVA model) to determine the speech motor system's sensitivity to auditory errors that are relevant vs. irrelevant and introduced gradually or suddenly. I applied formant perturbations (first and second formants of /ɛ/ were shifted toward formants of /æ/) to generate auditory errors. Then I measured subjects' adaptive responses to the formant perturbations. I examined (a) the accuracy of models in explaining the adaptive responses (b) the relationship between the models' parameters and the adaptive responses. My results showed that both models predict the adaptive responses to errors. However, the models' parameters differently correlated with the adaptive responses, suggesting that while the models perform similarly, they provide different insights about adaptive responses to auditory errors. These results have important implications for speech motor learning and production models and shed light on neural processes involved in generating adaptive responses.

Keywords: speech, feedforward, feedback, adaptation, modeling

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor Dr. Ayoub Daliri for his assistance, support, and patience. Besides my advisor, I would like to thank my thesis committee Dr. Corianne Rogalsky and Dr. Xin Luo, for their unwavering support.
I would like to thank my fantastic labmates, especially Sara Chao, who helped and supported me since the first day. I would also like to thank ASU students who participated in this research.

My sincere thanks also goes to Dr. Megan Petrov, who constantly believed in me; all her passion for research and teamwork led me to dive into the research with more enthusiasm.

I would like to thank all my good friends in Arizona who supported me, especially during the COVID-19 pandemic.

I would like to thank my parents for their endless love and support, which without them, this would not have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Suppose to put a Fresnel prism (displacing prism) on my friend’s eyes and ask her to reach a visual target. She would initially miss the target, but she would adapt and learn the process after some trials and errors. The angular difference that this prism makes is a perturbation, and after a while, she will adapt to this perturbation. If this prism being removed, she will displace the target’s direction angle as much as in the beginning but in the opposite direction. Motor adaptation as a form of motor learning is an error cancellation process in which the nervous system leans to ignore the constantly changing environment's effect while producing accurate movement. Speech motor control shares many features with other sensory-motor systems, such as limb motor control. There are several stages to make an idea into the final words that convey it to the listeners in acoustic waves for someone to speak. One stage is to translate thoughts into linguistic representation (the speech product). The second stage is how this representation is built (the speech production process). This message is transmitted to muscles, which activate the vocal tract and articulators to produce speech. (Berg & Levelt, 1990). During speech production, the target is to make an articulatory movement in order to produce a specific acoustic sound (Houde & Jordan, 1998). The speech motor system relies on the auditory and somatosensory systems to produce speech accurately. Studying the auditory and somatosensory feedback is essential to understand speech production and speech learning and adaptation systems in speakers with normal or disordered speech. For this purpose, computational models are an important, repeatable and testable mechanism, and by comparing different models, each model’s advantages and disadvantages would be highlighted and can be used for future investigations. (Parrell & Houde, 2019a).

Recent theories of speech (F. H. Guenther, 2016; Houde & Nagarajan, 2011) suggest that speech motor system employs two control mechanisms: feedforward and feedback. The feedforward control system produces motor commands aimed at desired sensory goals (auditory and somatosensory goals), and articulators perform the motor commands. The speech motor system also predicts the sensory consequences of the motor commands. The feedback control system scans the sensory outcomes of the motor commands to check production accuracy. After speech production, the speech motor system compares the previous sensory prediction to the sensory feedback to estimate the potential prediction errors (the mismatch between prediction and feedback). The speech motor system uses the prediction error to adapt its motor commands and modify its feedforward control system. Overall, prediction error plays an important role in ensuring the accuracy of speech production.

In post-lingually deaf speakers, pitch and loudness control quickly decreases after hearing loss, but they still can produce intelligible speech for decades even though their auditory feedback is different (Cowie & Douglas-Cowie, 1992). However, we cannot assume speech production is solely relying on the feedforward process. For example, children who become deaf before learning to produce speech do not naturally learn how to speak (Oller & Eilers, 1988). This evidence and more similar evidence show that auditory feedback is crucial to learn speaking. In an early study, Lee discussed that auditory feedback is not ignored after the speaking skill has been learned. He showed that when a delay is heard in one's speech (delayed auditory feedback), the speech becomes less fluent (B. S. Lee, 1950). Thus, one way to study how the sensory control mechanisms working is to apply auditory feedback perturbation and generate an error (Ingo Titze, 1994).

