Bayesian inference of state feedback control parameters for f_o perturbation responses in cerebellar ataxia

Jessica L. Gaines[1*], Kwang S. Kim[2], Ben Parrell[3], Vikram Ramanarayanan[4,5], Alvincé L. Pongos[1], Srikantan S. Nagarajan[4,6], John F. Houde[4]

1 UC Berkeley - UCSF Graduate Program in Bioengineering, University of California-San Francisco, San Francisco, California, USA

2 Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, Indiana, USA

3 Department of Communication Sciences and Disorders, University of Wisconsin-Madison, Madison, Wisconsin, USA

4 Department of Otolaryngology, University of California-San Francisco, San Francisco, California, USA

5 Modality.ai, San Francisco, California, USA

6 Department of Radiology, University of California-San Francisco, San Francisco, California, USA

* jessica.gaines@berkeley.edu

Abstract

Behavioral speech tasks have been widely used to understand the mechanisms of speech motor control in healthy speakers as well as in various clinical populations. However, determining which neural functions differ between healthy speakers and clinical populations based on behavioral data alone is difficult because multiple mechanisms may lead to the same behavioral differences. For example, individuals with cerebellar ataxia (CA) produce abnormally large compensatory responses to pitch perturbations in their auditory feedback, compared to controls, but this pattern could have many explanations.

Here, computational modeling techniques were used to address this challenge. Bayesian inference was used to fit a state feedback control (SFC) model of voice fundamental frequency (f_o) control to the behavioral pitch perturbation responses of individuals with CA and healthy controls. This fitting process resulted in estimates of posterior likelihood distributions of five model parameters (sensory feedback delays, absolute and relative levels of auditory and somatosensory feedback noise, and controller gain), which were compared between the two groups. Results suggest that the CA group may proportionally weight auditory and somatosensory feedback differently from the control group. Specifically, the CA group showed a greater relative sensitivity to auditory feedback than the control group. There were also large group differences in the controller gain parameter, suggesting increased motor output responses to target errors in the CA group. These modeling results generate hypotheses about how CA may affect the speech motor system, which could help guide future empirical investigations in CA. This study also demonstrates the overall proof-of-principle of using this Bayesian inference approach to understand behavioral speech data in terms of interpretable parameters of speech motor control models.

Author summary

Cerebellar ataxia is a condition characterized by a loss of coordination in the control of muscle movements, including those required for speech, due to damage in the cerebellar region of the brain. Behavioral speech experiments have been used to understand this disorder's impact on speech motor control, but the results can be ambiguous to interpret. In this study, we fit a computational model of the neural speech motor control system to the speech data of individuals with cerebellar ataxia and that of healthy controls to determine what differences in model parameters best explain how the two groups differ in their control of vocal pitch. We found that group differences may be explained by increased sensitivity to *auditory feedback prediction errors* (differences between the actual sound speakers hear of their own speech as they produce it and the sound they expected to hear) and increased motor response in individuals with cerebellar ataxia. These computational results help us understand how cerebellar ataxia impacts speech motor control, and this general approach can also be applied to

study other neurological speech disorders.

Introduction

Altered auditory feedback experiments have been widely used to probe the mechanisms of speech motor control. In this class of experiments, participants speak while listening to a digitally-altered version of their own voice through headphones. This allows experimenters to observe how the speech motor system responds to a perceived error in some acoustic property of the voice such as pitch, formant frequencies, or loudness (for a review, see [1]). In a pitch perturbation task, the auditory feedback of participants' production of a sustained vowel sound is, at an unexpected time and for a brief period (usually a fraction of a second), digitally altered to have a higher or lower fundamental frequency (f_o; which is perceived as vocal pitch) than was actually produced. Participants tend to compensate for this perceived error within the ongoing production by shifting their produced pitch in the opposite direction of the perturbation, demonstrating that people use auditory feedback for online control during speech. Variations of this task have been used to inform the understanding of speech motor control in typical speakers [2–4] and, more recently, to compare the responses of healthy speakers with those of various clinical populations (e.g., Alzheimer's disease [5]; Parkinson's disease [6]; Hyperfunctional voice disorders [7]; Laryngeal dystonia [8]; Cerebellar ataxia [9, 10]).

However, due to the complexity of the speech motor system, it remains difficult to determine what specific mechanisms may lead to the observed differences in behavioral results in different populations. A difference in how two groups respond to a mid-trial pitch perturbation may create multiple hypotheses about the differences in the speech motor control system between the two groups. A mechanistic computational model can thus be a powerful tool in evaluating these hypotheses by simulating the effects of specific model changes on the observable output. Tuning the parameters of a computational model to fit observed behavioral data has the potential to distinguish how different model components contribute to the observed effect [11–16].

A recent example of mechanistic ambiguity in behavioral speech data is the increased response to pitch perturbations observed in individuals with cerebellar ataxia

(CA; [9, 10]). CA is a condition characterized by a loss of coordination in limb movements, eye movements, and gait, as well as speech symptoms such as articulatory impreciseness, harsh or breathy voice, slowed speech, hypernasality, excessive variation of loudness, scanning speech, and/or impaired timing of voice onset [17, 18]. In CA these movement-related symptoms are caused by cerebellar lesions or degeneration, which is consistent with the idea of the cerebellum as a likely neural substrate for *internal models*, or neural representations of the body used in neuromotor control [19, 20]. However, open questions remain about the precise impact of cerebellar lesions on internal models and the speech motor control system.

Previous pitch feedback perturbation studies involving individuals with CA have examined the effects of CA on speech motor control. Houde et al. [9] found that individuals with CA displayed a heightened response to a 400 ms, mid-utterance pitch perturbation of 100 cents, with the peak of the CA group average response observed to be twice as high as that of the control group average. Around 300 ms after the end of the perturbation, however, the response of the CA group had fallen such that there was no significant difference between the average normalized pitch of the two groups during the time period from 0.7 to 1.0 s. The latency of the peak response was approximately the same for the two groups. These results are reproduced in Fig 1. Similar results were observed by Li et al. [10] using a slightly different paradigm with a 200 ms perturbation of 200 or 500 cents.

Houde et al. [9] proposed two possibilities to explain these findings:

Hypothesis 1: Compared to the control group, the CA group exhibits increased reliance on auditory feedback in comparison with somatosensory feedback.

Hypothesis 2: Compared to the control group, the CA group displays an increased reliance on all types of sensory feedback collectively due to impairment of the feedforward system.

The current study further investigated these two hypotheses by estimating parameter values to fit a computational model of voice f_o to the empirical pitch perturbation responses of CA and control groups (as observed in Houde et al. [9]). A recently developed Bayesian inference method called simulation-based inference (SBI; `https://sbi-dev.github.io/sbi/`; [21, 22] was used to estimate posterior likelihood distributions for each model parameter based on each data set. The above hypotheses
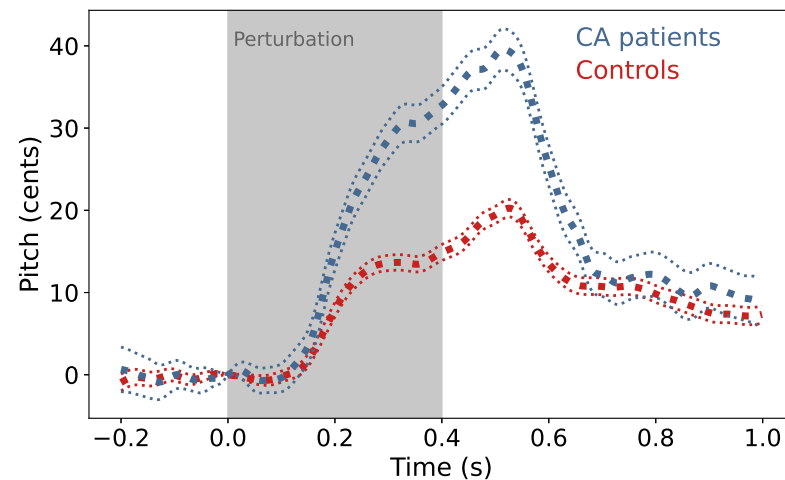
Fig 1. Behavioral data. Group averaged response (thick dotted line) to a 400 ms pitch perturbation of 100 cents for the CA group (blue) and control group (red) [9]. The thin dotted line indicates standard error. The CA group response showed a significantly larger magnitude than that of the control group with no change in peak latency.

were investigated by comparing posterior probability distributions of parameters between the two groups and observing the impact of parameter ablation on the quality of model fit.

