



## OPEN ACCESS

## EDITED BY

Jeffery A. Jones,  
Wilfrid Laurier University, Canada

## REVIEWED BY

David Jackson Morris,  
University of Copenhagen, Denmark  
Nishant Rao,  
Haskins Laboratories, United States

## \*CORRESPONDENCE

Elaine Kearney  
ekearney@bu.edu

## SPECIALTY SECTION

This article was submitted to  
Speech and Language,  
a section of the journal  
Frontiers in Human Neuroscience

RECEIVED 27 April 2022

ACCEPTED 04 October 2022

PUBLISHED 02 November 2022

## CITATION

Kearney E, Nieto-Castañón A,  
Falsini R, Daliri A, Heller Murray ES,  
Smith DJ and Guenther FH (2022)  
Quantitatively characterizing reflexive  
responses to pitch perturbations.  
*Front. Hum. Neurosci.* 16:929687.  
doi: 10.3389/fnhum.2022.929687

## COPYRIGHT

© 2022 Kearney, Nieto-Castañón,  
Falsini, Daliri, Heller Murray, Smith and  
Guenther. This is an open-access  
article distributed under the terms of  
the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution  
or reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Quantitatively characterizing reflexive responses to pitch perturbations

Elaine Kearney<sup>1\*</sup>, Alfonso Nieto-Castañón<sup>1,2</sup>,  
Riccardo Falsini<sup>1</sup>, Ayoub Daliri<sup>3</sup>, Elizabeth S. Heller Murray<sup>4</sup>,  
Dante J. Smith<sup>5</sup> and Frank H. Guenther<sup>1,6,7</sup>

<sup>1</sup>Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA, United States, <sup>2</sup>The McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, United States, <sup>3</sup>College of Health Solutions, Arizona State University, Tempe, AZ, United States, <sup>4</sup>Department of Communication Sciences and Disorders, Temple University, Philadelphia, PA, United States, <sup>5</sup>Graduate Program for Neuroscience, Boston University, Boston, MA, United States, <sup>6</sup>Department of Biomedical Engineering, Boston University, Boston, MA, United States, <sup>7</sup>The Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, MA, United States

**Background:** Reflexive pitch perturbation experiments are commonly used to investigate the neural mechanisms underlying vocal motor control. In these experiments, the fundamental frequency—the acoustic correlate of pitch—of a speech signal is shifted unexpectedly and played back to the speaker via headphones in near real-time. In response to the shift, speakers increase or decrease their fundamental frequency in the direction opposing the shift so that their perceived pitch is closer to what they intended. The goal of the current work is to develop a quantitative model of responses to reflexive perturbations that can be interpreted in terms of the physiological mechanisms underlying the response and that captures both group-mean data and individual subject responses.

**Methods:** A model framework was established that allowed the specification of several models based on Proportional-Integral-Derivative and State-Space/Directions Into Velocities of Articulators (DIVA) model classes. The performance of 19 models was compared in fitting experimental data from two published studies. The models were evaluated in terms of their ability to capture both population-level responses and individual differences in sensorimotor control processes.

**Results:** A three-parameter DIVA model performed best when fitting group-mean data from both studies; this model is equivalent to a single-rate state-space model and a first-order low pass filter model. The same model also provided stable estimates of parameters across samples from individual subject data and performed among the best models to differentiate between subjects. The three parameters correspond to gains in the auditory feedback controller's response to a perceived error, the delay of this response, and the gain of the somatosensory feedback controller's "resistance" to this correction. Excellent fits were also obtained from a four-parameter model with an additional auditory velocity error term; this model was

better able to capture multi-component reflexive responses seen in some individual subjects.

**Conclusion:** Our results demonstrate the stereotyped nature of an individual's responses to pitch perturbations. Further, we identified a model that captures population responses to pitch perturbations and characterizes individual differences in a stable manner with parameters that relate to underlying motor control capabilities. Future work will evaluate the model in characterizing responses from individuals with communication disorders.

#### KEYWORDS

computational modeling, motor control, speech production, pitch, auditory feedback

## Introduction

Auditory perturbation paradigms have become an important experimental approach in uncovering the neural mechanisms underlying vocal motor control. First described by [Elman \(1981\)](#), these paradigms involve manipulating the frequency spectrum of someone's speech and playing it back to them via headphones in near real-time, such that they—often subconsciously—detect an error in their production. In pitch perturbation experiments specifically, the frequency spectrum is perturbed so that the fundamental frequency ( $f_0$ ; the acoustic correlate of pitch) is higher or lower than produced. In response to this manipulation, speakers will change their  $f_0$  in the direction opposite the perturbation, which makes what they hear in the headphones closer to what they intended to produce. When the perturbations are unexpected (for example, when applied randomly on a small percentage of trials or when applied at a random time during each trial), the compensatory response is referred to as reflexive; that is, the response is evident within a given perturbed trial but has a limited effect on subsequent trials. This contrasts with perturbations sustained over many trials that elicit both reflexive within-trial responses as well as adaptive across-trial responses ([Daliri, 2021](#)). The current work focuses on reflexive responses to pitch perturbations; we will use the term *pitch shift reflex* (PSR) to refer to such responses ([Kiran and Larson, 2001](#)).

There is a long history of utilizing reflexive responses as a diagnostic tool for probing neural function. For example, the pupillary light reflex was used by Claudius Galenus in the 2nd century to evaluate the visual capabilities of candidates for cataract surgery (see [Thompson, 2003](#) for a historical review). Since that time, scientists have characterized the pupillary light reflex in ever-increasing detail, and modern investigations often utilize pupillography to accurately measure the time course of the pupil's reaction to changes in light input. These studies have led to the parameterization of the temporal

profile of the pupillary light reflex (e.g., [Hall and Chilcott, 2018](#)) as well as parameterized mathematical models of the dynamics of the pupillary light reflex that capture individual differences ([Pamplona, 2008](#)). The different parameters in these characterizations correspond to different neural processes; thus, an individual's pupillary light reflex can be used to differentiate damage to one part of the nervous system from damage to another, in turn allowing clinicians to make informed decisions regarding treatment options. The dynamics of the pupillary light reflex are now used to gauge neural function in a wide range of disorders extending beyond impairment of the visual system, including concussion ([Master et al., 2020](#)), schizophrenia ([Bär et al., 2008](#)), Alzheimer's disease ([Tales et al., 2001](#)), Parkinson's disease ([Stergiou et al., 2009](#)), autism spectrum disorders ([Fan et al., 2009](#)), and alcoholism ([Rubin, 1980](#)). Against this background, a primary goal of the current study is to mathematically characterize the pitch reflex response using mathematical models with parameters that reflect the function of different neural subsystems involved in the control of voice.

Since the early application of the pitch perturbation paradigm, over 140 studies have used this paradigm to investigate various aspects of vocal motor control and across different populations. These studies have revealed several properties of the PSR. First, responses are typically in the direction opposite the perturbation, while a small percentage of responses occur in the same direction as the perturbation (e.g., [Burnett et al., 1998](#); [Franken et al., 2018](#)). Second, the compensation is usually incomplete, likely reflecting an interaction between the auditory and somatosensory control systems ([Smith et al., 2020](#)). Third, the responses occur in a variety of speech stimuli ([Natke and Kalveram, 2001](#); e.g., sustained vowels, syllables, running speech; [Chen et al., 2007](#); [Smith et al., 2020](#)). In addition, investigations of the PSR in speakers of tonal languages, such as Mandarin, show an interaction between the linguistic intent of an utterance

and perturbations, with larger responses evident when the perturbation changes the meaning of a word (Xu et al., 2004). Musicians and singers, who have higher-than-average experience controlling pitch, are also able to ignore large pitch perturbations ( $\sim 200$  cents) while they compensate more completely for smaller and shorter perturbations ( $\sim 25$  cents) (Zarate et al., 2010; Behroozmand et al., 2014; Parkinson et al., 2014).

While the majority of pitch-perturbation studies to date have focused on neurotypical adult speakers, a growing number of studies have examined responses in children and individuals with communication disorders. Reflexive perturbation responses in children are evident as young as age 3 years (Russo et al., 2008; Scheerer et al., 2013, 2016; Heller Murray and Stepp, 2020) but are associated with longer response latencies and greater variability compared to adult responses. Studies have also investigated responses in individuals with Parkinson's disease (Kiran and Larson, 2001; Liu et al., 2012; Abur et al., 2021a), Alzheimer's disease (Ranasinghe et al., 2017), cerebellar degeneration (Houde et al., 2019; Li et al., 2019), apraxia of speech (Ballard et al., 2018), aphasia (Behroozmand et al., 2018, 2022), hyperfunctional voice disorders (Abur et al., 2021b), 16p11.2 deletions (Demopoulos et al., 2018), autism (Russo et al., 2008), and in those who stutter (Loucks et al., 2012; Sares et al., 2018, 2020). Collectively, these studies shed light on the development of vocal motor control and the mechanisms underlying speech and voice disorders. In the future, these findings may inform novel treatments that directly target these mechanisms.

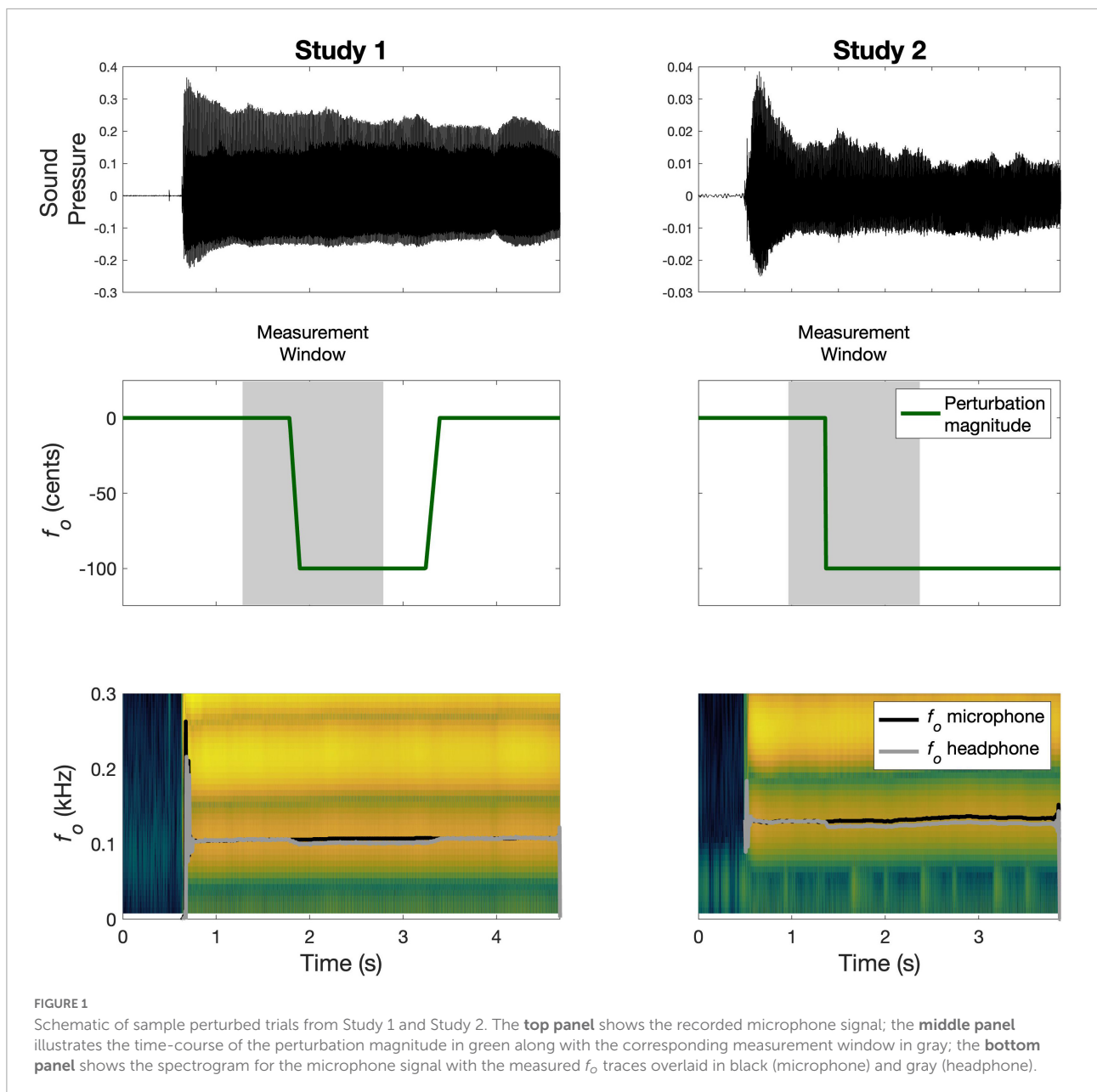
Compensatory responses to pitch perturbations rely on neural processes that compare the target pitch for a given utterance to the pitch as sensed through audition and apply corrections if and when an error is detected. We can use computational models to explicate these internal processes by specifying the processes with mathematical equations and evaluating how well the equations (i.e., the models) explain existing experimental data. There are several candidate model classes that may be used to model reflexive pitch perturbation data, including *Proportional-Integral-Derivative (PID)*, *State-Space (SS)*, and *Directions Into Velocities of Articulators (DIVA)* models.