The vocal tract generates acoustic resonance that can be represented on the sound spectrum; formant is acoustic wave concentration around a particular frequency in the spectrum. Applying auditory perturbation can be through generating unexpected formant perturbation during the production of the prolonged word or by applying perturbations systematically and constantly over several words (Fuchs, S., Cleland, J., & Rochet-Capellan, 2019). This experiment allows us to investigate the sensory-motor control system's adaptive response to perturbations. In the prism experiment, when the participant reaches a visual target (the motor action), she adapts to the visual target's shifted image through prisms (perturbed sensory feedback). This concept is called sensory-motor adaptation. In the speech domain, in the adaptation experiment, a formant is shifted along toward another formant (like wearing displacing prisms), and the sound has been perturbed. For instance, participants hear more /a/ when they produce /e/, so they change their production of /e/ to compensate for the altered feedback. At this moment, the participant feels a mismatch between what she expected to hear and what she hears, and as a response to this mismatch, she generates responses to correct it. This corrective feedback response does not happen after the first trial, since the error sensitivity is small (Daliri & Dittman, 2019a; Shadmehr & Mussa-Ivaldi, 2013), applying small changes in their adaptive feedforward control system leads to making small changes in ongoing trials so after some trials the change would be measurable.

Although prediction error is an important factor for both feedforward and feedback control systems, it is not clear (1) how much error sensitivity each of these control systems has, (2) whether the participant changes her corrective response based on the somatosensory feedback (since sensory feedback in speech is not limited to auditory domain and in these experiments, we only manipulate auditory feedback control system) and (3) although there is a little uncertainty on estimation from the different participant, but how much individual's differences would affect the experimental study, for instance, individual's learning rate.

One way to study sensorimotor adaptation is to design an experimental study to examine motor control behavior; the other approach is to generate computational models and simulate the speech production system to 1) better understand motor control behavior, 2) conduct multiple simulated experiments. To validate computational models, we need to examine them to compare them to the experimental data to examine their accuracy. Daliri & Dittman adopted a computational model (state-space-model), initially used in limb motor control (Shadmehr & Mussa-Ivaldi, 2013), to estimate error sensitivity in the adaptation paradigm using feedforward and feedback control system. This model's free parameter is associated with feedback sensitivity, error sensitivity, and feedback response magnitude (Daliri & Dittman, 2019a). Kearney et al. (Kearney et al., 2020) tested a simple 3-parameter mathematical model that quantified feedback and feedforward contribution in sensory-motor adaptation. This model is a simplified version of the DIVA model, so this model is called Simple DIVA. In this model, the free parameters are associate with the gain for the auditory feedback control system, the gain for the somatosensory control system, and the learning rate. Both model's parameters can be estimated by fitting the model to participants' responses. To compare these two models, the state-space model defined parameters regarding feedforward and feedback control systems and distinguished them in early and late time points, although it is focused on auditory feedback. On the other hand, the simple DIVA model focuses on sensory feedback (auditory and somatosensory). In the current study, we conducted an auditory perturbation experiment. We test whether the state-space model and simple Diva model can predict and estimate error sensitivity and sensory gain accurately; and whether these parameters can explain observed behavioral responses in the adaptation paradigm. The goal is to see how well these models can fit the real data.

METHOD

Participants

All participants signed the Arizona State University IRB consent form. We recruited 28 (M = 23.4 years, SD = 4.3 years). All participants were native speakers of American English. Each session lasts 2 hours. Participants did not have a history of neurological, psychological, or speech-language disorders (self-reported). All participants had hearing thresholds of 20 dB HL or less at all octave frequencies of 250 Hz to 8000 Hz.

Apparatus

The experimental apparatus is shown in figure 1. All experiments were conducted in a sound booth. Participants sat inside the sound booth in front of a computer monitor. Shure microphone (SM58) positioned 15 cm away from the moth corner at a 45-degree angle. The microphone signal was amplified via TubeOpto 8, ART preamplifier, and digitalized via an external audio interface, 8pre, MOTU soundcard with a sampling rate of 48 kHz, which transmitted the signal to the computer and transmitted back to the audio interface. The signal was then amplified via S.phone, Samson Technologies Corp, and played back binaurally through the earphones (ER-1, Etymotic Research Inc.). The auditory feedback was amplified 5 dB higher than the microphone signal's intensity, so we also calibrated the amplification levels of the microphone and earphone amplifiers before each session.

We used the Audapter software package (MATLAB-based) for near real-time tracking and shifting formant frequencies. For this purpose, this package uses linear predictive coding (LPC); for female participants, we used an LPC order of 15, and for male participants, an LPC order of 17. We recorded the signal from the microphone and auditory feedback signal (the played back sound in earphones) simultaneously on two separate channels using a digital audio recorder, Tascam DR-680MKII. The delay between these two channels is about 16.4 ms. To measure this delay, we used a 2-cc coupler (Type 4946, Bruel & Kjaer Inc.) connected to a sound level meter (Type 2250A, Bruel & Kjaer Inc.)(Daliri & Max, 2015).