Model

Overview of state feedback control

The computational model used in this investigation implements a state feedback control (SFC) architecture to simulate a participant's response to an unexpected pitch perturbation during a sustained vowel production. The theory of state feedback control has been well established as a plausible neural mechanism for non-speech motor tasks [23, 24] and has more recently been explored in the field of speech motor control [13, 25, 26].

This SFC model of voice control we consider is greatly simplified from the full task of voice control. Our model simulates only the control of f_o in ongoing voice output, which we idealize as controlling the rest-length of a spring in a single damped spring-mass system that is somewhat analogous to the cricothyroid muscle's control of pitch [25, 27]. The controls generated by laryngeal motor cortex are modeled as desired

changes in the rest-length of a single muscle determining f_o in our simplified larynx. 78

The SFC architecture for motor control of this simplified larynx is shown in Fig 2. 79
Motor controls are calculated by comparing the desired laryngeal state to an internal 80
estimate of the current state of the larynx, which is maintained by a process of state 81
prediction and subsequent correction based on auditory and somatosensory feedback 82
signals. The controls are scaled by a tunable controller gain g_c. An efference copy of 83
these commands is used to predict the state of the larynx for the following time step, 84
and consequently the expected auditory and somatosensory feedback. Meanwhile, the 85
actual sensory feedback is generated by the larynx model. This simulated feedback can 86
be altered to simulate the effects of a pitch perturbation experiment. Each sensory 87
feedback signal contains Gaussian noise with zero mean and tunable variance σ and has 88
some tunable delay Δ. The delayed and noisy feedback is compared with the predicted 89
feedback to generate an error signal. The error signals from each feedback modality are 90
used to correct the internal estimate of laryngeal state. The error signals are weighted 91
by a Kalman gain matrix, which is calculated based on the noise in each signal, to 92
determine the appropriate correction to the internal state estimate. Across multiple 93
frames, the auditory output of the larynx model can be compared to a behavioral pitch 94
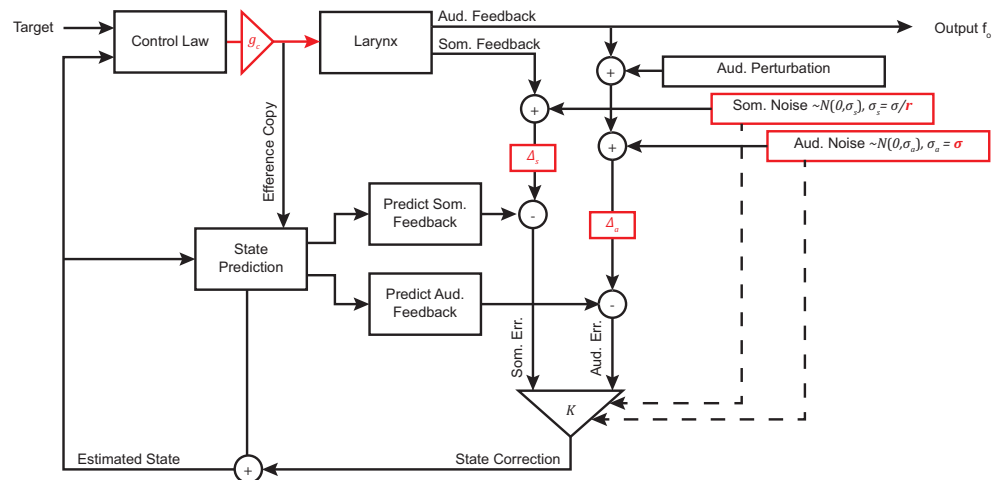perturbation response. 95

Fig 2. Overview of state feedback control. A state feedback model of vocal f_o control where Δ_a is auditory feedback delay, Δ_s is somatosensory feedback delay, σ is overall feedback noise variance, r is feedback noise ratio, and g_c is controller gain. Parameters tuned in this investigation are highlighted in red.

State space representation

The plant of the control system, the larynx, is modeled using a state space representation of a dynamical system (Eq 1,2; [28]). Specifically, the larynx is modeled as a simple spring-mass system in which the length of the spring represents muscle tension on the vocal folds, and is therefore linearly related to vocal pitch. There exist many more complicated and accurate models of the generation of voice by the complex system of muscles controlling the larynx (see [29, 30]); however, for the current study, a very simple model was used to represent broadly the dynamics of the plant that must be controlled.

$$\dot{x}[t] = Ax[t-1] + Bu[t-1] + w[t] \tag{1}$$

$$y[t] = Cx[t] + v[t] \tag{2}$$

In these equations, x is the state (position and velocity of the mass on the spring) of the larynx dynamical system (elsewhere referred to as "larynx"), t is the current time step, u is the control input, and w is Gaussian process noise. The variable y is a vector containing auditory and somatosensory feedback and v is Gaussian measurement noise. A and B are matrices encoding the spring constant, damping constant, and mass of the system, which would be determined by the anatomy of the individual, and C encodes the transformation from state to sensory feedback. These three matrices are constants of the larynx plant (Eq 3) and remain fixed throughout the study at the values used by Houde et al. [27].

$$A = \begin{bmatrix} 0 & 1 \\ -160000 & -1600 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 160000 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \tag{3}$$

Process noise w is defined such that

$$Q = E[ww^T] = \begin{bmatrix} \sigma_q & 0 \\ 0 & \sigma_q \end{bmatrix} \tag{4}$$

where σ_q is the variance of noise applied to each element of laryngeal state. σ_q was held constant at 1e-8 throughout the simulations, a value that was found to produce stable simulator output.

Measurement noise v is defined such that

$$R = E[vv^T] = \begin{bmatrix} \sigma_a & 0 \\ 0 & \sigma_s \end{bmatrix} \tag{5}$$

where σ_a is the variance of Gaussian noise on the auditory feedback signal and σ_s is the variance of Gaussian noise on the somatosensory signal.

Another state space system defined by $A = 0$, $B = 1$, $C = 1$ is connected in series preceding this main system, serving to integrate over the main system and thus produce a laryngeal position from the motor commands output by the main system. Additionally, the full continuous system is discretized using zero-order hold methods with a 0.004 s sampling time.

Pitch perturbation

The variable y is a two-element vector containing auditory and somatosensory feedback from the larynx. The pitch perturbation is implemented as an addend to the auditory component of the vector as follows:

$$y[t] = \begin{bmatrix} y_a[t] \\ y_s[t] \end{bmatrix} + \begin{bmatrix} -100 \\ 0 \end{bmatrix}, 0 \le t < 400\text{ms} \tag{6}$$

March 11, 2024

Controller

The controller command to change the state of the larynx is defined by

$$u[t] = g_c * (x_{\text{target}}[t] - x_{\text{estimate}}[t]) \tag{7}$$

where g_c is the controller gain, x_{target} is the desired laryngeal state and x_{estimate} is the internal estimate of the laryngeal state.