The PID model class was originally designed to mimic the steering strategy used by expert ship helmsmen (Minorsky, 1922) and is now commonly used in a wide range of engineering applications. This model class includes proportional (P) models, where the corrective command is proportional to the error signal; proportional-derivative (PD) models, where the proportional command is supplemented with a command that is formed by multiplying the derivative of the error signal by a gain; proportional-integral (PI) models, where the proportional command is supplemented with a command formed by multiplying the integral of the error signal by a gain; and finally, PID models that combine all three error

terms. SS models also originated in control engineering and have been widely applied in studies of limb motor control (Thoroughman and Shadmehr, 2000; Smith et al., 2006; Galea et al., 2015; Huberdeau et al., 2015). SS models model physical systems as a set of input, output, and state variables using first-order differential equations. The DIVA model is a prominent neural network model of speech motor control (Guenther, 2016; Kearney and Guenther, 2019). It is organized around three control subsystems, namely feedforward, auditory feedback, and somatosensory feedback control, and has been used to explain a wide number of speech phenomena. Although the SS and DIVA models have different theoretical motivations, they are closely related mathematically (as we will demonstrate) and will be treated together throughout this paper.

To the best of our knowledge, only one study to date has utilized a computational model to simulate responses to a reflexive pitch perturbation paradigm (Larson et al., 2000). Larson et al. (2000) implemented a model in which the  $f_o$  error was computed as the difference between the target  $f_o$  and actual  $f_o$  (following a 130 ms processing delay), partially integrated via a low-pass filter, and applied to the output. The model simulations were compared graphically to experimental data and approximated the overall timing and shape of the observed responses. The authors acknowledged that the model was likely an over-simplification of the underlying processes but nonetheless showed promise and feasibility for computational modeling of reflexive perturbation data. The current study extends this work by investigating a variety of models that utilize different numbers of free parameters to quantitatively fit pitch shift responses measured experimentally.

Both SS and DIVA models have been successfully used to simulate responses to adaptive perturbation paradigms (Daliri and Dittman, 2019; Kearney et al., 2020). Daliri and Dittman (2019) implemented an SS model with two parameters (an internal estimate forgetting factor and a sensory error weighting factor) that showed good fits to experimental data. Kearney et al. (2020) developed SimpleDIVA—a simplified version of the DIVA model—with three parameters that correspond to gains in the key subsystems involved in speech motor control (auditory feedback, somatosensory feedback, and feedforward control). SimpleDIVA also provides good fits to experimental data and is able to account for a number of variations in the sensorimotor adaptation paradigm (e.g., perturbing more than one dimension or using masking noise). An additional benefit of SimpleDIVA is that the model's parameters provide a mechanistic explanation of behavioral responses in terms of the neural control systems believed to be involved in controlling speech production. These adaptive models, however, are not immediately applicable to reflexive response data as the mechanisms underlying the responses are not the same. Specifically, because we do not expect trial-to-trial learning in a reflexive experiment (Daliri et al., 2020; cf. Hantzsch et al., 2022), we examine the within-trial responses averaged over



all perturbed trials in an experiment. Examining within-trial responses also means that we need to account for latencies associated with processing delays.

Several earlier PSR studies have observed that the compensatory response could occur on more than one time scale, resulting in a complex or multi-peaked response (Burnett et al., 1997, 1998; Larson, 1998; Hain et al., 2000). The first peak was described as a short-latency, rapid response occurring around 100–225 ms, and the second as a long-latency, slow response occurring around 250–600 ms. The simplest form of the DIVA/SS model produces only a single response peak. For this reason, we also investigate generalized versions of the

DIVA/SS model that are better able to capture multi-component responses.

To address our primary goal of developing a quantitative model of the PSR, we established a model framework that allows the specification of several model variations based on PID and SS/DIVA model classes. The performance of the different models was then compared by fitting them to datasets from two prior PSR studies (Heller Murray and Stepp, 2020; Smith et al., 2020). We operationally defined model validity in terms of the ability to capture population-level responses to pitch perturbation experiments as well as individual differences in sensorimotor control processes. That is, a valid model should be able to (1) explain group mean responses to pitch perturbations,

(2) have parameters that are stable across samples from an individual subject, and (3) have parameters that differentiate between individual subjects.

## Materials and methods

Our overall approach is to mathematically define a number of control models that each involve optimizable parameters. Each model generates a time series of  $f_o$  values, denoted by the variable  $f(t)$ , where  $t$  ranges from 0 to the trial length of the experiment being modeled. A particle swarm optimization procedure is used to find the optimal parameter values [in terms of minimizing root-mean-square error (RMSE)] for each model when fitting a particular data set, and the resulting fit is characterized in terms of RMSE, Akaike information criterion (AIC), and cross-validated classification scores. Model fits were performed separately for two datasets from different studies involving unpredictable perturbations of  $f_o$  (Heller Murray and Stepp, 2020; Smith et al., 2020) applied during extended vowel productions of young healthy adult speakers.

## Datasets

In Study 1 (Smith et al., 2020), a group of English speakers ( $N = 18$ ; aged 18–34) completed 80 trials, during which they sustained the vowel /a/ for four seconds. On a quarter (20) of all trials, an auditory perturbation of –100 cents was applied at a jittered point in time, 1,000–1,500 ms after the beginning of the trial. The perturbation was implemented as a time-domain/formant-adjusted shift using Audapter software (Cai et al., 2008); this process shifts only  $f_o$  while preserving the produced formants. The perturbation onset was characterized by a linear ramp that took 110 ms to reach the full perturbation magnitude. The perturbation remained on for a further 1,000–1,500 ms. The order of perturbed and control trials was pseudorandomized, with no consecutively perturbed trials.  $f_o$  trajectories (Hz) were extracted for the duration of the vowel using Praat (Boersma and Weenink, 2018), and then time-aligned to the beginning of the perturbation and parsed from –500 to +1500 ms in MATLAB. A schematic of a sample perturbed trial and corresponding data is shown in **Figure 1**. The data were normalized to the average of each subject's baseline. On average, subjects compensated for 48.8% ( $SD$ : 20.8) of the perturbation, calculated as change from baseline to the last 250 ms of a trial and expressed as a percentage of the maximum perturbation magnitude.

In Study 2, a group of English speakers ( $N = 20$ ; aged 18–28) completed 60 trials, during which they sustained the vowel /i/ for 3 s (Heller Murray and Stepp, 2020). On each trial, an auditory perturbation of +100 cents or –100 cents was applied at a jittered point in time, 500–1,000 ms after voice onset.

The perturbation was implemented as a full-spectrum shift by shifting the values and spacing of the vocal harmonics using Eventide Eclipse hardware (Eventide Inc., Little Ferry, NJ, USA; Heller Murray et al., 2019), thus shifting  $f_o$ . The perturbation onset was characterized by a step function (or sudden onset) and, once applied, the perturbation remained on for the rest of the trial. All trials in the experiment were perturbed, and the direction of the perturbation was pseudorandomized to ensure that no more than five consecutive trials were perturbed in the same direction. The intertrial interval was also jittered between 500 and 1000 ms to reduce anticipation of the next trial.  $f_o$  trajectories (Hz) were extracted for the duration of the vowel using Praat (Boersma and Weenink, 2018), and then time-aligned to the beginning of the perturbation and parsed from –400 to +1400 ms in MATLAB. A schematic of a sample perturbed trial and corresponding data is shown in **Figure 1**. To fit the models to data from both perturbation directions together, all data were normalized by dividing by each subject's baseline average, and then flipping the upshift data around the  $x$ -axis. On average, subjects compensated for 17.1% ( $SD$ : 14.4) of the perturbation, calculated as change from baseline to the last 250 ms of a trial and expressed as a percentage of the maximum perturbation magnitude.

## Assumptions and definitions for all control models

We use the variable  $f_T$  to represent the value of  $f_o$  that the controller is attempting to achieve; we assume this target is constant for a given speaker rather than a function of time since the experimental task being modeled involves attempting to maintain a constant pitch, and we equate  $f_T$  to the average  $f_o$  of the speaker prior to the onset of the perturbation (i.e., during the *baseline period* between 0 and 500 ms for Study 1 and 0 to 400 ms for Study 2). Next, we assume that the output of the controlled *plant* (corresponding to the vocal tract articulators and musculature) is updated based on the signal provided by the controller at each time point<sup>1</sup> as follows:

$$f(t) = f_T + \int_{\delta=0}^t \dot{f}_C(\delta) d\delta \quad (1)$$

where  $f(t)$  is the plant output (i.e., the actual  $f_o$  produced by the subject) at time  $t$ ,  $\dot{f}_C(t)$  is the controller output at time  $t$ , and  $\delta$  is a dummy variable for integration. This controller output represents a corrective command in response to the perceived error at time  $t$ . During the baseline period,  $\dot{f}_C(t)$  is set to 0 for all models, and the baseline period is accordingly not included in RMSE calculations.

<sup>1</sup> We use continuous time notation here for clarity, although the simulations utilize a discrete time representation with one time point per data sample. The data modeled here were sampled at 200 Hz; accordingly, the simulations utilize 5 ms time steps.

The auditory feedback of the produced sound available to the controller, corresponding approximately to the auditory cortical representation of the pitch/ $f_o$  of the produced sound, is defined as follows:

$$f_A(t) = f(t - \tau_A) \cdot (1 + P(t - \tau_A)) \quad (2)$$

where  $\tau_A$  is a delay parameter that is optimized (along with other model parameters) to fit a particular dataset, and  $P(t)$  is the size of the perturbation applied at time point  $t$ , expressed as a percentage of  $f(t)$  in decimal form (e.g.,  $P = 0.06$  corresponds to a 6% upward perturbation of  $f_o$ ). The delay  $\tau_A$  represents the combined delay of the perturbation processing software and hardware and the total neural processing delay from the auditory periphery to the corresponding motor output in the auditory feedback control system.

The DIVA model also includes a somatosensory representation of  $f_o$ , assumed to derive from laryngeal mechanoreceptors, which is related to the actually produced  $f_o$  as follows:

$$f_S(t) = f(t - \tau_S) \quad (3)$$

where  $\tau_S$  is a delay parameter (corresponding roughly to the transmission delay from the somatosensory periphery to somatosensory cortex) that can be optimized (along with other model parameters) to fit a particular dataset. This somatosensory representation can be shown to be closely related to the parameter  $A$  in a typical state-space model, which weights the degree to which the current state of the system contributes to the next state (see *Basic DIVA equation* below).