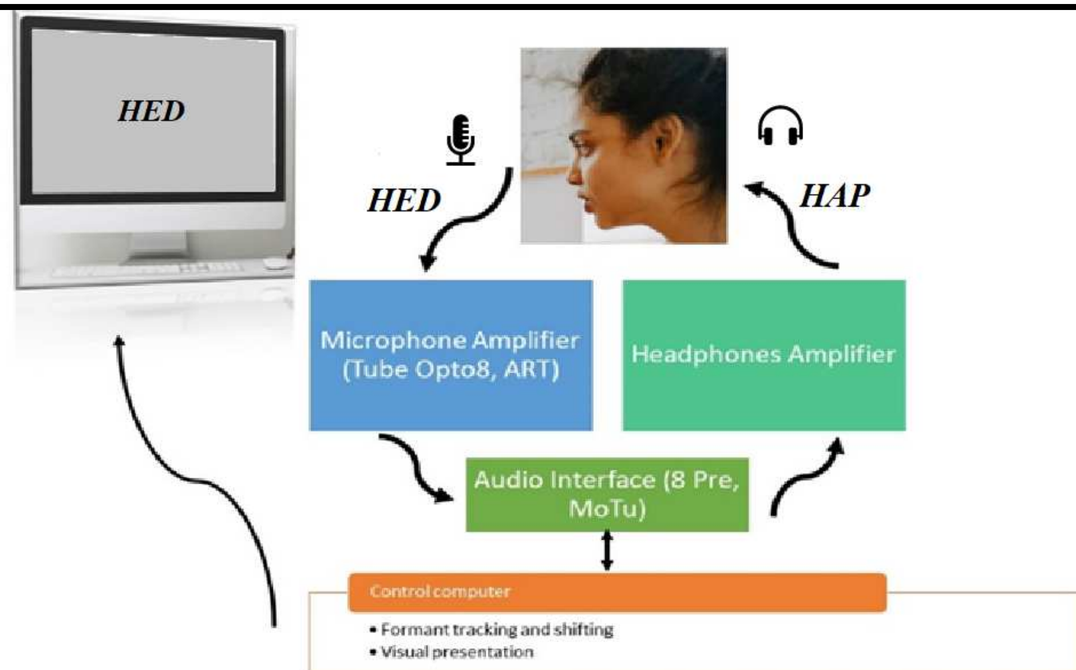

Figure 1. Experimental apparatus

Procedure

All participants completed the preparation task at the beginning, which included 30 trials. On the monitor in front of the participant, we showed the participant CVC (consonant-vowel-consonant) one syllabic word in black font and on a grey background which lasted 2.5 s each, and between two sequential trials, there was a 1-1.5 s break. The words were “Head,” “Hep,” and “Heck”; each word was randomly presented ten times. We wanted the production speech to be in desired intensity and duration (70–80 dB SPL; 400–600 ms), so feedback was shown on the monitor after each trial. There were two bars on the screen while producing speech; the top bar on the screen showed the speech duration, and the bottom bar showed its intensity. If they were too loud or long, they would get red bars, and if they were too soft or short, they would get blue bars. If participants were producing speech in the desired range, the bars would be green. This task lasted 2 minutes.

In the next task, the participant completed the vowel production task, which included 75 trials in which participants completed a word reading task in around 3.5 minutes in duration. “Hip,” “Hep," and “Hap” were randomly presented on the screen (25 times each), and participants received visual feedback only if they were out of range. This task aimed to identify each participant’s centroids (the middle of a vowel distribution in the F1-F2 coordinates) of /ɛ/ and /æ/. The Audapter software initially calculated first and second formant frequencies (F1 and F2); we also used a MATLAB script to get the first and second formants’ average for each production. Using mean formants, we detected each participant’s centroids of /ɛ/ and /æ/ and in the F1-F2 coordinates, we calculated ɛ-æ Euclidean distance (in Hz), and the angle between centroids. The ɛ-æ distance and angle were used in the adaptation task to determine participant-specific formant perturbations.

In the last task (adaptation task), we applied auditory feedback perturbation in four conditions: sudden shift, gradual shift, sudden clamp, and gradual clamp. In sudden perturbation (step), the formant perturbation is introduced suddenly at once. In gradual perturbation (ramp), formant perturbation is gradually applied in sequence during trials. As mentioned above, we found each participants’ vowel centroids; in clamp conditions, the auditory feedback was fixed on the vowel centroids, but in shift, the auditory feedback depended on the participants’ F1 and F2 in each trial. Each condition consisted of 216 trials. Both sudden and gradual conditions consisted of a baseline phase (first 36 trials, no perturbation), hold phase (144 trials of shift or clamp), and an end phase (36 trials, no perturbation). Perturbations were designed based on participant’s /ɛ-ea/ distance and angle to shift the F1 and F2 of /ɛ/ to F1 and F2 of /ea/.

Data analysis

In addition to Audapter formant tracking, we checked the formant's accuracy in each production on the spectrogram using MATLAB scripts. We selected the beginning and the end of each trial's vowels by demonstrating them on spectrogram and time-domain waveform. Then, we used this clean data to extract F1 and F2. For each trial, we extracted F1 and F2 and showed them in F1-F2 coordinates. Adaptation responses are calculated based on each subject's individual centroid of the reference point (here is /ε/). When the adaptation response is positive, the participant adapts to perturbed auditory feedback and changes her response toward /æ/. We did not include the deviation response in our analysis. Since the magnitude of perturbation for each participant was different, in order to compare responses among participants, we divided responses by the participant's perturbation magnitude to normalize responses.