Observer

The observer provides the internal estimate of the state of the larynx through iterative prediction of state and subsequent update of the prediction using sensory feedback. The state is predicted by

$$x_{\text{predicted}}[t] = A x_{\text{estimate}}[t-1] + B u[t-1] \tag{8}$$

where A and B matrices are identical to those used to define the larynx (Eq 3)and $u[t-1]$ is the efference copy of the controller commands from the previous timestep (Eq 7). Predicted state is used to predict the sensory feedback by

$$y_{\text{predicted}}[t] = C x_{\text{predicted}}[t] \tag{9}$$

where C is identical to the state-to-feedback transformation matrix of the laryngeal system (Eq 3). The error between the predicted sensory feedback and the actual sensory feedback is then determined by

$$y_{\text{error}}[t-\Delta] = y[t-\Delta] - y_{\text{predicted}}[t-\Delta], \text{ where } \Delta = \begin{bmatrix} \Delta_a \\ \Delta_s \end{bmatrix} \tag{10}$$

where Δ_a is auditory feedback delay and Δ_s is somatosensory feedback delay. Sensory feedback error is used to estimate the error in laryngeal state by

March 11, 2024

$$x_{\text{error}}[t] = K * y_{\text{error}}[t - \Delta] \tag{11}$$

where K is the steady-state Kalman gain. Kalman gain is the optimal gain matrix calculated by first solving the discrete-time algebraic Ricatti equation (Eq 12) using the Python Control Systems Library (python-control; [31]) for $\hat{P}$ where the equation is satisfied when $P = \hat{P}$.

$$A^T P A - P - A^T P B (B^T P B + R)^{-1} B^T P A + Q = 0 \tag{12}$$

In the above equation, A and B are defined by the larynx dynamical system (Eq 3) and Q and R are defined by noise variances (Eq 4, Eq 5). Optimal Kalman gain is then calculated as follows where C is defined by the larynx dynamical system (Eq 3).

$$K = \hat{P} C^T (R C \hat{P} C^T)^{-1} \tag{13}$$

Finally, the predicted laryngeal state is updated by the error to find the estimated state using

$$x_{\text{estimate}}[t] = x_{\text{predicted}}[t] + x_{\text{error}}[t] \tag{14}$$

Tunable Parameters

The values of a set of tunable parameters in the SFC model affect the shape of the model output, a simulated time course of vocal f_o in response to a mid-trial pitch perturbation. These parameters include the sensory delay parameters Δ_a and Δ_s, controller gain g_c, and sensory feedback noise variance parameters. In order to separate the effects of absolute noise level from the relative amount of noise between the two sensory feedback modalities, the noise variance parameters σ_a and σ_s were

parameterized to σ and r such that $\sigma_a = \sigma$ and $\sigma_s = \sigma/r$. Thus σ represents the variance of sensory feedback noise overall in both auditory and somatosensory feedback modalities and r represents the ratio of auditory feedback noise variance to somatosensory feedback noise variance. Overall feedback noise variance σ was explored on a $\log_{10}$ scale in order to more effectively search many orders of magnitude of noise variance. Thus the parameter set tuned in this investigation was $\theta = \{\Delta_a, \Delta_s, \sigma, r, g_c\}$

Expression of Study Hypotheses in the Context of the SFC Model

Houde et al. [9] proposed two possible explanations for the behavioral pitch perturbation response differences between CA and control groups: 1) an increase in the CA group's reliance on auditory feedback, or 2) an increase in the CA group's reliance on all types of sensory feedback collectively compared to that of the control group. These hypotheses can be examined using the feedback noise parameters, which are inversely related to reliance on sensory feedback through the calculation of Kalman gain K. An increase in the variance of the noise distribution of a particular feedback modality results in a lower Kalman gain on that feedback signal. Thus the hypotheses can be conceptualized in terms of the SFC model as follows:

Hypothesis 1: Compared to the control group, the CA group exhibits increased reliance on auditory feedback over somatosensory feedback. Mathematically, this can be expressed as:

$$\left(\frac{K_a}{K_s}\right)_{\text{CA}} > \left(\frac{K_a}{K_s}\right)_{\text{control}} \tag{15}$$

Since Kalman gain is inversely related to noise variance, this is equivalent to:

$$\left(\frac{\sigma_a}{\sigma_s}\right)_{\text{CA}} < \left(\frac{\sigma_a}{\sigma_s}\right)_{\text{control}} \tag{16}$$

or

$$r_{\text{CA}} < r_{\text{control}} \tag{17}$$

Hypothesis 2: Compared to the control group, the CA group displays an increased reliance on all types of sensory feedback collectively due to impairment of the feedforward system. Mathematically, this can be expressed as:

$$(K_a)_{\text{CA}} > (K_a)_{\text{control}} \quad \text{and} \quad (K_s)_{\text{CA}} > (K_s)_{\text{control}} \tag{18}$$

which is equivalent to:

$$(\sigma_a)_{\text{CA}} < (\sigma_a)_{\text{control}} \quad \text{and} \quad (\sigma_s)_{\text{CA}} < (\sigma_s)_{\text{control}} \tag{19}$$

or simply:

$$\sigma_{\text{CA}} < \sigma_{\text{control}} \tag{20}$$

Thus Hypothesis 1 can be examined by comparing the feedback noise ratio parameter r between the two groups, while Hypothesis 2 can be examined by comparing the overall feedback noise variance parameter σ between groups. The finding of a smaller value for r in the CA group compared with the control group would support Hypothesis 1, while a smaller value for σ in the CA group would support Hypothesis 2.

Results

Parameter effect size

Bayesian inference was used to fit a five-parameter SFC model of voice f_o to the empirical pitch perturbation responses of CA and control groups as observed in Houde et al. [9]. For each empirical data set, simulation-based inference (SBI;

https://github.com/sbi-dev/sbi/; [21, 22] was used to generate a posterior 200

likelihood distribution across values for each parameter. To improve robustness, the 201

posterior distributions from 10 repetitions of the inference procedure were combined. 202

Fig. 3 shows the combined posterior distributions generated for each parameter, with 203

the median and 95% Bayesian credible interval marked in each distribution. Since tests 204

of significance lose meaning under the high statistical power of simulated data [32], 205

effect size was used to quantify which parameters were most different between control 206

and CA groups. Glass's delta effect size was used because it is designed to compare 207

populations with unequal variances. The mean and standard error of effect size 208

calculated from 100 bootstrap samples of size 1000 are annotated on each subplot. 209

Controller gain and feedback noise ratio had the largest effect size, while somatosensory 210

feedback delay, auditory feedback delay and overall feedback noise variance had 211

relatively small effect size. Notably, feedback noise ratio had a much larger effect size 212

than the overall variance of sensory feedback noise, suggesting that the relative amount 213

of noise between feedback modalities contributed much more to group differences than 214

the overall level of feedback noise. All of the distributions are unimodal, which suggests 215

there was a single optimal parameter set rather than several local optima. 216
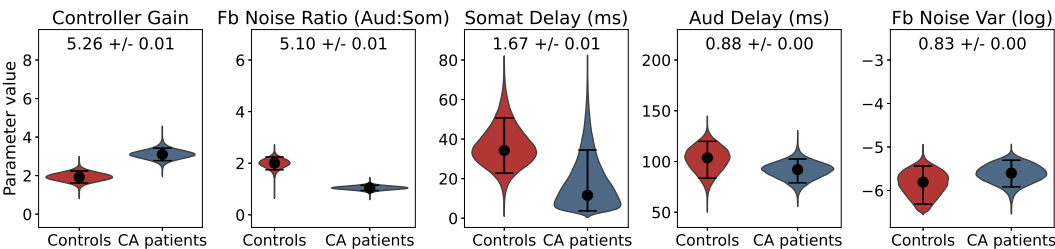

Fig 3. Parameter likelihood distributions. Posterior likelihood distributions (pooled from 10 repetitions of inference) for each parameter, with control group distributions on the left (red) and CA group distributions on the right (blue). The 95% Bayesian credible interval is indicated by horizontal bars and the median value of each distribution is indicated by a black dot. Glass's delta effect size (mean and standard error from bootstrap samples) is printed at the top of each subplot.