## Proportional-integral-derivative equation

A PID controller is defined by the following equation:

$$\begin{aligned} \dot{f}_C(t) = & \alpha_P \cdot (f_T - f_A(t)) + \alpha_I \\ & \cdot \int_{\delta=0}^t (f_T - f_A(\delta)) d\delta + \alpha_D \\ & \cdot \frac{d}{dt}(f_T - f_A(t)) \end{aligned}$$

which simplifies to:

$$\dot{f}_C(t) = \alpha_P \cdot (f_T - f_A(t)) + \alpha_I \cdot \int_{\delta=0}^t (f_T - f_A(\delta)) d\delta - \alpha_D \cdot \dot{f}_A(t) \quad (4)$$

where  $\alpha_P$ ,  $\alpha_I$ , and  $\alpha_D$  are optimizable gains for the position, integral, and derivative terms. We will simulate four models using this equation: a proportional model (P) in which  $\alpha_I$  and  $\alpha_D$  are fixed at 0, a proportional-integral (PI) model in which  $\alpha_D$

is fixed at 0, a proportional-derivative (PD) model where  $\alpha_I$  is fixed at 0, and a proportional-integral-derivative (PID) model in which all parameters are optimized.

## Basic directions into velocities of articulators/state-space equation

The DIVA model's feedback controller consists of both auditory and somatosensory feedback control components. The standard formulation of the DIVA model's feedback controller is:

$$\dot{f}_C(t) = \alpha_A \cdot (f_T - f_A(t)) + \alpha_S \cdot (f_T - f_S(t)) \quad (5)$$

where  $\alpha_A$  and  $\alpha_S$  are parameters denoting the gains of the auditory and somatosensory feedback control systems, respectively, and  $\tau_S$  is a delay parameter corresponding to the delay between an action and the corresponding somatosensory feedback signal in somatosensory cortex. When  $\tau_S$  is set to 0 [and therefore  $f_S(t) = f(t)$ ; see EQ3], EQ5 is mathematically equivalent to the following SS model<sup>2</sup>:

$$\dot{f}_C(t) = A \cdot f_C(t) + B \cdot (f_T - f_A(t))$$

where  $f_C(t) = f(t) - f_T$  (see Eq. 1),  $B$  is equal to  $\alpha_A$  in EQ5, and  $A$  is equal to  $-\alpha_S$  in EQ5. Preliminary simulations of the two models verified this mathematical equivalence and also indicated nearly identical performance for generalized versions of the DIVA/SS models described below. The model of EQ5 is also equivalent to the low-pass filter or "leaky integrator" model proposed by Larson et al. (2000), which is a special case of EQ5 with  $\alpha_A = \alpha_S$  and the time constant of the low-pass filter equal to our time step size (0.005 s) times  $1/\alpha_S$ . For simplicity, we will use the DIVA-based formulations for simulations herein as it provides a more direct physiological interpretation of model parameters than the SS or Larson et al. (2000) formulations.

## Generalized directions into velocities of articulators/state-space equations

The model of EQ5 can be generalized to include an  $f_o$  velocity target in addition to the  $f_o$  position target as follows:

$$\begin{aligned} \dot{f}_C(t) = & \alpha_A \cdot (f_T - f_A(t)) + \alpha_{Av} \cdot (\dot{f}_T - \dot{f}_A(t - \tau_{Av})) + \alpha_S \\ & \cdot (f_T - f_S(t)) + \alpha_{Sv} \cdot (\dot{f}_T - \dot{f}_S(t - \tau_{Sv})) \end{aligned}$$

<sup>2</sup> This model is also equivalent to the SS model of sensorimotor adaptation posed by Daliri and Dittman (2019) with the parameter  $a$  from that model equal to  $1 - \alpha_S$ , parameter  $b$  equal to  $\alpha_A$ , and removal of  $\tau_A$  from the current model since sensorimotor adaptation data were modeled on a trial-by-trial basis rather than a timepoint-by-timepoint basis by Daliri and Dittman. The model is also equivalent to a leaky integrator of the error signal with the leak rate parameter equal to  $\alpha_S$ .

where  $\dot{f}_T$  is the target velocity,  $\alpha_{Av}$  and  $\alpha_{Sv}$  are the auditory and somatosensory feedback control gains of the velocity-based response component, respectively, and  $\tau_{Av}$  and  $\tau_{Sv}$  represent the differential delays between the position and velocity components. Because subjects in the experiments being modeled were instructed to maintain a constant pitch,  $\dot{f}_T$  is set to 0 and this equation reduces to:

$$\dot{f}_C(t) = \alpha_A \cdot (f_T - f_A(t)) - \alpha_{Av} \cdot \dot{f}_A(t - \tau_{Av}) + \alpha_S \cdot (f_T - f_S(t)) - \alpha_{Sv} \cdot \dot{f}_S(t - \tau_{Sv}) \tag{6}$$

This characterization is approximately equivalent (though not identical) to a two-state (position and velocity error) SS model.

Alternatively, the model of EQ5 can be generalized to allow two different position-error-based responses that operate at different delays:

$$\dot{f}_C(t) = \alpha_A \cdot (f_T - f_A(t)) + \alpha_{As} \cdot (f_T - f_A(t - \tau_{As})) + \alpha_S \cdot (f_T - f(t)) + \alpha_{Ss} \cdot (f_T - f_S(t - \tau_{Ss})) \tag{7}$$

where  $\alpha_A$  and  $\alpha_S$  are the auditory and somatosensory feedback control gains of the faster response component,  $\alpha_{As}$  and  $\alpha_{Ss}$  are the auditory and somatosensory feedback control gains of the slower response component, and  $\tau_{As}$  and  $\tau_{Ss}$  represent the differential delay between the fast and slow components ( $\tau_{As}, \tau_{Ss} = 0$ ). In effect, this model is a quantification of the idea that the response to a pitch perturbation includes a relatively fast, automatic component (captured by the terms involving  $\alpha_A$  and  $\alpha_S$ ) and a slower component (captured by the terms involving  $\alpha_{As}$  and  $\alpha_{Ss}$ ) that may be under more conscious control than the faster component (Burnett et al., 1997, 1998; Larson, 1998; Hain et al., 2000). This characterization is also approximately equivalent to a two-state (fast and slow position error) SS model.

### Model versions used in simulations

A total of 19 different models were tested: 4 based on PID control (models P, PI, PD, and PID) and 15 based on the DIVA model and equivalent or near-equivalent state-space formulations (D1–D15). Table 1 lists the equations and optimized parameters for all models. All unused parameters from an equation were set to 0.

### Model parameter optimization

To fit a model to a particular dataset, a particle swarm optimization procedure was used to find optimized values of the free parameters of the model to fit a given dataset. The particle swarm optimization routine was chosen because it rapidly finds

solutions in high-dimensional workspaces such as those utilized here and makes no assumptions regarding differentiability of the optimization problem. In this procedure, the system is initialized with a population of 10,000 random sets of parameter values (“particles”) and iterated until convergence to obtain an optimized parameter set. In each iteration, all parameter sets are evaluated by computing the RMSE of their fits to the data, and a fraction of all sets is replaced by random linear combinations of those parameter sets currently producing the best fits. The procedure stops when all 10,000 parameter sets converge within a 1% range of the optimal solution or after 100 consecutive iterations without any improvement in the optimal fit to the data. When the procedure stops, the optimal parameter set among the 10,000 sets from the last iteration is selected as the solution. For each model fit, the optimization procedure was run 10 times in order to evaluate any potential residual variability due to initial conditions or local optima. The resulting parameter estimates were highly robust to the initial conditions of the swarm procedure, indicative of reaching the global minimum of the RMSE measure. The minimum-RMSE solution across all 10 repetitions was chosen as the optimized parameter set, and Pearson’s *r* was calculated for this solution to characterize fit quality.

The particle swarm optimization procedure requires upper and lower bounds for the optimized parameters in order to efficiently search the parameter space. The parameter ranges for the current simulations were chosen to be big enough that they

TABLE 1 List of models included in the simulations.

Name	EQ	# Parameters	Optimized parameters
P	EQ4	2	$\alpha_P, \tau_A$
PI	EQ4	3	$\alpha_P, \alpha_I, \tau_A$
PD	EQ4	3	$\alpha_P, \alpha_D, \tau_A$
PID	EQ4	4	$\alpha_P, \alpha_I, \alpha_D, \tau_A$
D1	EQ5	3	$\alpha_A, \tau_A, \alpha_S$
D2	EQ5	4	$\alpha_A, \tau_A, \alpha_S, \tau_S$
D3	EQ6	3	$\alpha_A, \tau_A, \alpha_{Av}$
D4	EQ6	4	$\alpha_A, \tau_A, \alpha_{Av}, \tau_{Av}$
D5	EQ6	4	$\alpha_A, \tau_A, \alpha_S, \alpha_{Av}$
D6	EQ6	5	$\alpha_A, \tau_A, \alpha_S, \tau_S, \alpha_{Av}$
D7	EQ6	6	$\alpha_A, \tau_A, \alpha_S, \tau_S, \alpha_{Av}, \tau_{Av}$
D8	EQ6	6	$\alpha_A, \tau_A, \alpha_S, \alpha_{Av}, \tau_{Av}, \alpha_{Sv}$
D9	EQ6	7	$\alpha_A, \tau_A, \alpha_S, \tau_S, \alpha_{Av}, \tau_{Av}, \alpha_{Sv}$
D10	EQ6	8	$\alpha_A, \tau_A, \alpha_S, \tau_S, \alpha_{Av}, \tau_{Av}, \alpha_{Sv}, \tau_{Sv}$
D11	EQ7	4	$\alpha_A, \tau_A, \alpha_{As}, \tau_{As}$
D12	EQ7	6	$\alpha_A, \tau_A, \alpha_S, \tau_S, \alpha_{As}, \tau_{As}$
D13	EQ7	6	$\alpha_A, \tau_A, \alpha_S, \alpha_{As}, \tau_{As}, \alpha_{Ss}$
D14	EQ7	7	$\alpha_A, \tau_A, \alpha_S, \tau_S, \alpha_{As}, \tau_{As}, \alpha_{Ss}$
D15	EQ7	8	$\alpha_A, \tau_A, \alpha_S, \tau_S, \alpha_{As}, \tau_{As}, \alpha_{Ss}, \tau_{Ss}$

did not exclude any reasonable solutions<sup>3</sup> but small enough to allow for relatively rapid convergence to the optimal solution. With this goal, the allowable range for all gain parameters was  $-0.1$  to  $1.1$  with the exception of  $\alpha_I$  in the PI and PID models, which used a range of  $-0.001$  to  $0.001$  (the  $\alpha_I$  parameter corresponds to the gain of the auditory error integral, which determines how much the corrective response increases as the error accumulates over the duration of the perturbation; preliminary simulations resulted in very tiny values for this parameter that did not always stabilize when using the larger range). A negative gain indicates a response that exacerbates, rather than corrects, the corresponding error; the negative gains allow us to model following responses. A gain of 1 corresponds to immediate full compensation for the corresponding error; gains significantly above 1 are therefore prone to instabilities and highly unlikely to represent optimal solutions. Delay parameters were limited to 0–500 ms except for the differential delays  $\tau_{Av}$  and  $\tau_{Sv}$ , which were limited to  $-100$  to 500 ms to allow for the possibility that the velocity error response is faster than the position error response. Preliminary simulations indicated that none of the optimized parameters were at one of the ends of the allowable range for any model; in other words, solutions were not artificially limited by the chosen bounds.

## Akaike information criterion calculations

Because adding more parameters will inevitably improve RMSE (even to the point of overfitting the data), for model comparisons we focus on AIC, which is designed to meaningfully compare models with different numbers of free parameters using the information theoretic criterion of minimum information loss<sup>4</sup>. The AIC for each model is defined by the equation  $AIC = 2k - 2\ln(L)$ , where  $k$  is the number of free parameters in the model and  $L$  is the maximum likelihood of the model. We estimated the optimal model parameters for each model by minimizing the residual mean square error between the model fit and the observed traces. Assuming that the trace residuals were normally distributed but potentially correlated across timepoints, the model log-likelihood could be approximated as  $\ln(L) = N/2 (-\ln(MSE) - 1 - \ln(2\pi))$ , where MSE is the mean square error of the model, and

<sup>3</sup> For example, it does not make much sense within the DIVA model for the auditory feedback gain to be less than 0 (which would exacerbate rather than correct auditory errors) or greater than 1 (which would overcompensate for auditory errors). The bounds used here are slightly larger than these to allow for random variation that may occur in any particular dataset.