Computational modeling

In this study, we used two computational models: the state-space model and the simple DIVA.

State-space model

Daliri and Dittman proposed a state space model that estimates the feedforward control mechanism's involvement in the adaptation paradigm (Daliri & Dittman, 2019b). The first premise in this model is that to create a feedforward motor command (F_{FF}) in a trial, one utilizes her estimation of perturbation (X_P) to produce an auditory target (F_T). After the feedforward motor command has been sent, we receive auditory feedback (F_{AF}). If we received any formant perturbation (F_P), auditory feedback contains it.

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

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

Prediction error can be calculated by comparing auditory feedback and auditory target. We can simplify this equation using the equation 1 and 2. So prediction error would be the difference between the current trial estimation of perturbation and the actual formant perturbation.

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

Using prediction error for the current trial, we can update the estimate of perturbation based on current trial information for the next trial. Prediction error for the next trial would be a weighted estimate of perturbation plus weighted prediction error. (β_{FF}) corresponds to each individual's sensitivity to prediction error; in some cases, it might be higher than others, which means they are more sensitive in predicting error. (α) corresponds to our estimation of perturbation, when α is higher, we can rely on the current perturbation estimation to update the estimated perturbation estimate for the next trial.

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

To update the feedback control system and the feedforward control system's command, we use prediction error to update the feedback control system (equation 5) (Kearney et al., 2020; Parrell & Houde, 2019b). Feedback error sensitivity (β_(FB)) is similar to feedforward but for feedback control systems.

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

Since from the beginning of speech production, it takes around 150 ms for us to hear back our production; in the earlier time point for each trial, the effect of the feedback control system is minimal, and we have the maximum contribution of the feedforward control system (F_(Early)). After 150 ms, we have the contribution of the feedback control system as well as feedforward (during the speech production, we constantly have a motor command from the feedforward control system, if it were at once, we would stop producing speech in the middle of trial), we call it F_{Late} which shows the contribution of two control systems. (Frank H. Guenther, 2016; Kearney et al., 2020; Parrell & Houde, 2019a). Since the feedforward does not immediately update during one trial, the difference between late and early responses would demonstrate the feedback control system.

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

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

Simple DIVA model

Kearney et al. 2020 revised the DIVA model (Frank H. Guenther, 2006, 2016) to demonstrate the feedforward control system's contribution, auditory and somatosensory feedback control system. $F1_{produced}$ is the final motor command reaches speech articulators. In one trial, $F1_{produced}$ is the sum of feedforward command ($F1_{FF}$) and correction of incoming sensory feedback ($\Delta F1_{FB}$ (n)). The simple DIVA model postulates that both somatosensory and auditory targets are equal to $F1_T$, the mean of the target $F1_{produced}$. $F1_T$ assumed to be constant in a task, so there is no change in the correct sound production. The following equation characterizes the feedback-based correction on a current trial:

$$F1_{produced}\ (n) = F1_{FF}\ (n) + \Delta F1_{FB}\ (n) \tag{1}$$

$$\Delta F1_{FB}\ (n) = \alpha_A * (F1_T - F1_{AF}\ (n)) + \alpha_S * (F1_T - F1_{SF}\ (n)) \tag{2}$$

$F1_{AF}$ and $F1_{SF}$ correspond to the current amount of auditory and somatosensory feedback before the feedback control system's effect. Equations 3 and 4 demonstrate that $F1_{AF}$ and $F1_{SF}$ are made of the sum of perturbation size in each domain and $F1_{FF}$. Here α_A and α_S represent the gain of auditory and somatosensory feedback, respectively.

$$F1_{AF} = F1_{FF} + \text{auditory perturbation size} \tag{3}$$

$$F1_{SF} = F1_{FF} + \text{somatosensory perturbation size} \tag{4}$$

While we apply perturbation, the participant intends to compensate the perturbation in the opposite direction using an auditory feedback control system, which means $F1_{AF}$ will change. Meanwhile, compensating auditory feedback affects the somatosensory control system as well; there is no perturbation in the somatosensory domain, so it tries to keep the vocal track and articulators normal. So, change of gaining in both domains can affect the compensatory auditory and somatosensory responses. The next equation demonstrates the feedforward command's updating process based on the previous trial:

$$F1_{FF}\,(n+1) = F1_{FF}\,(n) + \lambda_{FF} * \ \Delta F1_{FB}\,(n) \tag{5}$$

Here, λ_{FF} represents feedforward learning rate parameter. To fit these two computational models to a particular dataset, we need to use an optimization tool to find the optimized value of 3 free parameters in both models (α, $\beta_{FF,}$ and β_{FB} in state-space model and α_A and α_S and λ_{FF} in diva model). We have four conditions in our experimental data; after getting the mean data for each condition, we used the "fmincon" function in MATLAB to fit each model to each individual's response. At first, the function calculated early and late response simulation with a set of random numbers in the accepted range of the free parameters (0-1). Then we used the differences between these simulated responses and the participant's responses to make the optimization more accurate. We got the optimized parameters for each model, then we calculated the early and late simulated responses, and finally, we compared these responses to the actual data to see which model can optimize the data better (Daliri & Dittman, 2019a).