Model fit   217

The median of each marginal likelihood distribution was chosen as the inferred 218

parameter set for each participant group (see Table 1). The inferred parameter sets 219

were validated by using them in the SFC simulator to check that the results were 220

broadly similar to the empirical data from each group. The close alignment of the tuned $_{221}$ model with the empirical data can be seen in Fig 4. The mean of 100 simulator outputs $_{222}$ using each group's inferred parameter set were plotted to account for stochasticity $_{223}$ within the SFC simulator. The standard error of these simulator outputs was also $_{224}$ plotted, but was too small to be distinguished from the mean. Root mean square error $_{225}$ (RMSE) between the simulator output and the behavioral data was used to quantify the $_{226}$ quality of model fit for each group with a statistic not explicitly optimized during $_{227}$ training. RMSE (mean $\pm$ standard error across 100 simulations) was 0.8613 ± 0.0002 $_{228}$ cents for the control group (4.06% of the range of the data) and 1.5544 ± 0.0003 cents $_{229}$ for the CA group (3.84% of the range of the data). $_{230}$
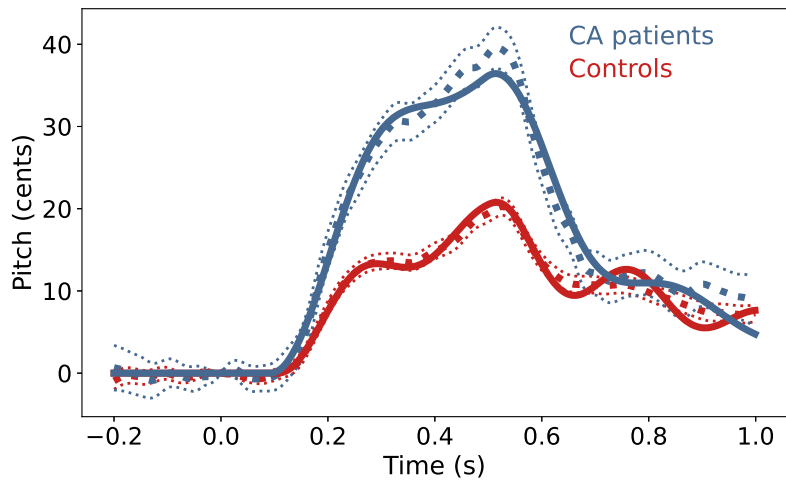

Fig 4. Model fit. The simulator output (solid lines) using the inferred parameter set for each group closely aligned with the behavioral data (dotted lines) previously seen in Fig 1. For simulator output and behavioral data, the mean is plotted with a thick line and standard error with a thin line, although standard error on simulator output was too small to visibly distinguish from the mean. Blue lines indicate data associated with the CA group and the corresponding simulator output, while red lines indicate those associated with the control group.

Table 1. Inferred values for control and CA groups.

Parameter	Aud Delay (ms)	Somat Delay (ms)	Fb Noise Var ($\log_{10}$ scale)	Fb Noise Ratio (Aud:Somat)	Controller Gain
Symbol	Δ_a	Δ_s	σ	r	g_c
Control	102.7	35.3	-5.8	2.0	1.9
CA	91.5	15.5	-5.6	1.0	3.1

Ablation study

Above, we have shown a successful model fit of the pitch perturbation responses for CA and control groups and quantified the parameter distribution differences between groups; however, the impact of these parameter differences on the simulator output are not immediately clear from previous results. Thus, ablation techniques [33] were used to understand the extent to which differences in inferred parameter values may translate to meaningful changes in model output. In this ablation study, the impact of each parameter on group differences was assessed by fitting the behavioral data set from the CA group with a series of reduced models. For each reduced model, one of the five original parameters was fixed to the control group's inferred value for that parameter (as listed in Table 1) and a model composed of the four remaining parameters was used to fit the behavioral data. This allowed us to quantify the impact of each parameter on group differences and simulate the pitch perturbation response we might expect to see if individuals with CA did not differ from controls in terms of each parameter.

As seen in Fig 5, the quality of the model fit for the CA group was most greatly impacted by fixing the feedback noise ratio parameter, followed by the controller gain parameter. This provided additional evidence that these parameters were the most impactful in explaining the differences between the pitch perturbation responses of the two groups. Fixing either the somatosensory feedback delay or feedback noise variance parameter did not seem to change the quality of model fit from that of the full five-parameter model, supporting the idea that these parameters have low impact in explaining the differences between CA and control groups. Once again, the impact of the feedback noise ratio parameter was much greater than that of the feedback noise variance parameter, showing that the absolute amount of noise in each feedback signal is less impactful than the relative amount of noise between signals. To validate the method, the reduced models were also used to fit the control group data. As expected, RMSE between the simulated and behavioral control group responses remained similar to that of the full model, showing that changes in the RMSE of the CA group in the reduced model fit are attributable to group differences in parameter values rather than an artifact of reducing the number of parameters.

For each four-parameter model, the optimized simulator output of control and CA
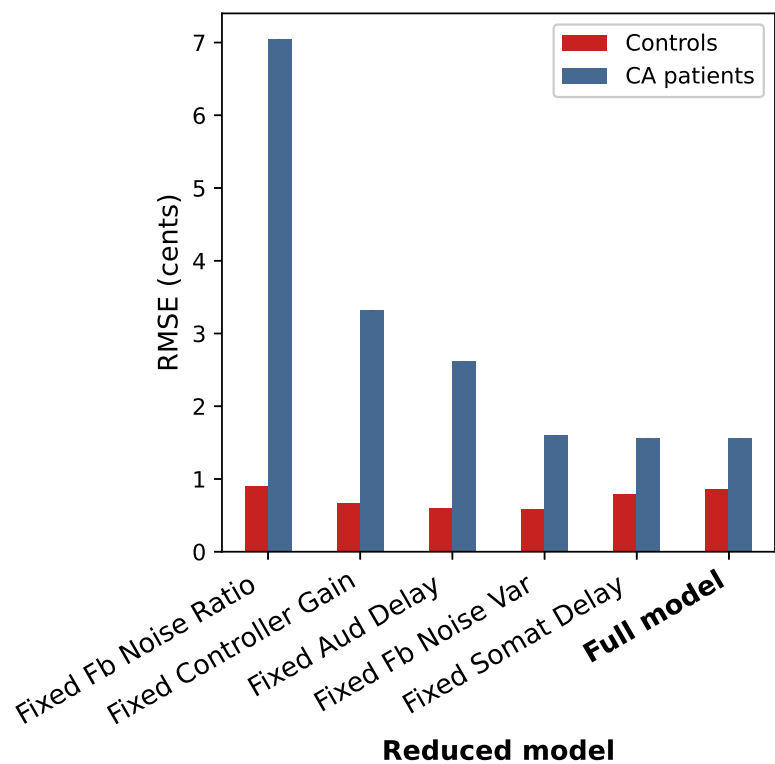
Fig 5. Quality of fit for reduced models. Quality of fit for each 4-parameter model with one parameter fixed to the inferred value for the control group. The quality of each model's fit is quantified by RMSE between empirical data and optimized model output (mean RMSE across 100 simulator runs).

group data are shown in Fig 6, along with the posterior distributions of the remaining parameters. The optimized simulator output for each four-parameter model provides insight regarding the impact of each parameter on the shape of the pitch perturbation response, while the distributions of the remaining parameters indicate which parameters may interact with the ablated parameter.