<sup>4</sup> We chose AIC here over the closely related Bayesian Information Criterion (BIC) because we anticipate that the model training datasets will generally be small; in such cases BIC tends to choose models that are too simple due to its use of a stronger penalty term for the number of model parameters (Burnham and Anderson, 2002).

$N$  is the effective degrees of freedom of the trace residuals (equal or smaller to the number of samples in the data). The degrees of freedom were computed using Satterthwaite–Welsh approximation (Satterthwaite, 1946) from the observed autocorrelation of the data before the onset of the perturbation (common to all models). Last, in order to facilitate comparisons of the resulting AIC measures across different datasets or with different studies, we reported corrected-AIC measures, dividing AIC by the data's effective degrees of freedom, leading to the combined equation:

$$cAIC = AIC/N = 2k/N + \ln(MSE) + 1 + \ln(2\pi) \quad (8)$$

When comparing two models, the relative likelihood of the two models can be computed from the difference in AIC values as  $\exp((AIC_{min} - AIC)/2)$ . To identify statistically significant differences in cAIC, we calculated the cAIC threshold necessary to support a 20:1 relative likelihood between the two models using the formula:

$$thr_{cAIC} = 2\ln(1/0.05)/N \quad (9)$$

A model whose cAIC is less than another model's cAIC by more than this threshold is, with 95% likelihood, the superior model.

## Cross-validated classification simulations

The last set of simulations further tested the models' abilities to characterize stable properties of each subject by optimizing the models using a subset of data from each subject (training trials) and then testing the models on the remaining trials (test trials). Specifically, for each model and each subject, 10 cross-validation iterations were performed, each involving a different random subset of 10 test trials (from a total of 13–20 trials per subject in Study 1 and 19–57 trials per subject in Study 2) used for testing, with the remaining trials for that subject used as the training set for optimizing model parameters (i.e., model parameters were optimized to fit the average trace of the training trials). The optimized model was then compared to the test trials to compute a combined cAIC value, using the same cAIC formula above as in the individual-trace analyses but setting MSE to the average of the MSE values across all of the individual test trials, and setting the data samples  $N$  to the average effective degrees of freedom across all the test trials multiplied by the total number of trials for that model/subject combination. This led to a single cAIC value for each model and each subject, characterizing the model's ability to predict the behavior of out-of-sample trials for an individual subject. The average of these cAIC values was then calculated across subjects for each model.

In addition, we wanted to evaluate, for each model, whether a subject's model parameter values could be used to uniquely



identify this subject's traces from different trials compared to the traces of other subjects. The models' abilities to correctly identify a subject were assessed from these same cross-validation iterations by first computing RMSE values comparing the average traces of one subject's test trials to the model traces obtained from fitting the training trials of the same (or a different) subject. From these comparisons we then determined *overall* and *pairwise* classification scores for each model, from a classification procedure that chose the subject with minimal RMSE as the most likely subject to have generated that mean test trace. All classification scores represent the percent correct identifications of a subject based on the mean of 10 test trials, averaged across the 10 cross-validation iterations and all appropriate between-subject comparisons. The *overall* classification scores represent the percentage of times the correct subject (i.e., the one who generated the test trials) had the lowest RMSE when compared to all other subjects for that same model, and they were computed as:

$$p_{overall} = \frac{1}{10N} \sum_{m=1}^N \sum_{i=1}^{10} \prod_{n \neq m}^N \frac{1}{10} \sum_{j=1}^{10} [RMSE_{imim} < RMSE_{imjn}]$$

where  $N$  is the number of subjects, and  $RMSE_{i,m,j,n}$  represents the RMSE value obtained when comparing the mean trace from the test trials of the  $i$ -th cross-validation iteration of subject  $m$  to the model traces obtained from fitting the training trials of the  $j$ -th cross-validation iteration of subject  $n$ . The *overall* classification scores for each model are reported in the "Overall" columns of **Table 2**. Study 1 involved 18 subjects and Study 2 involved 20, so chance performance on the classification task was 5.6% for Study 1 and 5% for Study 2.

*Pairwise* classification scores represent the percentage of times the correct subject had lower RMSE than another (randomly selected) incorrect subject for that same model, and they were computed as:

$$p_{pairwise} = \frac{1}{10N(N-1)} \sum_{m=1}^N \sum_{i=1}^{10} \sum_{n \neq m}^N \frac{1}{10} \sum_{j=1}^{10} [RMSE_{imim} < RMSE_{imjn}]$$

Classification accuracies for a given model were averaged across all pairs of subjects to obtain the scores listed in the "Pairwise" columns of **Table 2**; chance performance on this classification task is 50%.

Additionally, intraclass correlation coefficients (ICC) were calculated to quantify the reliability/stability of model parameters across the 10 cross-validation iterations. ICC values were calculated as:

$$ICC = \frac{\sigma_{between}^2}{\sigma_{between}^2 + \sigma_{within}^2}$$

where  $\sigma_{between}$  is the between-subject standard deviation for a given parameter and  $\sigma_{within}$  is the within-subject standard

deviation for a given parameter. ICC values fall between 0 and 1, with values  $< .5$  indicating poor reliability, values 0.5–0.75, 0.75–0.9, and 0.9–1 indicating moderate, good, and excellent reliability, respectively (Koo and Li, 2016).

## Results

**Table 2** summarizes the fit statistics for all models and simulations. The following subsections describe these results by simulation set: fits to study group means, fits to individual subject means, and cross-validated classification simulations.

### Fits to group means

The group mean trace for each study was formed by first calculating the mean  $f_o$  value at every time point for each individual subject (averaged across all that subject's trials), then averaging these individual subject means to form the group mean trace. The group mean traces are indicated by the solid blue lines in **Figure 2A** (Study 1) and **Figure 3A** (Study 2), with standard error of the mean (SEM) indicated by blue shading. Full compensation, the inverse of the perturbation magnitude, is shown in green. Full compensation illustrates what a 100% compensation for the perturbation would look like, although this is rarely achieved in auditory perturbation studies.

The columns labeled "Group" in **Table 2** indicate the RMSE for each model's fit to the group mean trace as well as the cAIC value resulting from comparing the model fit to the individual subject mean traces. The lowest RMSE and cAIC values for each study are indicated in boldface. Blue shading indicates cAIC values that are within the cAIC threshold of the lowest cAIC value; in other words, the models with no shading are inferior to the best (boldfaced) model according to the cAIC criterion, whereas the models with blue shading are not significantly different (at the  $p < 0.05$  false positive level) from the best model. For both studies, the three-parameter model D1 provided the best fit according to cAIC, with the three-parameter model PI also falling within the cAIC threshold, along with several four-parameter models (PID, D2, D5, and D11), a five-parameter model (D6), and several six-parameter models (D7, D8, D11, D12, and D13). For the remainder of this article, we will refer to models within the cAIC threshold of the best model collectively as the "best models."

When multiple models fall within the AIC threshold of the top model, there is not enough empirical evidence to support the selection of an individual model among them. In these cases, and until more evidence becomes available, it is reasonable to give preference to the model with the fewest parameters amongst these models. Thus, according to the cAIC criterion, the models providing the best fits to the group mean data are the three-parameter models D3 and PI, followed by the 4-parameter

TABLE 2 Fit statistics for all simulated models.

Model	Study 1 group		Study 2 group		Study 1 subject		Study 2 subject		Study 1 Xval classification			Study 2 Xval classification		
	RMSE	cAIC	RMSE	cAIC	RMSE	cAIC	RMSE	cAIC	cAIC	Overall	Pair	cAIC	Overall	Pair
P	0.00312	-6.33001	0.00198	-7.1133	0.00381	-6.1104	0.00237	-5.4655	-1.6882	23.39%	84.23%	-1.3595	18.25%	72.11%
PI	0.00083	-6.41743	0.00073	-7.1847	0.00201	-6.1983	0.00130	-5.4930	-1.7201	35.94%	88.10%	-1.3708	24.14%	74.59%
PD	0.00312	-6.32608	0.00197	-7.1110	0.00355	-6.1208	0.00234	-5.4644	-1.6870	26.82%	85.53%	-1.3587	18.21%	72.11%
PID	0.00065	-6.41588	0.00071	-7.1820	0.00158	-6.2102	0.00120	-5.4926	-1.7206	38.76%	88.77%	-1.3701	24.16%	74.58%
D1	0.00059	<b>-6.42106</b>	0.00039	<b>-7.1932</b>	0.00187	-6.2049	0.00112	-5.4950	-1.7228	35.70%	88.15%	-1.3713	22.69%	73.66%
D2	0.00059	-6.41672	0.00039	-7.1901	0.00174	-6.2023	0.00100	-5.4960	-1.7187	35.72%	88.28%	<b>-1.3715</b>	24.72%	74.68%
D3	0.00312	-6.32608	0.00197	-7.1110	0.00355	-6.1208	0.00234	-5.4644	-1.6870	26.82%	85.53%	-1.3587	18.21%	72.11%
D4	0.00116	-6.40619	0.00083	-7.1780	0.00183	-6.1964	0.00145	-5.4848	-1.7157	37.75%	88.42%	-1.3669	22.04%	73.25%
D5	0.00053	-6.41737	0.00026	-7.1921	0.00151	-6.2129	0.00105	-5.4941	<b>-1.7233</b>	38.55%	88.74%	-1.3706	23.07%	73.78%
D6	0.00052	-6.41321	0.00021	-7.1895	0.00137	-6.2106	0.00092	-5.4951	-1.7195	38.73%	88.89%	-1.3708	25.56%	74.83%
D7	0.00046	-6.40946	0.00017	-7.1867	0.00113	-6.2103	0.00081	-5.4942	-1.7147	39.49%	89.00%	-1.3697	25.65%	74.93%
D8	0.00052	-6.40886	0.00017	-7.1866	0.00117	-6.2093	0.00092	-5.4918	-1.7142	38.53%	88.78%	-1.3688	23.70%	73.94%
D9	0.00052	-6.40451	0.00017	-7.1835	0.00103	-6.2065	0.00076	-5.4930	-1.7110	40.20%	89.18%	-1.3689	26.05%	75.06%
D10	0.00048	-6.40055	0.00013	-7.1806	0.00100	-6.2014	0.00071	-5.4922	-1.7070	40.37%	89.14%	-1.3679	26.12%	75.13%
D11	0.00104	-6.40902	0.00040	-7.1900	0.00147	<b>-6.2129</b>	0.00097	<b>-5.4964</b>	-1.7227	38.57%	88.93%	-1.3713	25.38%	75.08%
D12	0.00051	-6.40890	0.00027	-7.1857	0.00111	-6.2110	0.00078	-5.4948	-1.7129	39.06%	88.94%	-1.3699	25.70%	75.24%
D13	0.00051	-6.40889	0.00027	-7.1857	0.00116	-6.2099	0.00082	-5.4944	-1.7160	39.92%	89.03%	-1.3698	25.07%	75.15%
D14	0.00051	-6.40454	0.00013	-7.1838	0.00105	-6.2064	0.00076	-5.4933	-1.7110	<b>40.71%</b>	89.18%	-1.3690	25.01%	75.16%
D15	0.00035	-6.40165	0.00012	-7.1807	0.00093	-6.2027	0.00066	-5.4924	-1.7058	40.57%	<b>89.22%</b>	-1.3682	<b>26.32%</b>	<b>75.27%</b>

Xval, cross-validation; RMSE, root-mean-square error; cAIC, corrected Akaike information criterion; Overall, overall accuracy (%); Pair, pairwise accuracy (%).

