

RESEARCH ARTICLE

A Bayesian Account of Vocal Adaptation to Pitch-Shifted Auditory Feedback

Richard H. R. Hahnloser*, Gagan Narula*

Institute of Neuroinformatics and Neuroscience Center Zurich, University of Zurich and ETH Zurich, Zurich, Switzerland

* rich@ini.ethz.ch (RH); gnarula@ini.ethz.ch (GN)



Abstract

Motor systems are highly adaptive. Both birds and humans compensate for synthetically induced shifts in the pitch (fundamental frequency) of auditory feedback stemming from their vocalizations. Pitch-shift compensation is partial in the sense that large shifts lead to smaller relative compensatory adjustments of vocal pitch than small shifts. Also, compensation is larger in subjects with high motor variability. To formulate a mechanistic description of these findings, we adapt a Bayesian model of error relevance. We assume that vocal-auditory feedback loops in the brain cope optimally with known sensory and motor variability. Based on measurements of motor variability, optimal compensatory responses in our model provide accurate fits to published experimental data. Optimal compensation correctly predicts sensory acuity, which has been estimated in psychophysical experiments as just-noticeable pitch differences. Our model extends the utility of Bayesian approaches to adaptive vocal behaviors.

OPEN ACCESS

Citation: Hahnloser RHR, Narula G (2017) A Bayesian Account of Vocal Adaptation to Pitch-Shifted Auditory Feedback. PLoS ONE 12(1): e0169795. doi:10.1371/journal.pone.0169795

Editor: Eleni Vasilaki, University of Sheffield, UNITED KINGDOM

Received: June 19, 2016

Accepted: December 21, 2016

Published: January 30, 2017

Copyright: © 2017 Hahnloser, Narula. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper.

Funding: Support was provided by the Swiss National Science Foundation Project, 31003A_156976/1 [<http://www.snf.ch/en/funding/projects/projects-in-all-disciplines/Pages/default.aspx>]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Humans and animals can quickly adapt their behaviors to changes in the environment. For example, running barefoot or in shoes through a forest or on sand, all require different gait patterns that are effortlessly recruited when needed. Whether to adapt or not to an unexpected sensory event often depends on whether the latter is self-caused or not. For example, is the cracking sound of a piece of wood caused by one's own foot step or that of a predator? In the latter case, running away may be a good option whereas in the former case doing nothing might be better. Optimal motor adaptation seems to suggest that the brain must determine the cause of sensory input [1] and the degree to which the input is commensurate with one's own motor actions.

The problem of estimating self-caused sensory feedback becomes challenging when sensors are noisy, are subject to damage (e.g. hearing loss due to a ruptured eardrum), or undergo age-related degradation (e.g. presbyopia of the human eye). In order to deal with inaccurate sensory information during possibly imprecise movements, the brain must optimally integrate sensory information with motor plans by weighing the reliability of a given source of

information, in a fashion similar to statistically optimal cue combination [2] or “causal inference” models in Bayesian cognitive psychology [1,3].

A nontrivial relationship between sensory error and motor compensation occurs in the visuo-motor domain: When humans perform arm reaching movements in the presence of force fields (proprioceptive error) or under visual rotations (visual error) that shift the trajectory of the hand away from the straight-line target, subjects partially compensate for the induced errors by selectively adapting the target reaches: large errors induce smaller relative compensation than small errors [3–8].

Another prominent example of partially compensated errors occurs during vocal control by auditory feedback. It has been shown that acoustic feedback has a strong influence on the adaptive control of human voice fundamental frequency: When fundamental frequency (or pitch) of vowels in speaking subjects is slowly shifted without their awareness (using earphones and bone oscillators), subjects compensate partly for the shift within tens of trials [9,10]. This compensation suggests that subjects interpreted the perceived shift to have occurred partially as the result of a production error that requires correction. As documented by a large body of literature, the percent compensation applied by subjects in such experiments declines monotonically with increasing shift in fundamental frequency [11–13].

Humans are not the only ones who compensate partially in such pitch-shift experiments. When pitch feedback into the ears of singing Bengalese finches is shifted using miniature headphones, birds compensate only partially for the shift, and the larger the shift the lesser the birds compensate [14]. All these experiments demonstrating a reduced compensation to large errors raise the possibility of a universal explanatory mechanism.