The results of the ablation of the feedback noise ratio parameter are shown in Fig 6A. We can see from the posterior distributions that while the somatosensory feedback delay, feedback noise variance, and auditory feedback delay parameter distributions remained mostly similar to those of the full model, the controller gain parameter distribution for the CA group shifted up to a much larger median value. This allowed the simulated response to nearly reach the peak response of the behavioral CA data, but it dipped far below the behavioral CA data during the period from 0.7 to 1.0 s after perturbation onset.
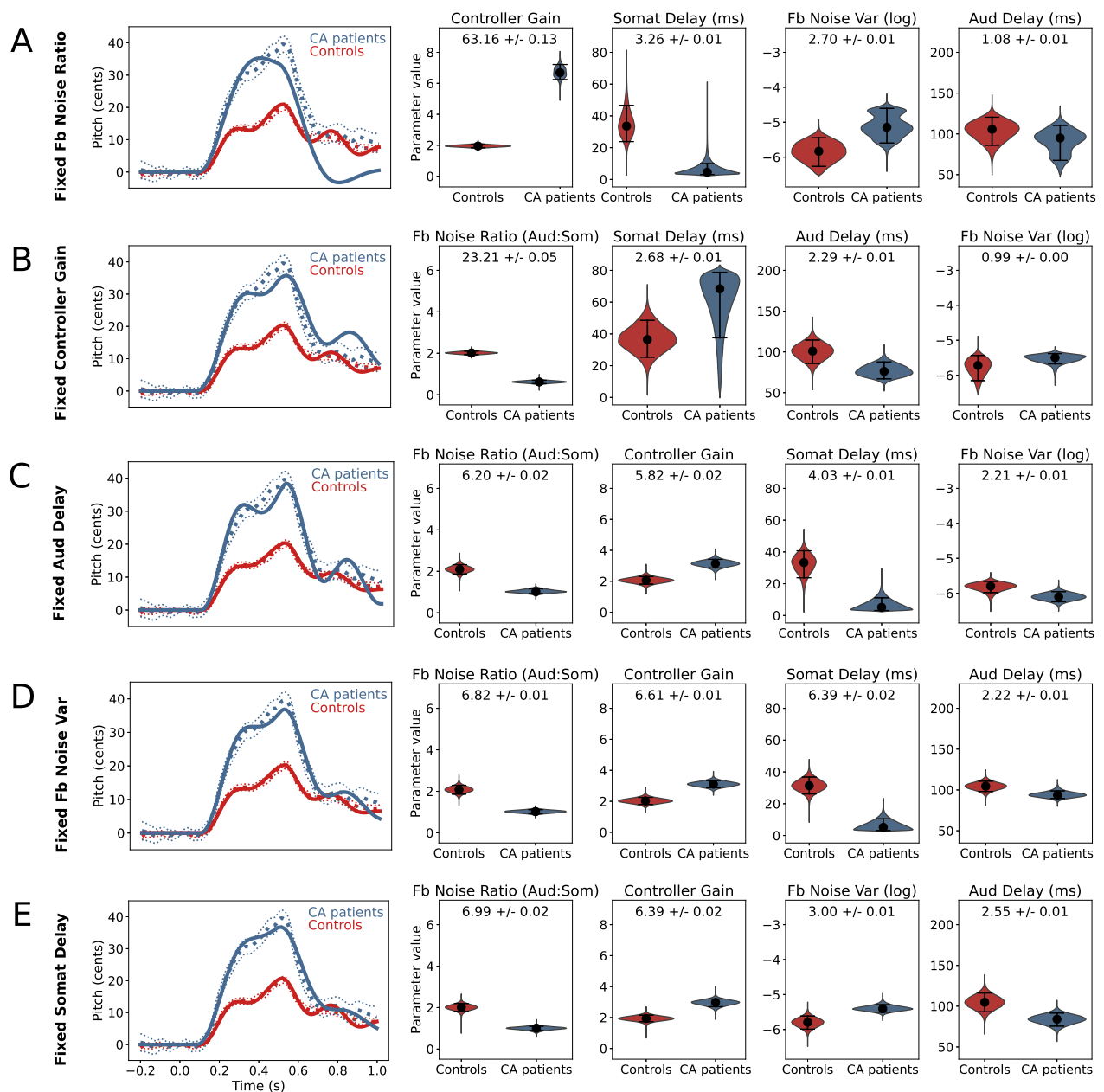
Fig 6. Reduced models: simulator output and posterior distributions of remaining parameters. Inference results for each of five reduced models in which a single parameter is fixed to the inferred value of the control group from the full model fit and the remaining four parameters are used to fit the behavioral pitch perturbation data. Shown for each reduced model are the posterior distributions of the four remaining parameters and the simulator output (mean of 100 simulations) using the inferred reduced parameter set.

Alternately, when the controller gain parameter was ablated, the feedback noise ratio greatly increased in effect size as the median value of the CA group distribution shifted down from 1.04 to 0.62. The median value of the somatosensory delay parameter also

increased but the effect size of this parameter increased only slightly due to the large variance of the distribution. These changes allowed the simulator output to nearly reach the peak magnitude of the behavioral data of the CA group, but with the controller gain fixed at the value of the control group, the optimized simulator output was higher and more oscillatory than the behavioral data during the period from 0.7 to 1.0 s after perturbation onset. Thus contributions from both the feedback noise ratio and controller gain parameters are needed to replicate both the large magnitude of the CA group response and the flat, moderate time course of f_o after the end of the perturbation.

When auditory feedback delay was fixed to the inferred value for the control group (Fig 6C), the CA group response became more oscillatory, contributing to a small increase in error from the behavioral data. Fixing either somatosensory feedback delay or overall feedback noise variance, meanwhile, did not appear to change the model fit. It can be seen in (Fig 6D) that the ablation of feedback noise variance caused a change in the distribution of the somatosensory feedback delay parameter, and vice versa in (Fig 6E). Thus these two parameters, interestingly, seem to have similar effects on model output.

Discussion

In this study, simulation-based Bayesian inference was used to disambiguate possible mechanisms underlying the observed differences in the pitch perturbation responses of individuals with CA and healthy controls by identifying speech motor control model parameter differences between these two groups. The results indicate that most of the differences between the pitch perturbation responses of individuals with CA and healthy controls can be explained by differences in the feedback noise ratio and controller gain. These parameters showed both (a) the largest effect size between groups in comparing the posterior distributions and (b) the greatest loss of fit accuracy for the CA group when fixed to the inferred value of the control group.

Our finding that the feedback noise ratio was lower in the CA group than in the control group with substantial effect size supports Hypothesis 1, which suggests increased reliance on auditory feedback relative to somatosensory feedback in the CA group. In contrast, we did not find evidence in support of Hypothesis 2, the idea that

CA group displays an increased reliance on all types of sensory feedback collectively. The small effect size for the overall sensory feedback noise parameter and the minimal change when this parameter was ablated were inconsistent with this hypothesis.

Here we will discuss how these findings fit into the previous literature regarding pitch perturbation response in individuals with CA, new hypotheses generated by the model, and the strengths and limitations of the modeling techniques used in this analysis.

Evidence for overreliance on auditory feedback

The high effect size of the feedback noise ratio suggests that the CA group may show increased reliance on auditory feedback relative to somatosensory feedback. This may appear to conflict with the results of Li et al. [10], which showed in individuals with spinocerebellar ataxia a decreased cortical P2 response in the right superior temporal gyrus (STG), primary auditory cortex (A1), and supramarginal gyrus (SMG) during a pitch perturbation task. However, while this finding may provide evidence against the idea that the absolute value of auditory feedback error gain is larger in the CA group than the control group $((K_a)_{\text{CA}} > (K_a)_{\text{control}})$, we argue that it does not rule out the idea of increased sensitivity to auditory feedback as we have defined it in this study, that is, relative to somatosensory feedback $(\left(\frac{K_a}{K_s}\right)_{\text{CA}} > \left(\frac{K_a}{K_s}\right)_{\text{control}}$; Hypothesis 1). If both auditory and somatosensory gains are smaller in the CA group compared to controls, it is possible that the resulting ratio between auditory gain and somatosensory gain may still be larger in the CA group compared to the control group (i.e., if the somatosensory gain in the CA group were much smaller compared that of the control group). Defining feedback noise in terms of a ratio between sensory modalities has been previously used by Crevecoeur et al. [23] to model relative sensitivity to visual and proprioceptive feedback in arm reaching. Furthermore, the P2 response may not be a direct measure of gain on auditory feedback error; the functional relevance of P2 response remains controversial and has been suggested to have contributions from multiple sensory modalities including auditory and somatosensory [34]. Although this view is highly speculative, we argue that the measurement of decreased P2 response in the CA group [10] does not completely rule out Hypothesis 1 and that our findings warrant further investigations.