Boldface type indicates the model with lowest RMSE and cAIC for each study.

Blue shading indicates cAIC values that are within the cAIC threshold of the lowest cAIC value.

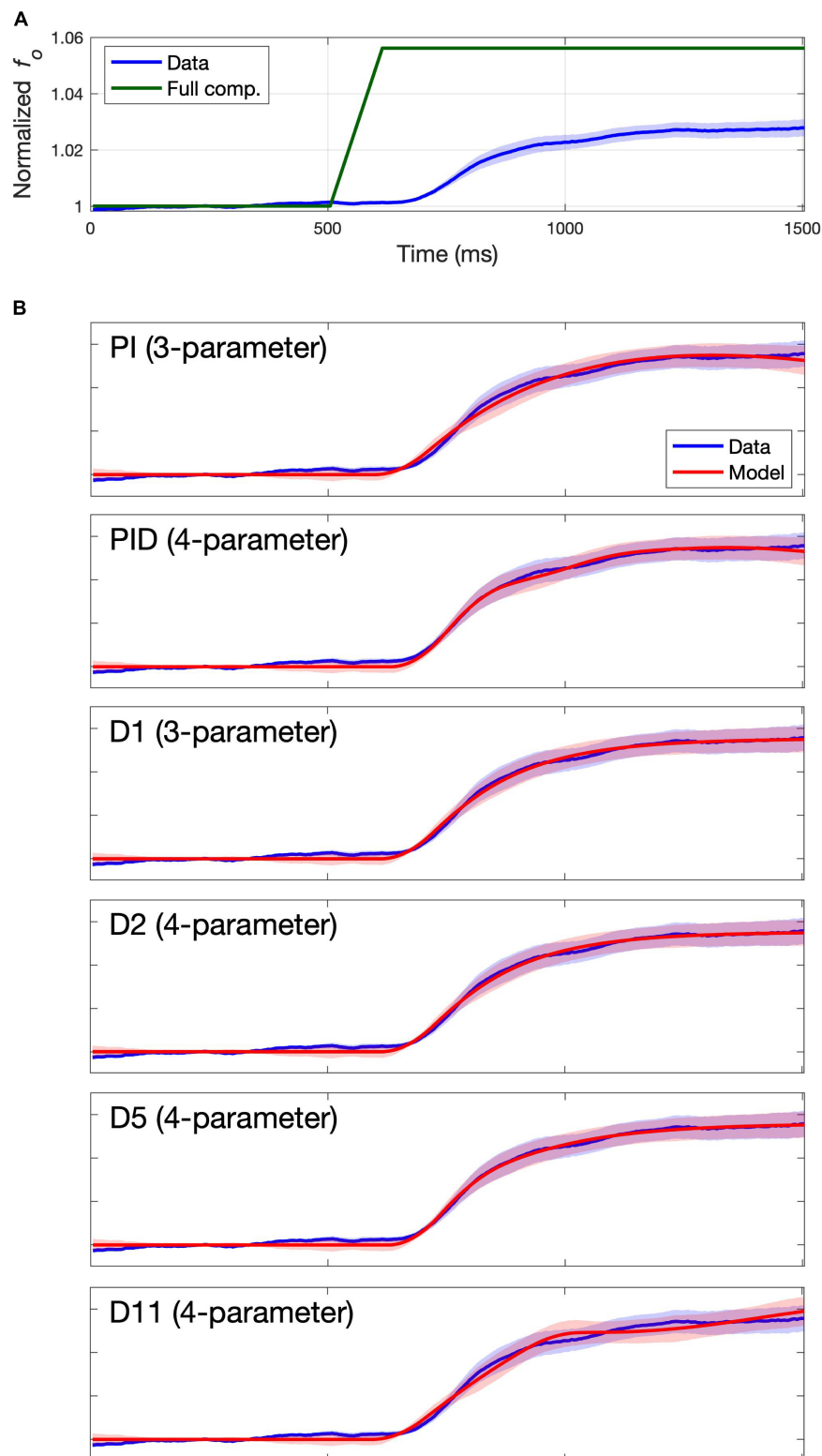


FIGURE 2

Group mean data and model fits for Study 1. Group mean data and standard error of the mean are shown with a blue line and shading.

(A) Group mean data shown relative to full compensation in green. Full compensation is the inverse of the perturbation magnitude and illustrates what 100% compensation would look like. (B) Group mean data shown relative to model fit (red line) and standard error of the model fit (red shading) for models providing best fits to the group mean data.

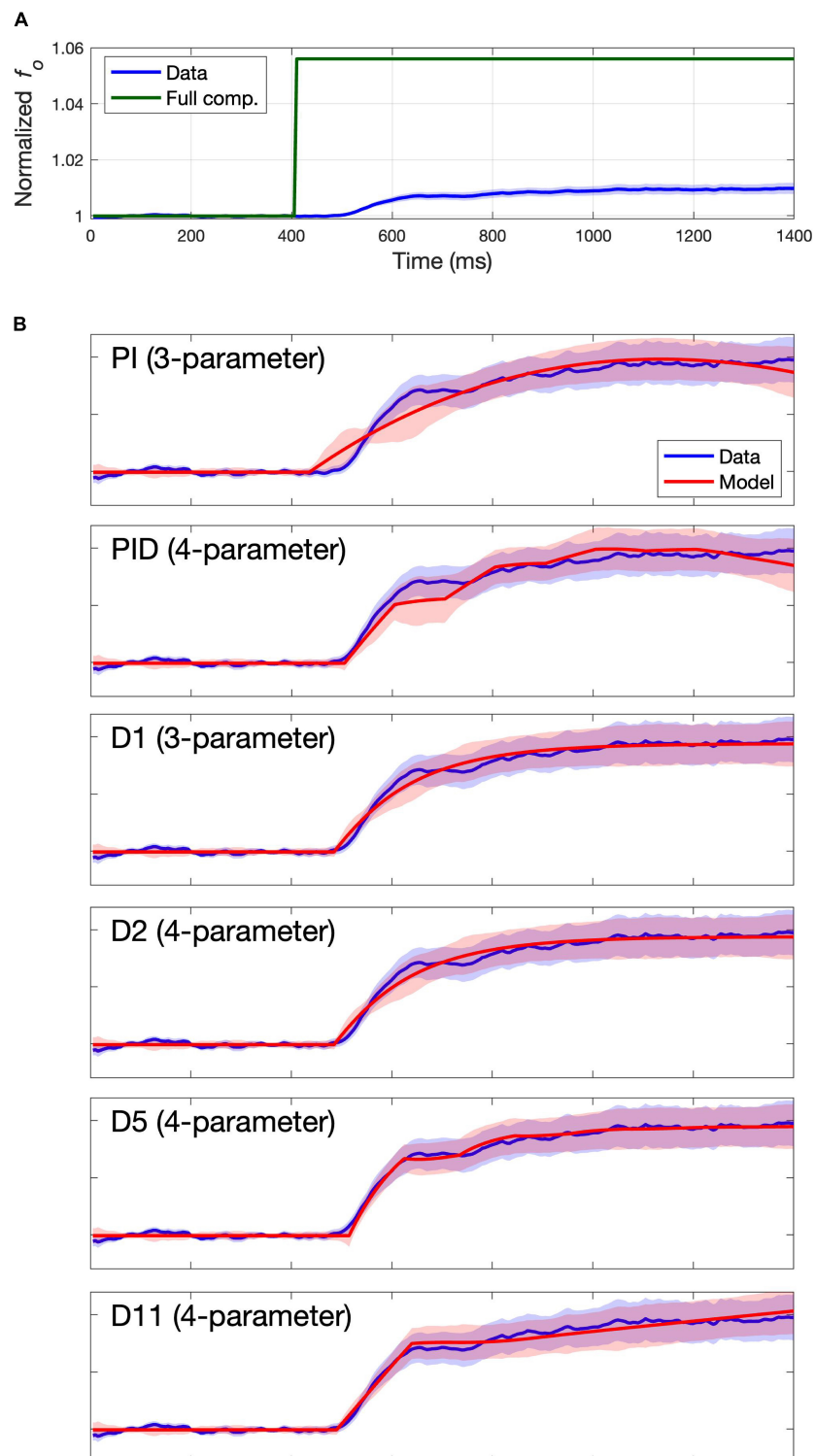


FIGURE 3

Group mean data and model fits for Study 2. Group mean data and standard error of the mean are shown with a blue line and shading. **(A)** Group mean data shown relative to full compensation in green. Full compensation is the inverse of the perturbation magnitude and illustrates what 100% compensation would look like. **(B)** Group mean data shown relative to model fit (red line) and standard error of the model fit (red shading) for models providing best fits to the group mean data.

models PID, D2, D5, and D11. **Figures 2B, 3B** plot the model fit (red line) and standard error of the model fit (red shading), along with the group mean (blue line) and SEM (blue shading) of the experimental data for these models. (Model fits for all models are provided in the **Supplementary materials**). Despite falling within the cAIC threshold of the best model D1, the PI and PID models produce traces that poorly match the overall shape of the data trace, casting doubt as to whether they are effectively capturing the physiological mechanisms responsible for the subjects' productions. This is particularly clear from the Study 2 fits in **Figure 3**. The four-parameter model D5 appears to best capture the overall shape of the data trace, which shows an initial plateau followed by a second rise approximately 100 ms after the start of the plateau (again more clearly visible in **Figure 3**). As noted in the *Introduction*, a two-component response pattern has been noted in prior pitch perturbation experiments (e.g., [Larson, 1998](#); [Hain et al., 2000](#)). Models D1, D2, and D11 capture the overall shape of the data trace reasonably well, but they fail to properly capture the shape of the plateau and second rise.

In sum, the three-parameter model D1 provides the best fit of the group mean traces according to the cAIC criterion while also capturing the overall shape of the experimentally measured response reasonably well. The four-parameter model D5 best captures the overall shape of the data traces amongst the three- and four-parameter models and falls within the cAIC threshold of model D1. The additional parameters of models with more than four parameters appear to provide little additional improvement.

The optimized values of all parameters for all models are provided in the **Supplementary materials**. For the basic DIVA best models (D1 and D2), the parameter values were very similar between models for a given dataset. For Study 1, the mean values (across the two models) were 0.011 for  $\alpha_A$ , 0.013 for  $\alpha_S$ , 115 ms for  $\tau_A$ , and 130 ms for  $\tau_S$ . For Study 2, they were 0.006 for  $\alpha_A$ , 0.033 for  $\alpha_S$ , 93 ms for  $\tau_A$ , and 54 ms for  $\tau_S$ . These parameters had similar values in the generalized DIVA best models (D5–D8 and D11–D13), whereas the additional parameters in the generalized DIVA models were considerably more variable across models.

## Fits to individual subjects

The second set of simulations compared the models on their ability to fit individual subject data using parameters optimized for the individual subject rather than the group mean. These simulations gauge how well the models can account for individual differences through subject-specific parameterizations. For each subject, model parameters were optimized to fit the subject's mean trace (averaged across trials). The RMSE values of these fits are provided in the columns labeled "Subject" in **Table 2**, along with the cAIC values resulting

from comparing the models' fits to the individual subject mean traces. With the exception of models P, PD, D3, and D4, all models fell within the cAIC threshold of the best model (D11 for both studies).

## Cross-validated classification simulations

The columns labeled "Xval Classification" in **Table 2** provide cAIC, overall classification accuracy, and pairwise classification accuracy for each model in each study. The models within the cAIC threshold of the best cAIC value for both studies were models PI, PID, D1, D2, D5–D7, and D11–D13. The highest overall classification accuracies were 40.71% for model D14 in Study 1 (chance level of 5.6%) and 26.32% for model D15 in Study 2 (chance level 5%). Even the worst-performing models had overall accuracies that were well above chance: 23.39% for model P in Study 1 and 18.21% for models PD and D3 in Study 2. Pairwise classification accuracies were also well above chance (50%) for all models, ranging from 84.23% (model D1) to 89.22% (model D15) for Study 1, and from 72.11% (models P and D1) to 75.27% (model D15) in study 2.