Two candidate arguments have been proposed to elucidate why both singing birds and humans compensate only partially for perturbed pitch. Liu & Larson argue that failure to correct for large-magnitude errors evidences a self-protection mechanism that prevents environmental sounds from exerting aberrant influence on vocal output [11]. Accordingly, humans subconsciously may interpret the perceived shift as resulting from a change in the environment that is not caused by them and thus needs no correction. Sober and Brainard [14] propose a somewhat different line of reasoning based on their finding that the amount of compensation is proportional to the overlap between the distributions of produced pitch and perceived pitch. They argue that sensory errors drive learning best when these fall within the range of production variability, explaining why large shifts outside the production range are ineffective in driving adaptation. Thus, the former explanation relates partial compensation to inferred external causes of the shift, and the latter relates it to constraints inherent to vocal learning. These explanations qualitatively differ in that the former argues for optimality of the adaptive behavior (there is no need to adapt to external causes), whereas the latter emphasizes a limitation of the adaptive system (constraint of learning). Our aim is to formalize the Liu and Larson’s proposal into a set of equations that can be used to fit data, similarly to Sober and Brainard’s model.

Computational models accounting for partial motor compensation have been previously proposed. Commonly, sensorimotor adaptation is explained using forward models [4,15]. Briefly, the idea is that the brain possesses an internal model of the system dynamics (e.g. hand position in a reaching task or fundamental frequency in a vocalization task) called the “forward model”. The forward model is able to generate an estimate of future sensory feedback based on a current motor command, hence the use of the term “forward”. If there are unexpected deviations between the estimate and the feedback, then motor commands are corrected dynamically in proportion to the perceived error [16,17]. Recently, [3] and [4] have extended this approach to account for the decreased compensation under large sensory errors by explicitly modeling the gain associated with integrating any particular sensory error as a decreasing function of the error (specifically, this appears through a sub-quadratic loss function).

We are taking a more principled approach to partial compensation in which we estimate the produced motor output (pitch in our case) based on a statistically optimal *linear combination* of expected and actual feedback which is then used (after non-linear weighting) to generate a “sensory error” for subsequent motor adaptation. The derived motor adaptation rule is quite similar to the Kalman filter, except that the gain term for the sensory error is obtained from a more principled perspective than in [4]. To test whether partial compensation is congruent with optimal behavioral strategies, we investigate a simple theoretical framework inspired by Bayesian theories of motor adaptation [8]. We show that published data in both humans and birds can be fit by a simple Bayesian decision model based on the idea that subjects respond optimally to the conflicting feedback, i.e., they adapt pitch up to an extent commensurate with known (or estimated) reliability of sensors and actuators. Our results suggest that both humans and birds behave optimally given knowledge of sensory, motor, and neural imperfections. And, in principle, given sufficient experimental data, the magnitude of these imperfections can be estimated from the data.

Results

We present our work in the nomenclature of birdsong, which is our primary expertise. In a typical experiment, the **impinging pitch** p' at the bird's ears is the sum of the **produced pitch** p and the artificially applied **pitch shift** p_{Δ} :

$$p_{\text{ear}} = p + p_{\Delta},$$

where pitch is reported in logarithmic coordinates of cents, $p = 1200 \log_2(F) - \kappa$, with F the pitch (or fundamental frequency) in *Hz* and κ an arbitrary constant. Note that if birds do not wear pitch-shifting earphones then the impinging pitch p_{ear} at the ears equals the produced pitch p .

It is well known that the produced pitch p of a song syllable is not constant but it fluctuates from one rendition of the syllable to another. We approximate the distribution of produced pitch p across syllable renditions by a Gaussian with **mean (produced pitch)** μ_m and **variance (of produced pitch)** σ_m^2 :

$$P(p) = N(\mu_m, \sigma_m^2),$$

where $P(p)$ is the probability of measuring the produced pitch p using a microphone in such an experiment and where $N(\mu, \sigma^2)$ denotes a Gaussian distribution with mean μ and variance σ^2 . For the motor standard deviation of bengalese finch song syllables we take $\sigma_m = 46$ cents. We obtained this value by multiplying the median absolute deviation (MAD) of pitch distributions reported in [18] by 1.4826, which is the known factor that relates MAD and Gaussian standard deviation, i.e., $\sigma_m = 31 * 1.4826 = 46$ cents.