RESULTS

Figures 2 and 3 show the participant's group-average adaptation responses in gradual and sudden perturbations. In these figures, the blue and pink lines indicate shift and clamp perturbation, respectively. In both gradual and sudden perturbation, there was a sudden change every 12 blocks (36 trials). Our analysis showed a statistical difference between shift and clamp perturbations at the end of the hold phase in blocks 48 to 60 for the sudden perturbations (p = .002). Similarly, we found a statistical difference between shift and clamp perturbations in blocks 48 to 60 for the gradual perturbations (p = .017).

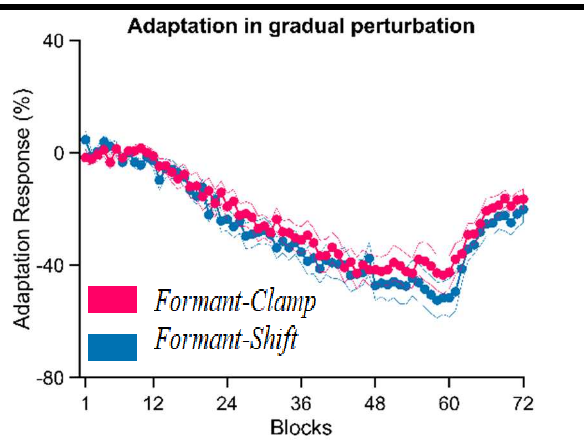

Figure 2, adaptation responses during gradual perturbation

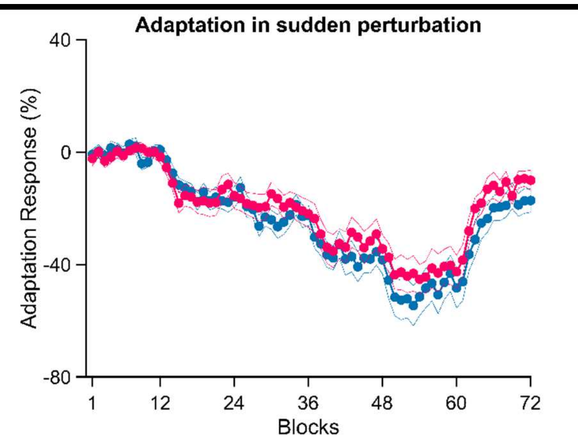

Figure 3, adaptation responses during sudden perturbation

Modeling results

We fitted the state-space model and simple DIVA model to the early responses of each participant to estimate each computational model's parameters. For each individual, the average of early adaptation responses was calculated. Figure 4 shows the group average simulated for gradual and sudden perturbation in both shift and clamp conditions for the state-space model.

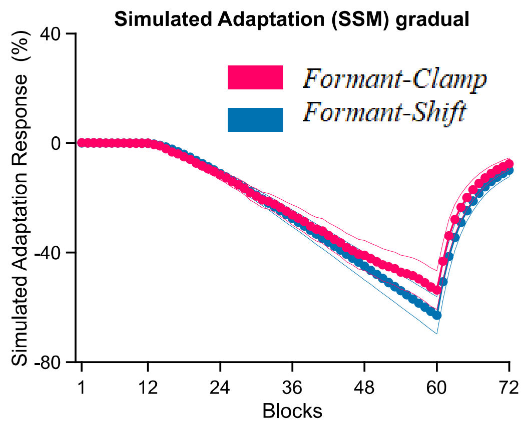

Figure 4. Simulated adaptation responses for gradual and sudden perturbations of formant shift and formant clamp. These results are based on the simulation of the state-space model.

Based on the fitted model for each participant, we extracted the model's parameters. These results are shown in Figure 5. Each grey line shows one participant, and the blue dots indicate the group average in each condition. Based on the fitted model for each participant, we extracted the model's parameters. These results are shown in Figure 5. Each grey line shows one participant, and the blue dots indicate the group average in each condition. These results showed that the prediction sensitivity was higher for formant shift (gradual or sudden) than for formant clamp ($p = .026$). The results for feedforward error sensitivity did not show statistically significant results as the parameter appeared to be the same in all conditions.