Lack of evidence for overall overreliance on feedback 338

Given the ties between internal models and the cerebellum [19,20], it is indeed somewhat 339
surprising that we did not find evidence of overreliance on sensory feedback overall. It 340
would be reasonable to expect that if the internal models, and thus the state prediction 341
process, are disrupted in CA, sensory feedback would be more heavily weighted to 342
compensate for impairment in the feedforward system [9]. However, the results indicate 343
similar weighting of sensory feedback overall between the two groups, with perhaps even 344
slightly lower weighting (increased noise) of sensory feedback in the CA group. We 345
speculate that perhaps the nature of the disruption of internal models impairs the 346
integration of sensory feedback, which has similar effects to a slightly decreased 347
weighting. Alternatively, perhaps the auditory nature of the task created a result in 348
which auditory feedback appears to be emphasized rather than overall feedback. 349

Novel hypothesis generated by the model: controller gain parameter 350 351

In addition to support for the previously stated hypothesis that individuals with CA 352
show increased reliance on auditory feedback relative to somatosensory feedback, our 353
results lead us to propose an additional hypothesis to be tested in future investigations. 354
The high impact of the controller gain parameter suggests a difference between 355
individuals with CA and healthy controls in the scaling of the motor command to the 356
larynx, which has likely neural substrate in the motor cortex [25]. Although any direct 357
ties between this parameter and the cerebellum are unknown, we speculate that perhaps 358
the increased value of controller gain in the CA group, which could also be interpreted 359
as increased sensitivity to target error (the difference between the intended production 360
and the actual production) may indicate a learned mechanism in the motor cortex to 361
compensate for deficits related to the disorder. Further study is needed to investigate 362
this hypothesis. 363

Efficacy of simulation-based inference for model parameter estimation

Simulation-based inference allowed us to obtain posterior likelihood distributions across each parameter rather than a single set of optimal values. This made it possible to take into account the spread of each distribution in addition to the inferred value and thus determine the effect size of the difference between groups for each parameter. Additionally, in a complex system it is possible to have many locally optimal parameter sets that can achieve high quality fit to the empirical data. Obtaining a unimodal posterior likelihood distribution for each parameter showed that the parameter set selected could approximate a globally optimal solution within the bounds of the prior. Simulation-based inference also offered advantages over other Bayesian methods since it does not require an analytical form of the model (allowing us to analyze a complex model like state feedback control) and is more computationally efficient than other numerical techniques such as Markov Chain Monte Carlo methods [35].

Limitations

It is important to note that the current work is beneficial for generating hypotheses, rather than drawing conclusions. The larynx was modeled here as a simple spring-mass system. While this implementation may approximate many of the dynamics of laryngeal movement, future studies should investigate how results may vary with more detailed models of laryngeal muscles. Additionally, real speakers vary in their vocal tract dimensions, damping, and other properties that are not possible to measure. The model cannot quantitatively reflect all of these properties to represent an actual speaker. Furthermore, these properties were left unchanged in the model throughout the present set of simulations, so the vocal tract properties of at most one speaker were represented. Meanwhile, the behavioral data used in the present study were group averages of many subjects with different physiologies. To account for the averaging of multiple observations, the simulator was run 100 times and the mean output was plotted. However, standard error on model output was so small (< 0.1 cent) that it could not be differentiated from the mean production on the plot. Standard error on the empirical mean was larger, in the range of 0-3 cents, further demonstrating that the model does

not account for the variation that is present among individual speakers. ₃₉₄

Additionally, the present study was limited to the model changes that could be ₃₉₅ captured by five model parameters. The same model architecture was used to model ₃₉₆ both control and CA groups, with all differences captured in the tuning of model ₃₉₇ parameters. Thus possible differences in the structure of neural systems between the ₃₉₈ groups were not tested in the current study. Furthermore, many of the model ₃₉₉ parameters tested here are abstract concepts whose precise neural implementation is not ₄₀₀ yet fully understood. Each parameter that is represented in the model as a single value ₄₀₁ may in fact be the result of many complex processes. Finally, only five parameters were ₄₀₂ considered in the present study. Since training data was generated by testing values for ₄₀₃ each parameter in combination with all other parameters, the number of parameters ₄₀₄ tested was in exponential tradeoff with the resolution of values tested and the range of ₄₀₅ the prior for each one. While the irrelevance of other untested model parameters cannot ₄₀₆ be proven, the tested parameters can be argued to be sufficient since the model output ₄₀₇ for each set of inferred values closely matched the behavioral data. ₄₀₈

Conclusion ₄₀₉

This work has shown that the controller gain and feedback noise ratio parameters have ₄₁₀ high effect size and large contributions to the group differences between the pitch ₄₁₁ perturbation responses of individuals with CA and healthy controls. These results (a) ₄₁₂ provide support for the previous hypothesis that individuals with CA show increased ₄₁₃ sensitivity to auditory feedback prediction error and (b) generate a new hypothesis of ₄₁₄ increased sensitivity to target error in the CA group. Furthermore, this work ₄₁₅ demonstrates the value of simulation-based inference methods in analyzing behavioral ₄₁₆ speech data using tunable parameters of a state feedback control model. ₄₁₇

Materials and Methods ₄₁₈

Simulation-based inference overview ₄₁₉

The SBI package in Python (`https://github.com/sbi-dev/sbi/`; [21, 22]) was used ₄₂₀ to obtain posterior likelihood distributions of five tunable parameters in the SFC model ₄₂₁

(see Model section for an overview of SFC) for the CA and control group average pitch perturbation responses observed in Houde et al. [9]. As detailed in Fig 7, SBI takes as input a computational model with a finite set of input parameters ("simulator"), a prior distribution for each parameter, and an empirical observation analogous to the output of the simulator. It generates a data set by running the simulator with inputs from the prior distributions of the parameters, and then, using the Sequential Neural Posterior Estimation (SNPE) option for inference, trains a deep neural density estimator to predict the posterior distribution of parameters given the empirical observation.
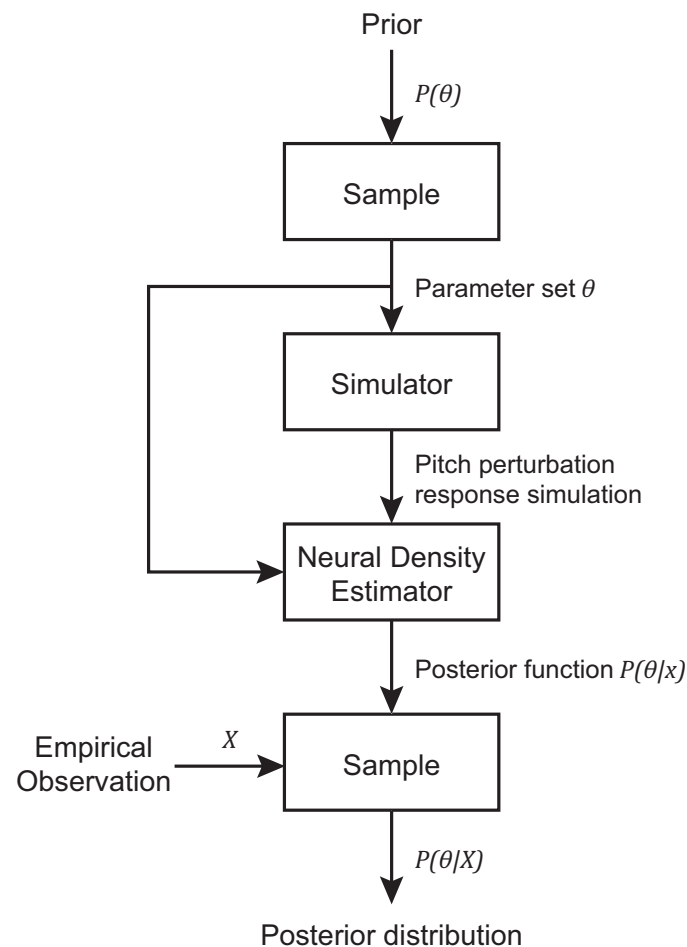

Fig 7. SBI overview. A pipeline for performing simulation-based inference [21, 22]. Parameter values are inferred for a particular empirical observation given a mechanistic model ("simulator") and a prior distribution for each tunable parameter.