In the subsequent treatment, μ_m has the role of the bird's current motor plan which together with the (constant) motor variability σ_m^2 is relayed to a sensory area as priors used for pitch estimation (we assume the sensory area has no information about the actual produced pitch p , only about its plan μ_m , the sensory feedback p_f and the typical motor and sensory noise variances σ_m^2 and σ_f^2 , respectively). In other words, we assume their perceptual system receives an internal expectation of pitch, which is not available when another bird sings. Evidence for such forward models (not of pitch per se, but of auditory input in more general terms) in both mammals and songbirds is provided by neurons in auditory brain areas that respond almost exclusively to feedback distortions [19,20]. We thus model pitch self-perception as the result of combined input from two separate sources, one being an internally expected feedback stemming from a forward model of the vocal organ and the other auditory feedback.

Given there must be noise in cochlear and neural processing of pitch, we assume the **auditory pitch feedback** p_f encoded in the brain's auditory processing stream to be a Gaussian distributed random variable with mean p_{ear} and **variance (of pitch feedback)** σ_f^2 :

$$P(p_f|p_{ear}) = N(p_{ear}, \sigma_f^2)$$

where $P(p_f|p_{ear})$ is the probability of observing pitch feedback p_f given impinging pitch at the ears p_{ear} . Unfortunately, the current birdsong literature documents no attempts of measuring the variance of pitch feedback σ_f^2 , which is why σ_f^2 is an unknown parameter in our model that we estimate through a simple fitting procedure described in Table 1. Given the two sources of noise, on average, birds try to reproduce a **pitch target** μ^* they have acquired from a tutor during a sensory song learning phase [21]. When birds are not subjected to pitch-shifted feedback ($p_\Delta = 0$), birds simply achieve a good copy of tutor song when the mean produced pitch equals the target, i.e., when $\mu_m = \mu^*$. However, under pitch-shifted feedback ($p_\Delta \neq 0$), the mean produced pitch μ_m deviates from the target by an amount ϵ :

$$\mu_m = \mu^* + \epsilon.$$

That amount, the **corrective pitch bias** ϵ is our quantity of interest. For a given fixed pitch shift p_Δ applied across many syllable renditions, we compute the equilibrium pitch bias $\langle \epsilon \rangle$ based on optimality criteria, as detailed in the following paragraphs.

One goal of the perceptual system is to decide whether the feedback p_f is within the expected range (within the known motor variance σ_m^2) of the expected value μ_m . The larger the sensory noise (the larger σ_f^2), the more likely will an imposed shift p_Δ be interpreted as being *self-caused*. By contrast, when the feedback p_f differs greatly from the expectation μ_m , then such a difference cannot be reconciled and the perceptual system decides that feedback is contaminated with another bird's vocalization or with some other environmental disturbance.

Thus, the critical problem is to estimate whether the feedback p_f is self-caused or not. The Bayes optimal approach is to compute the posterior probability of the source of a sensory event conditioned on the sensory input. Essentially, we can consider the source (S) of an event to be a Bernoulli random variable taking on one of two values $S \in \{s, e\}$ denoting 'self' and 'external', with prior probabilities that satisfy $P(s) + P(e) = 1$. In this approach $P(s|p_f)$ is the

Table 1. Stochastic algorithm for finding the motor bias ϵ in response to a given pitch shift p_Δ and for identifying optimal model parameters σ_f^2 and k by fitting motor bias (compensation) data.