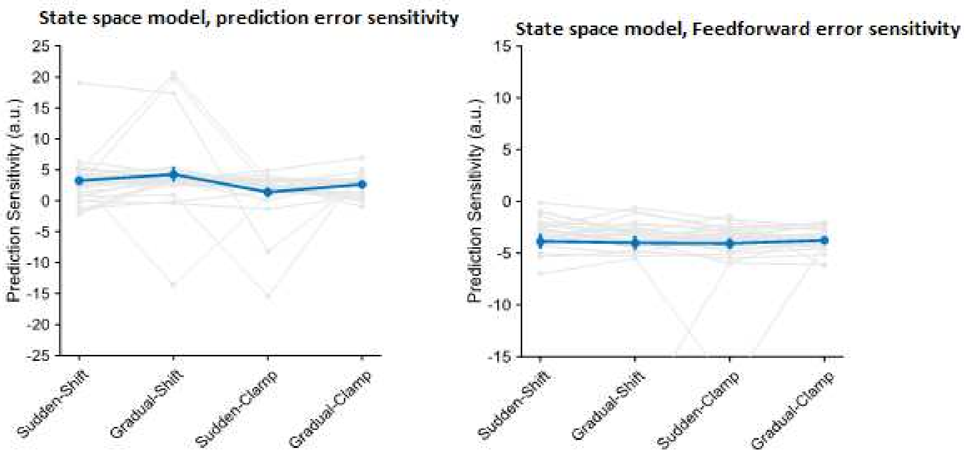

Figure 5. Model parameters (prediction sensitivity and feedforward error sensitivity) of the state-space model for all participants and all conditions.

We also fitted the simple DIVA to the data. Figures 6 and 7 show the average simulated data and model parameters in the four perturbation conditions. Similar to parameters of the state-space model, these model parameters were between 0 and 1, so we transformed the parameters using a logit transform to normalize the distribution. The simple DIVA model can accurately generate adaptive responses; however, this model did not generate adaptive responses for the formant-clamp as accurately as the responses for the formant-shift responses.

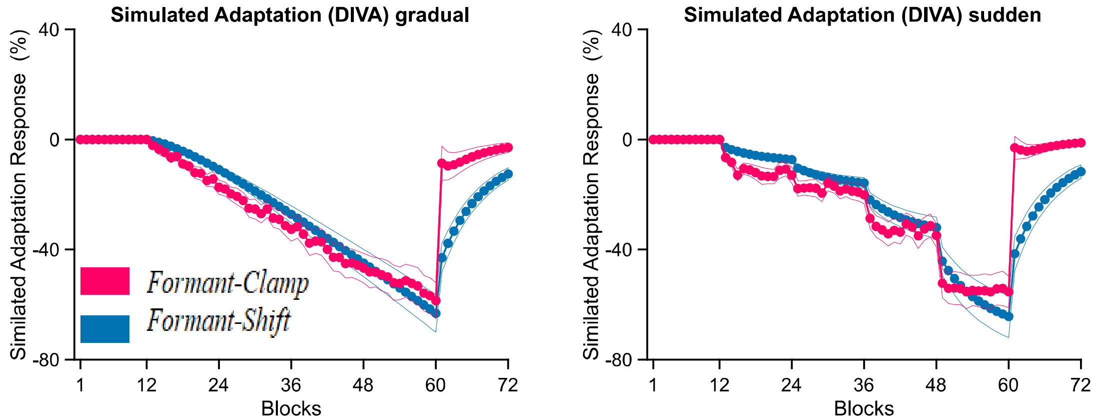

Figure 6. Simulated adaptation responses for gradual and sudden perturbations of formant shift and formant clamp. These results are based on the simulation of the simpleDIVA model.

The simple DIVA, unlike the state-space model, includes both auditory and somatosensory feedback. Therefore, the model would allow us to estimate the sensitivity of both the auditory and somatosensory systems to errors. Figure 7 shows all three parameters of the simple DIVA in all conditions. As shown in this figure, the learning rate was lower for formant shift, and both auditory and somatosensory feedback has higher sensitivity in the formant clamp conditions. While the results for the auditory gain were statistically significant, these results were not significant for prediction sensitivity and somatosensory sensitivity.

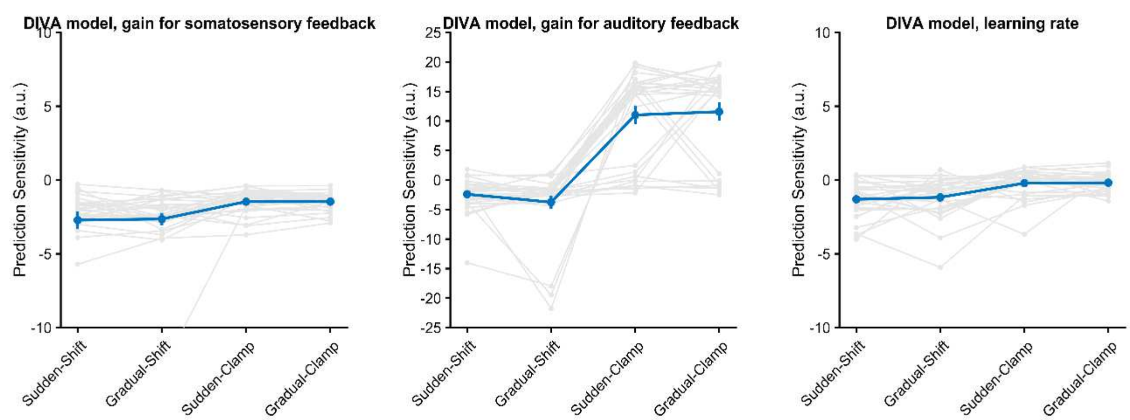

Figure 7. Model parameters (learning rate and somatosensory and auditory feedback gain) of the simple DIVA model for all participants and all conditions.