Simulator

A Python implementation (`https://github.com/jessicagaines/1d-sfc`) of a state feedback control model of f_o [25, 27] was used as the simulator (see Model section for more details). The input parameter set included the following parameters: auditory feedback delay Δ_a, somatosensory feedback delay Δ_s, controller gain g_c, feedback noise variance σ, and feedback noise ratio r. The parameters σ and r are parameterizations of auditory feedback noise variance σ_a and somatosensory feedback noise variance σ_s such that $\sigma_a = \sigma$ and $\sigma_s = \sigma/r$. This idea of exploring relative levels of feedback noise between different sensory modalities was also used by Crevecoeur et al. [23] in their state feedback control model of arm reaching. The output simulated a time course of voice f_o in response to a -100 cent, 400 ms, mid-utterance perturbation of f_o feedback. Random noise was added to model output during training as Jin et al. [36] found that this increased reconstruction accuracy. Uniform noise distributions of increasing width were tested and since the quality of model fit stopped improving for training noise distribution wider than 7 cents, noise with distribution $\sim U(-3.5, 3.5)$ was added to the simulator output during training.

Prior distribution of parameters

The SBI inference procedure requires the input of a prior distribution for each parameter, which defines the search space of the training data. A uniform prior was used for each parameter. For some parameters, the bounds of the prior could be estimated from measurable quantities. For example, Abbs and Gracco [37] indicate that latencies in response to a somatosensory perturbation are on the order of tens of milliseconds, so a prior of 3 to 80 ms was used for somatosensory delay. Latencies to auditory perturbations, meanwhile, have been reported in the range of 100-200 ms [38], so a prior of 50 to 200 ms was used. Measurable latencies may be greater than delays since they include motor response time, so the lower bounds were set lower than the measured response latencies. The lower bound for somatosensory delay is quite low, in the range of what is typically associated with non-cortical reflex [39], but since no minimum delay value can be definitively measured, we opted not to restrict the prior based on this information. The remaining parameters could not be estimated from

measurable quantities, so wide initial priors were selected to fully explore the space. ₄₆₀

Initial bounds of 0.1 to 10 were selected for the feedback noise ratio and controller gain ₄₆₁

parameters to include two orders of magnitude, and the range of 1e-10 to 1e-1 was ₄₆₂

selected for feedback noise variance. The auditory feedback noise variance parameter ₄₆₃

was converted to a base-10 logarithmic scale to search many orders of magnitude more ₄₆₄

effectively. For likely parameter sets in this regime, the simulator was found to be ₄₆₅

unstable for feedback noise variance less than 1e-6.5, and so the lower bound for this ₄₆₆

parameter was increased to 1e-6.5 to train the likelihood estimator on stable simulator ₄₆₇

outputs. Finally, the results of the wide prior showed that the tails of the posterior ₄₆₈

likelihood distributions of feedback noise variance, feedback noise ratio, and controller ₄₆₉

gain parameters were far from the upper bounds of each prior. The bounds were ₄₇₀

narrowed slightly to increase the search resolution for each parameter and decrease the ₄₇₁

computational resources needed. The final bounds selected are shown in Table 2. This ₄₇₂

choice of prior is validated by the result that the likelihood distributions for each ₄₇₃

parameter (see Fig 3) lie comfortably within these bounds, except for feedback noise ₄₇₄

variance, which was restricted for stability, and somatosensory delay, which by definition ₄₇₅

cannot be less than one frame of simulator operation. ₄₇₆

Table 2. Uniform priors with the following bounds were selected for each parameter.

Parameter	Aud Delay (ms)	Somat Delay (ms)	Fb Noise Var ($\log_{10}$ scale)	Fb Noise Ratio (Aud:Somat)	Controller Gain
Symbol	Δ_a	Δ_s	σ	r	g_c
Lower bound	50	3	-6.5	0.1	0.1
Upper bound	200	80	-3	6	8

Empirical observation ₄₇₇

Behavioral group average pitch perturbation responses from CA and control groups [9] ₄₇₈

as described in the Introduction were used as empirical observations to sample the ₄₇₉

posterior. Each behavioral data set was downsampled from 413 to 300 frames per 1.2 s ₄₈₀

trial to match the output of the simulator. ₄₈₁

Inference

10^5 simulations were used to train the neural density estimator [36]. The posterior was then sampled 10^4 times for each group. To improve robustness, this procedure was repeated 10 times and the samples from each repetition were pooled to obtain the final parameter distributions. A 95% Bayesian credible interval was calculated for each distribution. Glass's delta was used to calculate the effect size of each parameter between groups. To assess the quality of model fit, the median of each pooled distribution was considered the "inferred value" for each parameter and each inferred parameter set was supplied as input to the simulator. To reduce the effects of stochasticity within the simulator, 100 simulations were run with each inferred parameter set and the mean of these was plotted. The quality of the fit was assessed quantitatively by calculating the point-wise root mean square error (RMSE) between the model output and the empirical data. This statistic was not used in training the neural network and therefore provided a separate method of quantifying the success of the model fit.

Ablation study

Finally, an ablation study was used to further understand the impact of each parameter on model output [33]. One at a time, each parameter was ablated by fixing it to the inferred value (the median value of the pooled posterior distribution) of the control group and repeating the inference procedure to generate posterior distributions for the four remaining parameters. The medians of these distributions were once again used in the simulator to assess the quality of fit for each reduced model using RMSE and compare the result to that of the full model. A greater increase in error for a particular reduced model indicated that the parameter ablated in that model had greater impact on group differences.

References

1. Caudrelier T, Rochet-Capellan A. Changes in speech production in response to formant perturbations: An overview of two decades of research.. In: Fuchs S,

Cleland J, Rochet-Capellan A, editors. Speech production and perception: Learning and memory, Vol 6. Berlin, Germany: Peter Lang; 2019. p. 15-75. doi: 10.3726/b15982

2. Burnett TA, Freedland MB, Larson CR, Hain TC. Voice F0 responses to manipulations in pitch feedback. Journal of the Acoustical Society of America. 1998 Jun 1;103:3153-3161. doi: 10.1121/1.423073.

3. Elman J. Effects of frequency-shifted feedback on the pitch of vocal productions. Journal of the Acoustical Society of America. 1981 Jul 1;70(1):45-50. doi: 10.1121/1.386580.

4. Jones JA, Munhall KG. Perception calibration of F0 production: Evidence from feedback perturbation. Journal of the Acoustical Society of America. 2000 Sep 1;108:1246-51. doi: 10.1121/1.1288414.

5. Ranasinghe KG, Gill JS, Beagle AJ, Mizuiri D, Honma SM, Gorno-Tempini ML, et al. Abnormal vocal behavior predicts executive and memory deficits in Alzheimer's disease. Neurobiology of Aging. 2017 Apr;52:71-80. doi: 10.1016/j.neurobiolaging.2016.12.020

6. Liu H, Wang EQ, Verhagen Metman L, Larson CR. Vocal responses to perturbations in voice auditory feedback in individuals with Parkinson's disease. PLoS ONE. 2012 Mar 20;7(3):e33629. doi: 10.1371/journal.pone.0033629.

7. Abur D, Subaciute A, Kapsner-Smith M, Segina RK, Tracy LF, Noordzij JP, et al. Impaired auditory discrimination and auditory-motor integration in hyperfunctional voice disorders. Scientific Reports. 2021 Jun 23;11:13123. doi: 10.1038/s41598-021-92250-8.