1.	Pick a pitch target μ^* and choose model parameters σ_f^2 and k characterizing the bird's sensory noise and perceptual priors
2.	Choose a feedback shift p_Δ (in cents), set the corrective bias to zero initially, $\epsilon = 0$
3.	Draw a set of n (e.g., $n = 200$) random pitch samples p according to $N(\mu_m, \sigma_m^2)$ and associated random auditory feedback signals p_f according to $N(p + p_\Delta, \sigma_f^2)$ (assume the bird produces n renditions of the target syllable)
4.	Compute the one-source posteriors $P(s p_n)$ and the pitch deviations Δp (Eq 4) of these n renditions
5.	Change the corrective bias according to $\epsilon \rightarrow \epsilon - 0.001 \langle \Delta p \rangle_n$ and return to Step 3 until convergence.
6.	Compute the percent compensation as $c = 100 \langle \epsilon \rangle_n / p_\Delta$, where the running average $\langle \epsilon \rangle_n$ runs over the last set of renditions
7.	Go to Step 2 and pick a new feedback shift p_Δ .
8.	Evaluate the goodness of fit (compute the average fitting residual as a function of p_Δ , as in Fig 2a and 2b). If the average residual is larger than in the previous fit, reject the latest parameter choices and proceed to Step 1 to pick new parameters σ_f^2 and k (randomly deviating from the current best choice by up to one order of magnitude).

doi:10.1371/journal.pone.0169795.t001

posterior probability that the perceived pitch is ‘self’ generated and $P(e|p_f)$ that it is externally generated.

These posterior probabilities can be used to determine for example whether auditory input during a particular vocalization is self-caused, which is the case if $P(s|p_f) > P(e|p_f)$. For example, a good adaptive strategy could be to retune the produced pitch when the auditory input is self-caused, $P(s|p_f) > 1/2$, and its pitch is off target (not close to μ^*). However, in the following treatment we will use Bayesian inference to retune pitch, based not on a decision about the source but instead on pitch estimates weighted by their posterior source probabilities.

According to Bayes’ theorem, we can write the posterior source probability $P(s|p_f)$ in terms of the likelihood $P(p_f|s)$ of observing the feedback p_f given it is produced by the bird, as follows:

$$\begin{aligned}
 P(s|p_f) &= \frac{P(p_f|s)P(s)}{P(p_f)} = \frac{P(p_f|s)P(s)}{P(p_f|s)P(s) + P(p_f|e)P(e)} \\
 &= \frac{P(p_f|s)}{P(p_f|s) + \frac{P(p_f|e)P(e)}{P(s)}} \\
 &= \frac{P(p_f|s)}{P(p_f|s) + k}
 \end{aligned} \tag{1}$$

where k is a free parameter that depends on two factors: i) the unknown ratio of prior probabilities $P(s)$ and $P(e)$, and ii) the unknown pitch likelihood $P(p_f|e)$ given an external source. The parameter k is motivated by the idea that birds may have an accurate forward model of their own motor system but not of the world in general, which boils down to a uniform model and thus a single scalar parameter.

The likelihood $P(p_f|s)$ of (non-shifted, $p_\Delta = 0$) pitch feedback can easily be computed assuming independence of sensory and motor noise sources:

$$\begin{aligned}
 P(p_f|s) &= \int P(p_f|q, s)P(q) dq \\
 P(p_f|s) &= \frac{1}{\sqrt{2\pi(\sigma_f^2 + \sigma_m^2)}} e^{-\frac{(\mu_f - \mu_m)^2}{2(\sigma_f^2 + \sigma_m^2)}}
 \end{aligned} \tag{2}$$

Where we made use of the definitions $P(p_f|q, s) = \frac{1}{\sqrt{2\pi\sigma_f^2}} e^{-\frac{(p_f - q)^2}{2\sigma_f^2}}$ and $P(q) = \frac{1}{\sqrt{2\pi\sigma_m^2}} e^{-\frac{(\mu_m - q)^2}{2\sigma_m^2}}$ as previously introduced. The integral in Eq 2 arises from the fact that birds do not have access to the true produced pitch p but can only optimally infer its consequences by summing over all possibilities weighted by their probabilities. By inserting Eq 2 into Eq 1 we obtain a closed expression for the posterior probability $P(s|p_f)$.