DISCUSSION

In this study, two computational models were used to estimate different parameters attributed to the speech motor system by fitting each model to a particular data set. We used an auditory perturbation experiment to examine whether the state-space model's and the simple DIVA model's parameters can estimate the speech control system's different attributes: feedforward error sensitivity, prediction sensitivity, learning rate, and somatosensory-auditory gains. Participants completed adaptation tasks in which their /ɛ/ formant was shifted toward their /æ/. In this paper, only early adaptive responses were measured, which is the main contributing factor to the feedforward control system. We fitted the state-space model and simple DIVA model to each participant's adaptive responses and extracted all models' parameters. To test whether each model accounts for its parameters, we examined the simulated adaptive responses based on the models.

By examining the adaptation responses, we found that adaptation responses do not change in either clamp or shift condition in the gradual adaptation responses except in blocks 48-60, which has the highest perturbation magnitude. The mean adaptation response in the clamp condition was less than the formant shift condition in these trials (48-60). Since the amount of perturbation is highest in this block, participants might notice the difference between the clamp and shift. For example, they may evaluate the clamp feedback as less relevant feedback that is not their own production, and thus, they may respond less to the feedback perturbation. Moreover, in the formant shift perturbation, in these blocks, when the participants hear their shifted speech production, they may rely on the feedback more (e.g., evaluate it as their own). Examining the individual data showed that most participants adapt slightly more in the gradual shift than the gradual, sudden condition in all six blocks. In the first 36 trials (blocks 1-12), the auditory prediction error is zero, but when the perturbation was applied in the 37th trial, the predicted error and received perturbation are not equal anymore. Thus, participants will have larger auditory prediction errors, and the feedback control system may change its prediction and generate a larger response. As we applied sudden/gradual perturbation throughout the trials, the feedforward commands are gradually updated to minimize auditory prediction errors. In the end phase (the last 36 trials, 60 to 72 blocks), there is a large prediction error when the perturbation is removed, leading to a large response.

Our analysis of the state-space model showed that the model could predict adaptive responses in all perturbation conditions. However, a closer examination of the simulated responses showed that the model more closely predicted the responses in the gradual conditions and in the formant shift conditions. The model, similar to the empirical data, showed that simulated adaptation responses were higher in the shift condition than in the clamp condition in blocks 48-60. In general, the simulated gradual adaptation responses were closer to the empirical data than adaptation responses in the sudden condition. Examining the model parameters showed that feedforward error sensitivity remained the same across all four perturbation conditions. One interpretation of these results is that the speech motor system's sensitivity to error does not change in different conditions, and thus, it responds to the errors in the same way. The prediction sensitivity parameter was different across conditions: (1) prediction sensitivity was higher for formant shift conditions than formant clamp conditions, and (2) prediction sensitivity was higher in gradual conditions than sudden conditions. We speculate that the speech motor system uses its evaluation of sensory feedback in previous trials more strongly to determine the magnitude of its responses than its error sensitivity. One limitation of the state-space model is that it did not include somatosensory feedback. Another limitation of the model was that the simulated responses were only based on early adaptive responses and did not include the corrective responses during speech production. Overall, these limitations can be addressed in future studies by including feedback sensitivity and corrective responses in the analysis.

Similar to the state-space model, the simple DIVA model could predict adaptive responses in all perturbation conditions; however, there were several noticeable differences between the performance of the two models. The first major difference is related to the accuracy of the simpleDIVA in predicting the formant clamp conditions. The model's accuracy appears to be lower than its accuracy of predicting the formant shift conditions. We speculate that this difference is due to the structure of the simpleDIVA, as the model uses feedback responses to update the feedforward responses. Another major difference between the two models is related to how suddenly the simulated responses of the simple DIVA change in the sudden conditions. It appears that the simpleDIVA has a higher learning rate and is more sensitive to errors, and responds more quickly to the errors. Examining the parameters, we found that auditory feedback gain was the most sensitive measure and could differentiate the formant clamp from the formant shift. However, this was not the case for the somatosensory feedback gain. One explanation for this pattern of the results is that the experiment only included auditory feedback perturbation, and somatosensory feedback was not perturbed. As a result, the model's parameter related to the somatosensory feedback was not influenced significantly. Interesting, the learning rate was similar across the perturbation conditions.