Overall, these results indicate that reflexive responses to  $f_0$  perturbations are largely individual-specific, and a number of models perform nearly equivalently on the cross-validated classification tasks. For comparison, we also calculated cross-validated classification accuracy when we used the mean of the training trials for classification rather than one of the models. This resulted in overall and pairwise accuracies of 38.14 and 89.25%, respectively, for Study 1 and 25.89 and 75.38% for Study 2. These are similar to values obtained for the best-performing models in **Table 2**.

The cross-validation training iterations also provide information regarding the stability of model parameters across the 10 iterations for a given subject. In other words, do the 10 iterations yield approximately the same values for a given parameter (as would be expected if the parameter has a reliable physiological basis) or do they vary substantially across iterations (indicative of a model whose parameters do not have a reliable physiological interpretation)? To assess this, we calculated ICC for each parameter in each model for each data set. The mean parameter values and ICC values from the 10 cross-validation iterations are provided in **Table 3** (Study 1) and **Table 4** (Study 2). Boldface type indicates the model with the highest ICC value per parameter. Dark blue shading indicates ICC values greater than 0.75 (corresponding to good reliability), and light blue shading indicates ICC values between 0.5 and 0.75 (moderate reliability).

Generally speaking, parameter stability was higher for the PID-based models and models D1–D10 compared to models D11–D15. In particular, all PID models and models D1–D9 had highly reliable values for the auditory feedback control

TABLE 3 Study 1 mean values and ICC of optimized parameters in cross-validation simulations.

Model	$\alpha_P/\alpha_A$		$\alpha_S$		$\alpha_D/\alpha_{Av}/\alpha_{As}$		$\alpha_I/\alpha_{Sv}/\alpha_{Ss}$		$\tau_A$		$\tau_S$		$\tau_{Av}/\tau_{As}$		$\tau_{Sv}/\tau_{Ss}$	
	Mean	ICC	Mean	ICC	Mean	ICC	Mean	ICC	Mean	ICC	Mean	ICC	Mean	ICC	Mean	ICC
P	0.005	<b>0.956</b>							0.044	0.647						
PI	0.009	0.939					-4.9E-05	<b>0.826</b>	0.109	0.660						
PD	0.010	0.924			0.923	<b>0.592</b>			0.109	0.601						
PID	0.013	0.918			0.551	0.414	-1.0E-04	<b>0.842</b>	0.135	0.594						
D1	0.012	0.851	0.016	0.705					0.128	0.586						
D2	0.010	0.870	0.013	0.657					0.111	0.556	0.159	<b>0.470</b>				
D3	0.010	0.924			0.923	<b>0.592</b>			0.109	0.601						
D4	0.008	0.941			0.788	0.688			0.101	<b>0.690</b>			0.274	<b>0.582</b>		
D5	0.015	0.852	0.018	<b>0.772</b>	0.393	0.328			0.143	0.536						
D6	0.012	0.903	0.014	0.747	0.568	0.335			0.128	0.568	0.226	0.337				
D7	0.011	0.851	0.009	0.650	0.731	<b>0.662</b>			0.118	0.534	0.203	0.369	0.162	0.455		
D8	0.014	0.827	0.011	0.699	0.789	0.468	0.170	0.317	0.123	0.502			0.218	0.469		
D9	0.012	0.823	0.008	0.576	0.648	<b>0.565</b>	0.403	0.412	0.124	0.500	0.152	0.419	0.240	0.350		
D10	0.012	0.836	0.008	0.619	0.734	0.644	0.444	0.296	0.123	0.488	0.143	0.318	0.201	0.257	0.122	<b>0.268</b>
D11	0.013	0.662			-0.010	0.584			0.120	0.613			0.296	<b>0.510</b>		
D12	0.018	0.545	0.069	0.657	0.013	<b>0.694</b>			0.110	0.397	0.080	0.235	0.307	0.358		
D13	0.033	0.592	0.123	0.762	0.020	0.686	0.573	0.206	0.120	0.354			0.300	0.396		
D14	0.022	0.448	0.070	0.611	0.014	0.575	0.618	0.472	0.119	0.391	0.078	0.226	0.319	0.309		
D15	0.015	0.666	0.035	0.555	0.011	0.607	0.697	0.440	0.114	0.380	0.052	0.277	0.287	0.301	0.185	0.264

Only one parameter listed per column is optimized in a given model. For example, the PID models have an  $\alpha_P$  parameter whereas the DIVA models have an  $\alpha_A$  parameter. See Table 1 for a complete list of parameters included in each model. ICC, intraclass correlation coefficient.

Boldface type indicates the model with the highest ICC value per parameter.

Light blue shading indicates ICC values are between 0.5 and 0.75 (moderate reliability).

Dark blue shading indicates ICC values are > 0.75 (good-excellent reliability).

TABLE 4 Study 2 mean values and ICC of optimized parameters in cross-validation simulations.

Model	$\alpha_P/\alpha_A$		$\alpha_S$		$\alpha_D/\alpha_{Av}/\alpha_{As}$		$\alpha_I/\alpha_{Sv}/\alpha_{Ss}$		$\tau_A$		$\tau_S$		$\tau_{Av}/\tau_{As}$		$\tau_{Sv}/\tau_{Ss}$	
	Mean	ICC	Mean	ICC	Mean	ICC	Mean	ICC	Mean	ICC	Mean	ICC	Mean	ICC	Mean	ICC
P	0.001	0.969							0.031	0.761						
PI	0.003	<b>0.972</b>					-2.6E-05	<b>0.955</b>	0.062	0.810						
PD	0.003	0.969			1.019	0.597			0.055	0.793						
PID	0.005	0.954			0.811	0.678	-3.4E-05	<b>0.947</b>	0.091	<b>0.823</b>						
D1	0.009	0.881	0.098	0.725					0.110	0.746						
D2	0.005	0.896	0.035	0.543					0.088	0.746	0.109	0.744				
D3	0.003	0.969			1.019	0.597			0.055	0.793						
D4	0.002	0.969			0.972	0.689			0.057	0.800			0.359	<b>0.747</b>		
D5	0.010	0.912	0.078	0.780	0.439	0.633			0.122	0.726						
D6	0.006	0.935	0.037	0.623	0.622	0.612			0.101	0.755	0.144	<b>0.758</b>				
D7	0.006	0.927	0.034	0.792	0.663	<b>0.762</b>			0.098	0.746	0.133	0.552	0.175	0.601		
D8	0.008	0.910	0.038	0.651	0.719	0.641	0.005	0.380	0.110	0.734			0.193	0.537		
D9	0.006	0.942	0.027	0.785	0.524	0.704	0.403	0.480	0.099	0.751	0.136	0.456	0.259	0.374		
D10	0.006	0.946	0.027	<b>0.797</b>	0.539	0.646	0.520	0.483	0.098	0.718	0.114	0.441	0.265	0.293	0.189	0.368
D11	0.004	0.702			-0.004	0.702			0.089	0.583			0.228	0.647		
D12	0.005	0.666	0.053	0.700	0.002	0.430			0.097	0.741	0.072	0.353	0.322	0.562		
D13	0.008	0.642	0.083	0.607	0.004	0.536	0.226	0.423	0.103	0.686			0.306	0.502		
D14	0.006	0.646	0.052	0.536	0.002	0.504	0.510	0.303	0.097	0.663	0.075	0.317	0.333	0.489		
D15	0.006	0.726	0.046	0.547	0.002	0.584	0.607	0.660	0.100	0.767	0.054	0.405	0.324	0.543	0.221	<b>0.480</b>

Only one parameter listed per column is optimized in a given model. For example, the PID models have an  $\alpha_P$  parameter whereas the DIVA models have an  $\alpha_A$  parameter. See Table 1 for a complete list of parameters included in each model. ICC, intraclass correlation coefficient.

Boldface type indicates the model with the highest ICC value per parameter.

Light blue shading indicates ICC values are between 0.5 and 0.75 (moderate reliability).

Dark blue shading indicates ICC values are > 0.75 (good-excellent reliability).

gain parameter ( $\alpha_P$  in PID models and  $\alpha_A$  in DIVA-based models) and moderately to highly reliable values for the auditory feedback control delay parameter in both studies. The somatosensory feedback control gain parameter was also moderately to highly reliable in all DIVA-based models (D1–D15) in both studies, and the parameter  $\alpha_I$  was highly reliable in the PI and PID models in both studies.

## Discussion

The primary goal of this study is the identification of a model that captures population responses in auditory perturbation experiments, and perhaps more importantly characterizes individual differences in a stable manner with parameters that relate to underlying motor control capabilities. The latter capability is particularly important if the model is to be used to characterize individuals with communication disorders for the purpose of providing individualized treatments that capitalize on the individual's strengths and weaknesses. For this approach to bear fruit, it is important that the behavioral responses exhibited by experimental subjects are reasonably stable and differ between individuals; if not, then no model will be capable of achieving our goal. A key finding from the current study (independent of any modeling) is that reflexive responses to  $f_o$  perturbations are largely individual-specific, providing optimism that such responses may reveal key insights into the individual's speech motor control processes. Although all subjects were healthy adults with no communication disorders (and therefore likely to have somewhat similar speech motor systems, in contrast to individuals with a speech disorder), the cross-validation classification analyses indicate that the mean of 10 reflexive responses from an individual is enough to distinguish that individual from another neurotypical individual with approximately 90% accuracy in Study 1 (see *pair* column in *Study 1 Xval Classification* section of **Table 2**) and 75% in Study 21 (see *pair* column in *Study 2 Xval Classification* section of **Table 2**). This highlights a rather remarkable property of the PSR independent of any modeling: an individual's pitch shift response is akin to a "fingerprint" that largely distinguishes them from other individuals (though not to the degree of an actual fingerprint). We expect that individuals with speech motor disorders will show much greater variability than our current healthy sample and therefore may be easier to distinguish based on their reflexive responses; verification of this expectation is an important topic for future research.

The three-parameter model D1 provided fits to group mean data with the lowest cAIC values of any model for both Study 1 and Study 2. Furthermore, this model was within the cAIC threshold of the lowest cAIC for individual subject fits and cross-validation simulations for both studies. Other models