To estimate the pitch discrepancy perceived by birds, we first consider the special case in which birds inferred the feedback stemming from one source, e.g. $P(s|p_f) = 1$. In this case, the (optimally) **combined pitch** p_{opt} given the noisy measurement p_f and the prior μ_m is given by the 1-step iteration of a Kalman filter:

$$p_{opt} = \operatorname{argmax}_p [P(p_f|p, s)P(p)] = \frac{\sigma_f^2 \mu_m + \sigma_m^2 p_f}{\sigma_f^2 + \sigma_m^2} \tag{3}$$

This formula also corresponds to the optimal combination of two noisy sensory cues, see e.g. [2]. As expected, the combined pitch p_o always lies somewhere in between p_f and μ_m . The inferred pitch discrepancy is $\Delta p = p_{opt} - \mu^*$. Naively, the adaptation strategy could be to apply a

pitch change ϵ such that the new combined pitch p_{opt} coincides with the target μ^* . However, such strategy would be optimal only if birds were certain about the feedback being self-caused, which we assume is not the case. To take into consideration the posterior probabilities of self-versus external sources, we perform Bayesian inference [22] according to which the least-square pitch estimator is the conditional mean pitch, which provides us with the following expression for the perceived pitch deviation of a given syllable rendition:

$$\Delta p = P(s|p_f)p_{opt} + P(e|p_f)\mu_m - \mu^* = \epsilon + (p_f - \mu_m)P(s|p_f) \frac{\sigma_m^2}{\sigma_f^2 + \sigma_m^2} \tag{4}$$

The remaining idea is that birds compute a motor bias ϵ such that on average (across syllable renditions), the pitch deviation vanishes, $\langle \Delta p \rangle = 0$. Unfortunately, because we cannot isolate ϵ in Eq 4 algebraically, it is not possible to provide a closed-form solution for the motor bias ϵ . To find this bias in practice, we iteratively change the corrective pitch bias ϵ (in small steps) with each set n of syllable renditions as follows:

$$\epsilon \rightarrow \epsilon - \eta \langle \Delta p \rangle_n,$$

where η is a small constant, e.g. 0.001. Using this procedure, for a fixed pitch shift p_Δ , the running average of pitch deviation $\langle \Delta p \rangle_n$ converges to zero and the running average $\langle \epsilon \rangle$ converges to the corrective motor response we are interested in. A schematic of the entire model is presented in Fig 1 and the algorithm for finding the motor bias ϵ is described in Table 1.

The model comprising two parameters σ_f^2 and k summarized in Table 1 provides very good fits to Bengalese finch data that we manually digitized from [14], Fig 2.