8. Kothare H, Schneider S, Mizuiri D, Hinkley L, Bhutada A, Ranasinghe K, et al. Temporal specificity of abnormal neural oscillations during phonatory events in laryngeal dystonia. Brain Commun. 2022 Feb 11;4(2):fcac031. doi: 10.1093/braincomms/fcac031.

9. Houde JF, Gill JS, Agnew Z, Kothare H, Hickok G, Parrell B, et al. Abnormally increased vocal responses to pitch feedback perturbations in patients with

cerebellar degeneration. Journal of the Acoustical Society of America. 2019 May 14;145(5):EL372-EL378. doi: 10.1121/1.5100910.

10. Li W, Zhuang J, Guo Z, Jones JA, Xu Z, Liu H. Cerebellar contribution to auditory feedback control of speech production: Evidence from patients with spinocerebellar ataxia. Human Brain Mapping. 2019 Jul 31;40(16):4748-58. doi: 10.1002/hbm.24734.

11. Civier O, Tasko SM, Guenther FH. Overreliance on auditory feedback may lead to sound/syllable repetitions: Simulations of stuttering and fluency-inducing conditions with a neural model of speech production. Journal of Fluency Disorders. 2010 Sep;35(3):246-79. doi: 10.1016/j.jfludis.2010.05.002.

12. Kearney E, Nieto-Castañón A, Falsini R, Daliri A, Heller Murray ES, Smith DJ, et al. Quantitatively characterizing reflexive responses to pitch perturbations. Frontiers in Human Neuroscience. 2022 Nov 2;16. doi: 10.3389/fnhum.2022.929687.

13. Kim KS, Gaines JL, Parrell B, Ramanarayanan V, Nagarajan SS, Houde JF. Mechanisms of sensorimotor adaptation in a hierarchical state feedback control model of speech. PLOS Computational Biology. 2023 Jul 28;19(7):e1011244. doi: 10.1371/journal.pcbi.1011244.

14. Larson CR, Altman KW, Liu H, Hain TC. Interactions between auditory and somatosensory feedback for voice F0 control. Experimental Brain Research. 2008 Mar 14;187:613-21. doi: 10.1007/s00221-008-1330-z.

15. Patri J-F, Diard J, Perrier P. Modeling sensory preference in speech motor planning: A Bayesian modeling framework. Frontiers in Psychology. 2019 Oct 25;10:2339. doi: 10.3389/fpsyg.2019.02339.

16. Zimmet AM, Cao D, Bastian AJ, Cowan NJ. Cerebellar patients have intact feedback control that can be leveraged to improve reaching. eLife. 2020 Oct 7;9:e53246. doi: 10.7554/eLife.53246.

17. Ackermann H, Hertrich I. Voice onset time in ataxic dysarthria. Brain and Language. 1997 Feb 15;56(3):321-33. doi: 10.1006/brln.1997.1740.

18. Manto M, Bower JM, Bastos Conforto A, Delgado-García JM, Farias da Guarda SN, Gerwig M, et al. Consensus paper: Roles of the cerebellum in motor control–The diversity of ideas on cerebellar involvement in movement. Cerebellum. 2012 Dec 13;11(2):457-487. doi: 10.1007/s12311-011-0331-9.

19. Ito M. Control of mental activities by internal models in the cerebellum. Nature Reviews Neuroscience. 2008 Apr;9:304-13. doi: 10.1038/nrn2332.

20. Wolpert DM, Miall RC, Kawato M. Internal models in the cerebellum. Trends in Cognitive Sciences. 1998 Sep 1;2(9):338-47. doi: 10.1016/s1364-6613(98)01221-2.

21. Cranmer K, Brehmer J, Louppe G. The frontier of simulation-based inference. In: Proceedings of the National Academy of Sciences. 2020 May 29;117(48):30055-62. doi: 10.1073/pnas.1912789117.

22. Tejero-Cantero A, Boelts J, Deistler M, Lueckmann J-M, Durkan C, Gonçalves PJ, et al. Sbi: A toolkit for simulation-based inference. Journal of Open Source Software. 2020 Aug 21;5(52):2505. doi: 10.21105/joss.02505.

23. Crevecoeur F, Munoz DP, Scott SH. Dynamic multisensory integration: Somatosensory speed trumps visual accuracy during feedback control. The Journal of Neuroscience. 2016 Aug 17;36(33):8598-8611. doi: 10.1523/JNEUROSCI.0184-16.2016.

24. Todorov E, Jordan MI. Optimal feedback control as a theory of motor coordination. Nature Neuroscience. 2002 Nov 1;5:1226–1235. doi: 10.1038/nn963.

25. Houde JF, Nagarajan SS. Speech production as state feedback control. Frontiers in Human Neuroscience. 2011 Oct 25;5(82):1-14. doi: 10.3389/fnhum.2011.00082.

26. Parrell B, Ramanarayanan V, Nagarajan SS, Houde JF. The FACTS model of speech motor control: Fusing state estimation and task-based control PLoS Computational Biology. 2019 Sep 3;15(9):e1007321. doi: 10.1371/journal.pcbi.1007321

27. Houde JF, Niziolek CA, Kort N, Agnew, Z, Nagarajan SS. Simulating a state feedback model of speaking. In: 10th International Seminar on Speech

Production; 2014 May 5-8; Cologne, Germany.

`https://www.researchgate.net/publication/284486657_Simulating_a_`

`state_feedback_model_of_speakisp`.

28. Friedland, B. Control system design: an introduction to state-space methods. New York: Dover Publications; 2012.

29. Lucero JC. Dynamics of the two-mass model of the vocal folds: Equilibria, bifurcations, and oscillation region. Journal of the Acoustical Society of America. 1993 Dec 1;94(6):3104-3111. doi: 10.1121/1.407216.

30. Story BH, Titze IR. Voice simulations with a body-cover model of the vocal folds. Journal of the Acoustical Society of America. 1995 Feb 1;97(2):1249-60. doi: 10.1121/1.412234.

31. Fuller S, Greiner B, Moore J, Murray R, van Paassen R, Yorke R. The Python Control Systems Library (python-control). In: 60th IEEE Conference on Decision and Control (CDC). 2021 Dec 14-17;Austin, TX, USA:4875-4881. doi: 10.1109/CDC45484.2021.9683368.

32. White JW, Rassweiler A, Samhouri JF, Stier AC, White C. Ecologists should not use statistical significance tests to interpret simulation model results. Oikos. 2013 Nov 29;123:385-388. doi: 10.1111/j.1600-0706.2013.01073.x.

33. Newell A. A Tutorial on Speech Understanding Systems. In: Reddy DR, editor. Speech Recognition: Invited Papers Presented at the 1974 IEEE Symposium. New York: Academic; 1975. p. 43.

34. Crowley KE, Colrain IM. A review of the evidence for P2 being an independent component process: Age, sleep and modality. Clinical Neurophysiology. 2004 Apr;115:732-44. doi: 10.1016/j.clinph.2003.11.021.

35. Boelts J, Lueckmann J-M, Gao R, Macke JH. Flexible and efficient simulation-based inference for models of decision-making. eLife. 2022 Jul 27;11:e77220. doi: 10.7554/eLife.77220.

36. Jin H, Verma P, Jiang F, Nagarajan SS, Raj A. Bayesian inference of a spectral graph model for brain oscillations. Neuroimage. 2023 Oct 1;279:120278. doi: 10.1016/j.neuroimage.2023.120278.

37. Abbs JH, Gracco VL. Sensorimotor actions in the control of multi-movement speech gestures. Trends in Neuroscience. 1983;6:391-95. doi: 10.1016/0166-2236(83)90173-X.

38. Liu H, Larson CR. Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex Journal of the Acoustical Society of America. 2007 Dec 1;122(6):3671-77. doi: 10.1121/1.2800254.

39. Gracco VL, Abbs JH. Dynamic control of the perioral system during speech: Kinematic analyses of autogenic and nonautogenic sensorimotor processes. Journal of Neurophysiology. 1985 Aug 1;54(2):418-32. doi: 10.1152/jn.1985.54.2.418.