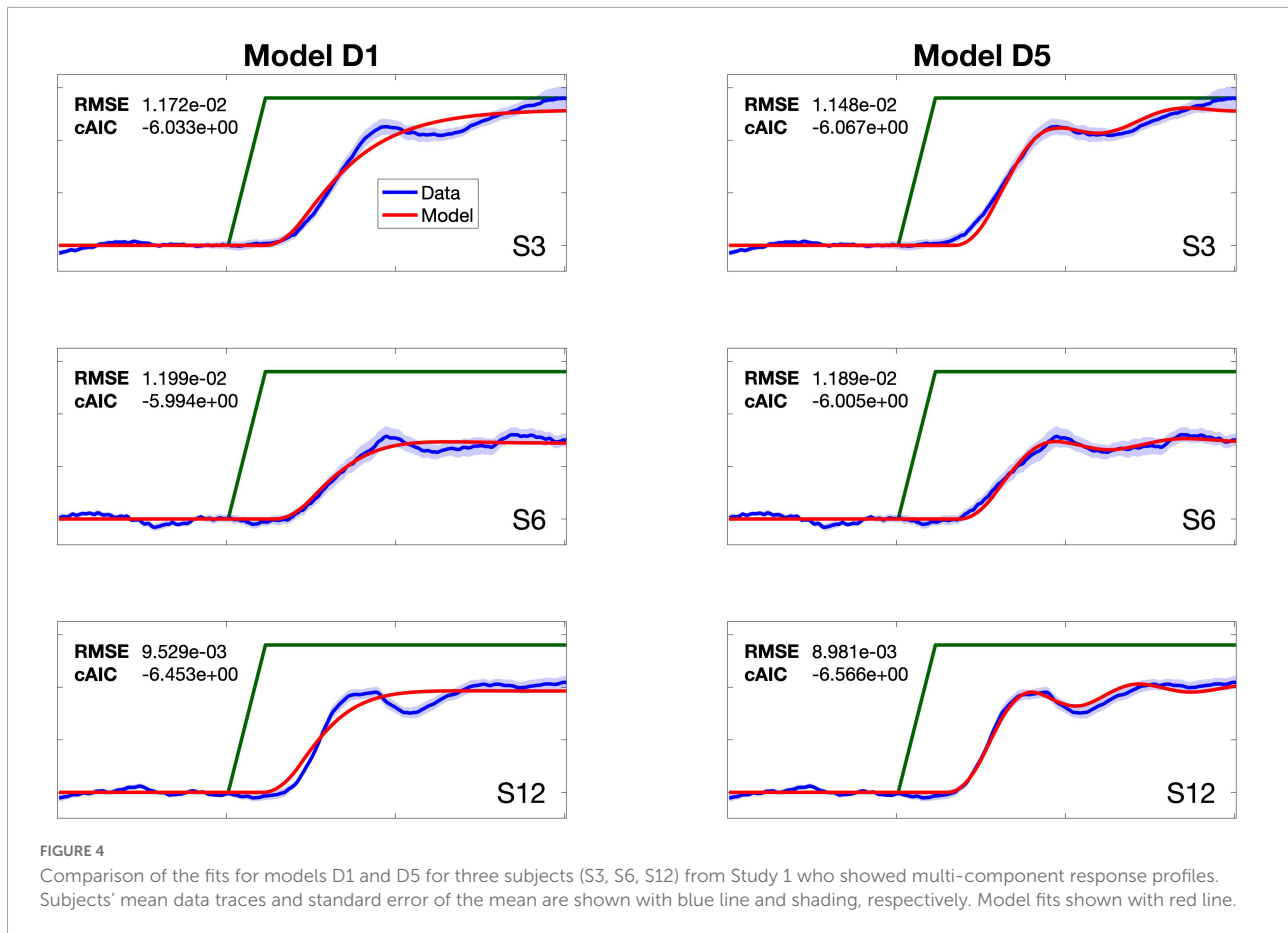
that fell into the best model category (i.e., those within the cAIC threshold of the lowest cAIC value) for all simulations were PI, D2, D5, D6, D7, D11, D12, and D13. Model D1 also had amongst the most stable parameters across cross-validation iterations as measured by ICC (see **Table 3**), and its pairwise classification scores were within 1–2% of the best-performing model.

Concerning the three-parameter model PI, although this model performed well according to the cAIC, parameter stability, and cross-validated classification criteria, the overall shape of the responses of the PI model differed considerably from the shape of the subject responses (compare the fits of models PI and D1 in **Figures 2B, 3B**). The anomalous response shape for the PI model is the result of the fact that the optimized values for the parameter  $\alpha_I$ , which determines how much the corrective response increases as error accumulates, were negative (see **Tables 3, 4**), indicating that the correction actually *decreased* with accumulating error. This is contrary to the theoretical motivation for this term (which is to increase the correction if the error keeps accumulating) and results in the "inverted U" shape of the PI model responses in **Figures 2, 3** that is not found in the data traces nor in model D1. It is also worth noting that fixing  $\alpha_I$  at 0 so it will not go negative reduces the PI and PID models to the poorly performing P and PD models, respectively.

For these reasons, we conclude that the best 3-parameter model for characterizing reflexive responses to  $f_o$  perturbations is D1 (EQ5), which has free parameters  $\alpha_A$ ,  $\tau_A$ , and  $\alpha_S$ . These parameters have straightforward interpretations:  $\alpha_A$  (which corresponds to the parameter B in a state-space formulation—see *Basic DIVA/SS equation* in the "Materials and Methods" section) is the gain of the auditory feedback controller's response to a perceived error,  $\tau_A$  is the delay of this response, and  $\alpha_S$  is the gain of the "resistance" to this correction. Within the DIVA model, this latter parameter corresponds to the gain of the somatosensory feedback controller, which is attempting to keep  $f_o$  (as detected through somatic sensation, which is not perturbed in the current experiment) at the target level.  $\alpha_S$  is related to the parameter A in a state-space formulation (specifically,  $A = -\alpha_S$ ); this parameter similarly acts to resist changes due to perceived auditory error, though it is not typically specifically associated with somatosensory feedback control. Model D1 is also equivalent to a low-pass filter/leaky integrator model, as proposed by [Larson et al. \(2000\)](#).

A more general interpretation of  $\alpha_S$ , which is consistent with both the DIVA and state-space formulations is that it reflects the influence of non-auditory-based motor subsystems on the overall motor output. This can include both feedforward control mechanisms and somatosensory feedback control mechanisms. Indeed, the estimate of the somatosensory state in DIVA is envisioned as a combination of an efference copy of the motor command (which provides a predictive estimate of





somatosensory state) and incoming somatosensory information (see for example [Figure 1](#) in [Guenther et al., 1998](#)). The use of a predictive estimate of the sensory state within a sensory feedback control architecture (see also [Houde and Nagarajan, 2011](#)) is, in essence, a form of feedforward control since it does not depend on sensory feedback for generating control signals.

Amongst the four-parameter models (PID, D2, D4, D5, and D11), models D5 and D11 were within the cAIC threshold of the lowest cAIC for all simulations for both studies (shaded cells in [Table 2](#)), and both of these models exhibited relatively high parameter stability ([Table 3](#)). Of these two models, D5 produced fits that better captured the overall shapes of the response profiles ([Figures 2, 3](#)). Although the cAIC values for models D5 and D11 were in no cases significantly better than the 3-parameter model D1, it is noteworthy that models D5 and D11 (as well as most of the models with five or more parameters) are better capable of accounting for multi-component response profiles. This is illustrated in [Figure 4](#), which compares the fits of models D1 and D5 to individual subjects from Study 1 who exhibited multi-component responses. Multi-component responses have also been reported in several prior PSR studies ([Burnett et al., 1997, 1998](#); [Larson, 1998](#); [Larson](#)

[et al., 2000](#); [Hain et al., 2000](#)), and it appears that the second response component is under more conscious control than the earlier “automatic” component; for example, the second component is much more influenced by instructions provided to subjects regarding whether they should attempt to oppose or follow the perturbation direction ([Hain et al., 2000](#)). The 4th parameter in model D5 is an auditory velocity error gain,  $\alpha_{AV}$ . This term has the effect of resisting any perceived changes in pitch (beyond the abrupt change at perturbation onset, which is ignored by the model), in keeping with the fact that subjects are attempting to maintain a constant pitch, as they were instructed to do in the studies modeled here.

Despite D5 better capturing multicomponent responses, D5 was not superior to D1 according to the cAIC criterion in any of the simulations; in other words, the reduction in RMSE afforded by the 4th parameter in D5 was offset by the AIC penalty term for increasing the number of model parameters by 1. This suggests that the secondary responses, which are better characterized by D5, are quite variable compared to the primary response, which is captured well by both D1 and D5. While the later components could be more influenced by cognitive variables such as attention level and conscious intent

(Burnett et al., 1997, 1998; Hain et al., 2000), modeling the contribution of those processes was beyond the scope of the current study. Additional parameters beyond 4 provide little additional improvement.

It is reasonable to wonder what is gained from characterizing and individual's reflexive responses to  $f_o$  perturbations with a parameterized model, given that the average of a set of training traces provides classification results that are on par with the best model characterizations. The key difference is that *a model whose parameters correspond to physiological motor control processes provides a quantitative assessment of an individual's motor speech capabilities*. For example, a past pitch perturbation study involving individuals with Parkinson's disease indicated greater compensation than age-matched controls (Liu et al., 2012). By itself, this observation is of limited value for characterizing the motor control processes of an individual with Parkinson's disease since a larger response might indicate enhanced auditory feedback control or, alternatively, degraded somatosensory feedback control. In contrast, the optimal fit of model D1 to the subject's response traces provides values of  $\alpha_A$  and  $\alpha_S$  that best capture the subject's response. These values can be compared to normative values to separately assess the integrity of the auditory and somatosensory feedback control subsystems. If, for example, an individual with Parkinson's disease has an abnormally low  $\alpha_S$  with normal  $\alpha_A$ , a clinician may favor approaches that leverage intact auditory feedback control capabilities to overcome deficient somatosensory feedback control capabilities. In contrast, the parameters in the PID models are interpreted relative to error correction (and whether that correction is proportional to the error, or an integral or derivative of the error). This interpretation does not convey information about the mechanisms driving the correction and may limit how that information could be used in a therapeutic context. Although much work remains to be done to verify the veracity of the D1 model's characterization, such an approach holds the promise of informing personalized therapeutic interventions, much like other reflexes such as the pupillary light reflex have proven useful for characterizing the integrity of the nervous system in cases of neurological impairment.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Author contributions

EK and FG conceptualized and designed the study. FG and AD developed the computational models. EK, AN-C, and RF developed the software. EH and DS collected and processed the data. EK, AN-C, and FG wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the final submitted version.

## Funding

This research was supported by NIH grants: R01 DC002852 (FG, PI), R01 DC016270 (FG and C. Stepp, PIs), F31 DC016197 (EH, PI), and R21 DC017563 (AD, PI).

## Acknowledgments

We are grateful to Cara Stepp for kindly sharing data with us for this study.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2022.929687/full#supplementary-material>