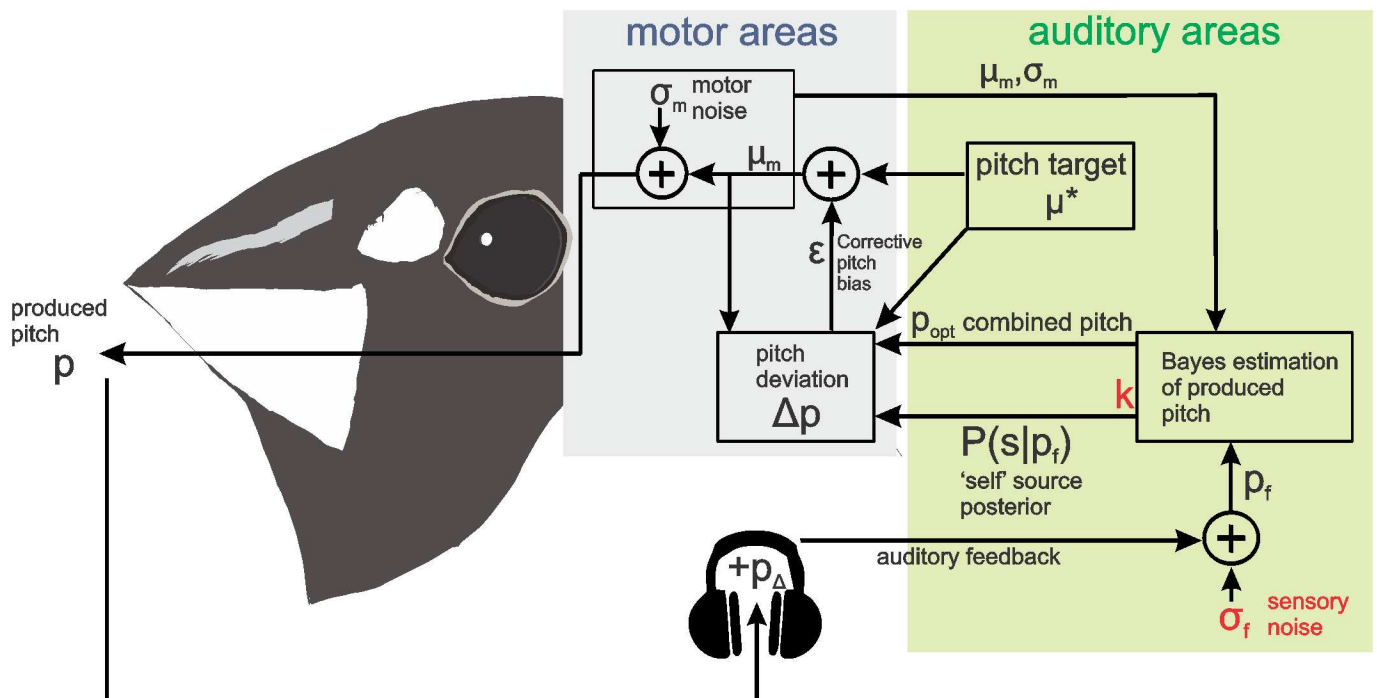


Fig 1. Model of optimal pitch adaptation. Motor areas in the brain generate a motor plan μ_m by integrating a desired pitch μ^* and pitch adaptation ϵ . The produced pitch suffers from motor noise. Auditory areas optimally combine the motor plan with corrupted feedback p_f , then reweight the estimate by the probability of feedback being self-caused $P(s|p_f)$ to produce a final pitch deviation Δp relative to the desired pitch μ^* . The two free parameters highlighted in red are estimated by fitting pitch compensation data from Bengalese finches and humans (Fig 2).