This result was similar to the result for the feedforward error sensitivity of the state-space model.

Overall, the simple DIVA model accurately predicted the results, and the performance of the two models was similar. The models' parameters provided additional insight into how the speech motor system responded to various auditory perturbations. These results have important implications for speech motor learning and production models and shed light on neural processes involved in generating adaptive responses. For example, by studying different parameters, we could determine how speech disorders influence the speech motor system and which components are more influenced by the disorders. Then we can develop behavioral or neural treatments to target those components more efficiently.

REFERENCES

B. S. Lee. (1950). Some effects of side-tone delay. Acoust. Soc. J., 22, 639–640 (1950) Am.

Berg, T., & Levelt, W. J. M. (1990). Speaking: From Intention to Articulation. The American Journal of Psychology, 103(3), 409. https://doi.org/10.2307/1423219

Cowie, R., & Douglas-Cowie, E. (1992). Postlingually acquired deafness: Speech deterioration and the wider consequences. In Postlingually Acquired Deafness: Speech Deterioration and the Wider Consequences. https://books.google.com/books?hl=en&lr=&id=jVYyB6CtvfcC&oi=fnd&pg=PA1&dq=Postlingually+Acquired+Deafness:+Speech+Deterioration+and+the+Wider+Consequences.+Hawthorne,+NY:+Mouton+de+Gruyter.&ots=7VaQbl4key&sig=sQvrh3vknOWOA67qF2CMHpf5Ebg#v=onepage&q&f=fals

Daliri, A. (2021). A Computational Model for Estimating the Speech Motor System's Sensitivity to Auditory Prediction Errors. Journal of Speech, Language, and Hearing Research, 64(6), 1841–1854. https://doi.org/10.1044/2021_JSLHR-20-00484

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

Daliri, A., & Dittman, J. (2019a). 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., & Dittman, J. (2019b). 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. (2015). 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., Wieland, E. A., Cai, S., Guenther, F. H., & Chang, S. E. (2018). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. Developmental Science, 21(2), 1–11. https://doi.org/10.1111/desc.12521

Fuchs, S., Cleland, J., & Rochet-Capellan, A. (2019). speech production and-perception learning and memory.

Guenther, F. H. (2016). Neural Control of Speech - Frank H. Guenther - Google Books. https://books.google.com/books?hl=en&lr=&id=aRSvDAAAQBAJ&oi=fnd&pg=P

R7&dq=Guenther,+F.+H.+(2016).+Neural+control+of+speech.+MIT+Press.&ots=ObaBg6LbC0&sig=YvI65TVyNpsSXytreqYUPgepQW8#v=onepage&q=Guenther%2C F. H. (2016). Neural control of speech. MIT Press

Guenther, Frank H. (2006). Cortical interactions underlying the production of speech sounds.
Journal of Communication Disorders, 39(5), 350–365. https://doi.org/10.1016/j.jcomdis.2006.06.013

Guenther, Frank H. (2016). Neural control of speech. MIT Press.
Guidelines for Manual Pure-Tone Threshold Audiometry. (2005). https://doi.org/10.1044/policy.GL2005-00014

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(OCTOBER), 1–14. https://doi.org/10.3389/fnhum.2011.00082

Ingo Titze. (1994). Principles of voice production (Book, 1994) [WorldCat.org]. Englewood Cliffs, N.J. : Prentice Hall, ©1994. https://www.worldcat.org/title/principles-of-voice-production/oclc/925118893

Kearney, E., Nieto-Castañón, A., Weerathunge, H. R., Falsini, R., Daliri, A., Abur, D., Ballard, K. J., Chang, S. E., Chao, S. C., Heller Murray, E. S., Scott, T. L., & Guenther, F. H. (2020). A Simple 3-Parameter Model for Examining Adaptation in Speech and Voice Production. Frontiers in Psychology, 10(January), 1–16. https://doi.org/10.3389/fpsyg.2019.02995

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

Oller, D. K., & Eilers, R. E. (1988). The Role of Audition in Infant Babbling. Child Development, 59(2), 441. https://doi.org/10.2307/1130323

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. https://doi.org/10.1523/JNEUROSCI.3363-16.2017

Parrell, B., & Houde, J. (2019a). Modeling the role of sensory feedback in speech motor control and learning. In Journal of Speech, Language, and Hearing Research (Vol. 62, Issue 8S, pp. 2963–2985). American Speech-Language-Hearing Association. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0127

Parrell, B., & Houde, J. (2019b). Modeling the role of sensory feedback in speech motor control and learning. In Journal of Speech, Language, and Hearing Research (Vol. 62, Issue 8S, pp. 2963–2985). American Speech-Language-Hearing Association. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0127

Shadmehr, R., & Mussa-Ivaldi, S. (2013). Biological Learning and Control. In Biological Learning and Control. The MIT Press. https://doi.org/10.7551/mitpress/9780262016964.001.0001

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