## References

- Abur, D., Subaciute, A., Daliri, A., Lester-Smith, R. A., Lupiani, A. A., Cilento, D., et al. (2021a). Feedback and feedforward auditory-motor processes for voice and articulation in Parkinson's disease. *J. Speech Lang. Hear. Res.* 64, 4682–4694. doi: 10.1044/2021\_JSLHR-21-00153
- Abur, D., Subaciute, A., Kapsner-Smith, M., Segina, R. K., Tracy, L. F., Noordzij, J. P., et al. (2021b). Impaired auditory discrimination and auditory-motor integration in hyperfunctional voice disorders. *Sci. Rep.* 11:13123. doi: 10.1038/s41598-021-92250-8
- Ballard, K. J., Halaki, M., Sowman, P. F., Kha, A., Daliri, A., Robin, D., et al. (2018). An investigation of compensation and adaptation to auditory perturbations in individuals with acquired apraxia of speech. *Front. Hum. Neurosci.* 12:510. doi: 10.3389/fnhum.2018.00510
- Bär, K.-J., Boettger, M. K., Schulz, S., Harzendorf, C., Agelink, M. W., Yeragani, V. K., et al. (2008). The interaction between pupil function and cardiovascular regulation in patients with acute schizophrenia. *Clin. Neurophysiol.* 119, 2209–2213. doi: 10.1016/j.clinph.2008.06.012
- Behroozmand, R., Bonilha, L., Rorden, C., Hickok, G., and Fridriksson, J. (2022). Neural correlates of impaired vocal feedback control in post-stroke aphasia. *Neuroimage* 250:118938. doi: 10.1016/j.neuroimage.2022.118938
- Behroozmand, R., Ibrahim, N., Korzyukov, O., Robin, D. A., and Larson, C. R. (2014). Left-hemisphere activation is associated with enhanced vocal pitch error detection in musicians with absolute pitch. *Brain Cogn.* 84, 97–108. doi: 10.1016/j.bandc.2013.11.007
- Behroozmand, R., Phillip, L., Johari, K., Bonilha, L., Rorden, C., Hickok, G., et al. (2018). Sensorimotor impairment of speech auditory feedback processing in aphasia. *Neuroimage* 165, 102–111. doi: 10.1016/j.neuroimage.2017.10.014
- Boersma, P., and Weenink, D. (2018). *Praat: Doing phonetics by computer [Computer software]*. Available online at: <http://www.praat.org/>
- Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *J. Acoust. Soc. Am.* 103, 3153–3161. doi: 10.1121/1.423073
- Burnett, T. A., Senner, J. E., and Larson, C. R. (1997). Voice F0 responses to pitch-shifted auditory feedback: A preliminary study. *J. Voice* 11, 202–211. doi: 10.1016/S0892-1997(97)80079-3
- Burnham, K., and Anderson, D. (2002). *Model selection and multimodel Inference: A practical information-theoretic approach*, 2nd Edn. New York, NY: Springer.
- Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., and Perkell, J. S. (2008). “A system for online dynamic perturbation of formant frequencies and results from perturbation of the Mandarin triphthong /iaul/” in *Proceedings of the 8th intl. seminar on speech production* (Strasbourg), 65–68.
- Chen, S. H., Liu, H., Xu, Y., and Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *J. Acoust. Soc. Am.* 121, 1157–1163.
- Daliri, A. (2021). A computational model for estimating the speech motor system's sensitivity to auditory prediction errors. *J. Speech Lang. Hear. Res.* 64, 1841–1854. doi: 10.1044/2021\_JSLHR-20-00484
- Daliri, A., and Dittman, J. (2019). Successful auditory motor adaptation requires task-relevant auditory errors. *J. Neurophysiol.* 122, 552–562. doi: 10.1152/jn.00662.2018
- Daliri, A., Chao, S.-C., and Fitzgerald, L. C. (2020). Compensatory responses to formant perturbations proportionally decrease as perturbations increase. *J. Speech Lang. Hear. Res.* 63, 3392–3407. doi: 10.1044/2020\_JSLHR-19-00422
- Demopoulos, C., Kothare, H., Mizuir, D., Henderson-Sabes, J., Fregeau, B., Tjernagel, J., et al. (2018). Abnormal speech motor control in individuals with 16p11.2 deletions. *Sci. Rep.* 8:1274. doi: 10.1038/s41598-018-19751-x
- Elman, J. L. (1981). Effects of frequency-shifted feedback on the pitch of vocal productions. *J. Acoust. Soc. Am.* 70, 45–50. doi: 10.1121/1.386580
- Fan, X., Miles, J. H., Takahashi, N., and Yao, G. (2009). Abnormal transient pupillary light reflex in individuals with autism spectrum disorders. *J. Autism Dev. Disord.* 39, 1499–1508. doi: 10.1007/s10803-009-0767-7
- Franken, M. K., Acheson, D. J., McQueen, J. M., Hagoort, P., and Eisner, F. (2018). Opposing and following responses in sensorimotor speech control: Why responses go both ways. *Psychon. Bull. Rev.* 25, 1458–1467. doi: 10.3758/s13423-018-1494-x
- Galea, J. M., Mallia, E., Rothwell, J., and Diedrichsen, J. (2015). The dissociable effects of punishment and reward on motor learning. *Nat. Neurosci.* 18, 597–602. doi: 10.1038/nn.3956
- Guenther, F. H. (2016). *Neural control of speech*. Cambridge, MA: MIT Press.
- Guenther, F. H., Hampson, M., and Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychol. Rev.* 105, 611–633. doi: 10.1037/0033-295X.105.4.611-633
- Hain, T., Burnett, T., Kiran, S., Larson, C., Singh, S., and Kenney, M. (2000). Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Exp. Brain Res.* 130, 133–141. doi: 10.1007/s002219900237
- Hall, C. A., and Chilcott, R. P. (2018). Eyeing up the future of the pupillary light reflex in neurodiagnostics. *Diagnostics* 8:19. doi: 10.3390/diagnostics8010019
- Hantzsch, L., Parrell, B., and Niziolek, C. A. (2022). A single exposure to altered auditory feedback causes observable sensorimotor adaptation in speech. *Elife* 11:e73694.
- Heller Murray, E. S., and Stepp, C. E. (2020). Relationships between vocal pitch perception and production: A developmental perspective. *Sci. Rep.* 10, 1–10. doi: 10.1038/s41598-020-60756-2
- Heller Murray, E. S., Lupiani, A. A., Kolin, K. R., Segina, R. K., and Stepp, C. E. (2019). Pitch shifting with the commercially available eventide eclipse: Intended and unintended changes to the speech signal. *J. Speech Lang. Hear. Res.* 62, 2270–2279. doi: 10.1044/2019\_JSLHR-S-18-0408
- Houde, J. F., and Nagarajan, S. S. (2011). Speech production as state feedback control. *Front. Hum. Neurosci.* 5:82. doi: 10.3389/fnhum.2011.00082
- Houde, J. F., Gill, J., Agnew, Z., Kothare, H., Hickok, G., Parrell, B., et al. (2019). Abnormally increased vocal responses to pitch feedback perturbations in patients with cerebellar degeneration. *J. Acoust. Soc. Am.* 145, EL372–EL378. doi: 10.1121/1.5100910
- Huberdeau, D. M., Krakauer, J. W., and Haith, A. M. (2015). Dual-process decomposition in human sensorimotor adaptation. *Curr. Opin. Neurobiol.* 33, 71–77. doi: 10.1016/j.conb.2015.03.003
- Kearney, E., and Guenther, F. H. (2019). Articulating: The neural mechanisms of speech production. *Lang. Cogn. Neurosci.* 34, 1214–1229. doi: 10.1080/23273798.2019.1589541
- Kearney, E., Nieto-Castañón, A., Weeratunge, H., Falsini, R., Daliri, A., Abur, D., et al. (2020). A simple 3-parameter model for examining adaptation in speech and voice production. *Front. Psychol.* 10:2995. doi: 10.3389/fpsyg.2019.02995
- Kiran, S., and Larson, C. R. (2001). Effect of duration of pitch-shifted feedback on vocal responses in patients with Parkinson's disease. *J. Speech Lang. Hear. Res.* 44, 975–987. doi: 10.1044/1092-4388(2001)076
- Koo, T. K., and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* 15, 155–163. doi: 10.1016/j.jcm.2016.02.012
- Larson, C. R. (1998). Cross-modality influences in speech motor control: The use of pitch shifting for the study of F0 control. *Clin. Commun. Disord.* 31, 489–503. doi: 10.1016/S0021-9924(98)00021-5
- Larson, C. R., Burnett, T. A., Kiran, S., and Hain, T. C. (2000). Effects of pitch-shift velocity on voice F0 responses. *J. Acoust. Soc. Am.* 107, 559–564. doi: 10.1121/1.428323
- Li, W., Zhuang, J., Guo, Z., Jones, J. A., Xu, Z., and Liu, H. (2019). Cerebellar contribution to auditory feedback control of speech production: Evidence from patients with spinocerebellar ataxia. *Hum. Brain Mapp.* 40, 4748–4758. doi: 10.1002/hbm.24734
- Liu, H., Wang, E. Q., Metman, L. V., and Larson, C. R. (2012). Vocal responses to perturbations in voice auditory feedback in individuals with Parkinson's disease. *PLoS One* 7:e33629. doi: 10.1371/journal.pone.0033629
- Loucks, T., Chon, H., and Han, W. (2012). Audiovocal integration in adults who stutter. *Int. J. Lang. Commun. Disord.* 47, 451–456. doi: 10.1111/j.1460-6984.2011.00111.x
- Master, C. L., Podolak, O. E., Ciuffreda, K. J., Metzger, K. B., Joshi, N. R., McDonald, C. C., et al. (2020). Utility of pupillary light reflex metrics as a physiologic biomarker for adolescent sport-related concussion. *JAMA Ophthalmol.* 138, 1135–1141. doi: 10.1001/jamaophthalmol.2020.3466
- Minorsky, N. (1922). Directional stability of automatically steered bodies. *J. Am. Soc. Nav. Eng.* 34, 280–309. doi: 10.1111/j.1559-3584.1922.tb04958.x
- Natke, U., and Kalveram, T. (2001). Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables. *J. Speech Lang. Hear. Res.* 44, 577–584. doi: 10.1044/1092-4388(2001)045
- Pamplona, V. F. (2008). *Photorealistic models for pupil light reflex and iridal pattern deformation*. Available online at: <https://www.frontiersin.org>

//lume.ufrgs.br/bitstream/handle/10183/15309/000677246.pdf;jsessionid=0CBDC969CA82E2EC4E26AD99A72AD927?sequence=1 (accessed April 12, 2022).

- Parkinson, A., Behroozmand, R., Ibrahim, N., Korzyukov, O., Larson, C., and Robin, D. (2014). Effective connectivity associated with auditory error detection in musicians with absolute pitch. *Front. Neurosci.* 8:46. doi: 10.3389/fnins.2014
- Ranasinghe, K. G., Gill, J. S., Kothare, H., Beagle, A. J., Mizuiri, D., Honma, S. M., et al. (2017). Abnormal vocal behavior predicts executive and memory deficits in Alzheimer's disease. *Neurobiol. Aging* 52, 71–80. doi: 10.1016/j.neurobiolaging.2016.12.020
- Rubin, L. S. (1980). Pupillometric studies of alcoholism. *Int. J. Neurosci.* 11, 301–308. doi: 10.3109/00207458009147594
- Russo, N., Larson, C. R., and Kraus, N. (2008). Audio-vocal system regulation in children with autism spectrum disorders. *Exp. Brain Res.* 188, 111–124. doi: 10.1007/s00221-008-1348-2
- Sares, A. G., Deroche, M. L. D., Ohashi, H., Shiller, D. M., and Gracco, V. L. (2020). Neural correlates of vocal pitch compensation in individuals who stutter. *Front. Hum. Neurosci.* 14:18. doi: 10.3389/fnhum.2020.00018
- Sares, A. G., Deroche, M. L. D., Shiller, D. M., and Gracco, V. L. (2018). Timing variability of sensorimotor integration during vocalization in individuals who stutter. *Sci. Rep.* 8:16340. doi: 10.1038/s41598-018-34517-1
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin* 2, 110–114.
- Scheerer, N. E., Jacobson, D. S., and Jones, J. A. (2016). Sensorimotor learning in children and adults: Exposure to frequency-altered auditory feedback during speech production. *Neuroscience* 314, 106–115. doi: 10.1016/j.neuroscience.2015.11.037
- Scheerer, N. E., Liu, H., and Jones, J. A. (2013). The developmental trajectory of vocal and event-related potential responses to frequency-altered auditory feedback. *Eur. J. Neurosci.* 38, 3189–3200. doi: 10.1111/ejn.12301
- Smith, D. J., Stepp, C. E., Guenther, F. H., and Kearney, E. (2020). Contributions of auditory and somatosensory feedback to vocal motor control. *J. Speech Lang. Hear. Res.* 63, 2039–2053. doi: 10.1044/2020\_JSLHR-19-00296
- Smith, M. A., Ghazizadeh, A., and Shadmehr, R. (2006). Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS Biol.* 4:e179. doi: 10.1371/journal.pbio.0040179
- Stergiou, V., Fotiou, D., Tsiptsios, D., Haidich, B., Nakou, M., Giantselidis, C., et al. (2009). Pupillometric findings in patients with Parkinson's disease and cognitive disorder. *Int. J. Psychophysiol.* 72, 97–101. doi: 10.1016/j.ijpsycho.2008.10.010
- Tales, A., Troscianko, T., Lush, D., Haworth, J., Wilcock, G. K., and Butler, S. R. (2001). The pupillary light reflex in aging and Alzheimer's disease. *Aging (Milan)* 13, 473–478.
- Thompson, H. S. (2003). The vitality of the pupil: A history of the clinical use of the pupil as an indicator of visual potential. *J. Neuroophthalmol.* 23, 213–224. doi: 10.1097/00041327-200309000-00007
- Thoroughman, K. A., and Shadmehr, R. (2000). Learning of action through adaptive combination of motor primitives. *Nature* 407, 742–747. doi: 10.1038/35037588
- Xu, Y., Larson, C. R., Bauer, J. J., and Hain, T. C. (2004). Compensation for pitch-shifted auditory feedback during the production of mandarin tone sequences. *J. Acoust. Soc. Am.* 116, 1168–1178. doi: 10.1121/1.1763952
- Zarate, J. M., Wood, S., and Zatorre, R. J. (2010). Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. *Neuropsychologia* 48, 607–618. doi: 10.1016/j.neuropsychologia.2009.10.025