doi:10.1371/journal.pone.0169795.g001

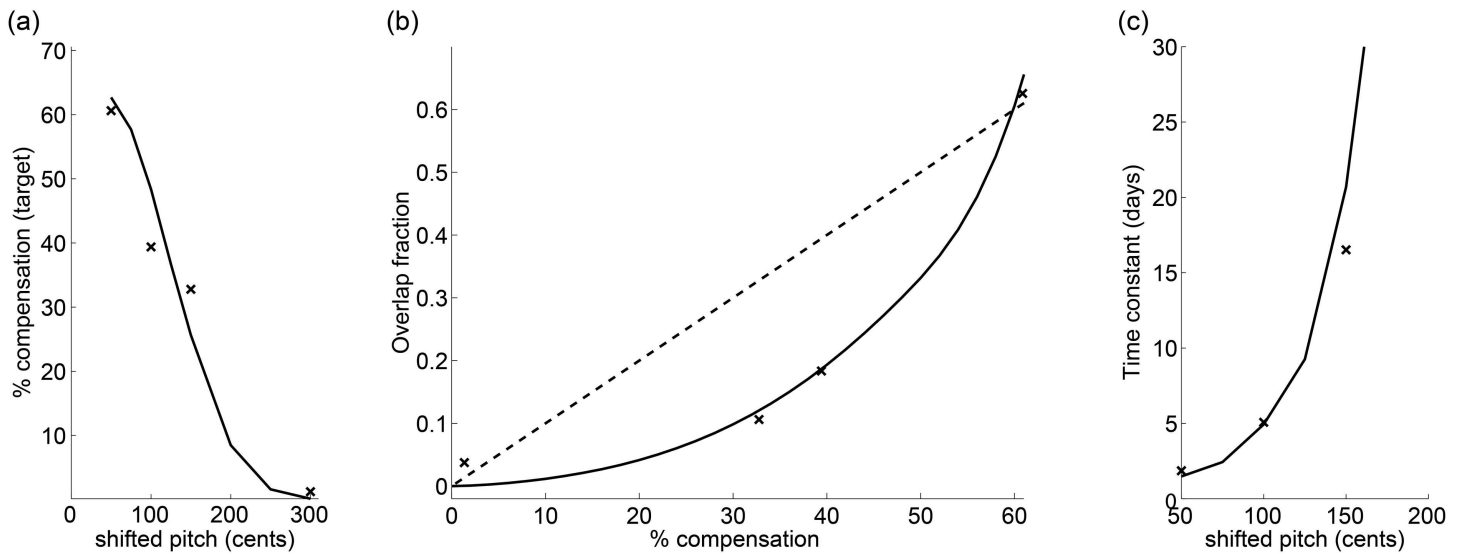


Fig 2. Model fits (black lines) to Bengalese finch data (crosses) digitized from [14]. Best fits to compensation data (a) and to overlap-fraction data (b) are achieved for $\sigma_f = 23$, $k = 1.5 \times 10^{-4}$. For comparison, the dashed line in (b) is the fit to the data provided by the overlap model in [14]. (c) The learning time constant (in days) was estimated as $\tau = q \langle P(e|p_f) \rangle / \langle P(s|p_f) \rangle$, i.e. as the ratio of the self-versus external-source posterior probabilities (learning occurs mainly during inferred self-produced syllable renditions), q is a parameter estimated using a least-squared error fit.

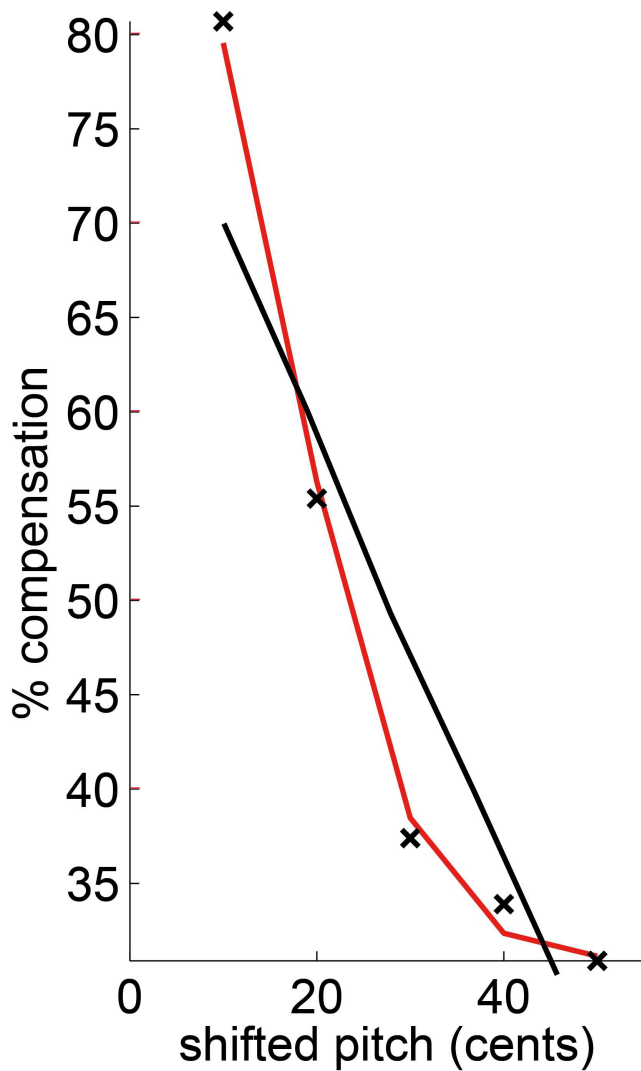
doi:10.1371/journal.pone.0169795.g002

To test whether the model is also able to reproduce pitch adaptation in humans, we digitized the pitch-shift compensation curves in [11] and produced model fits using parameters k , σ_f , and σ_m (motor variability was not reported in [11]). We found that model fits were good but not excellent, Fig 3a. We speculated that the main source of discrepancy between model and data could be the manner in which Liu and Larson quantified the motor response as the transient peak compensation (assessed relative to pitch traces in non-shifted control trials). Because of spontaneous pitch fluctuations, the peak transient compensation must be biased (any non-constant function exhibits local maxima that deviate from the mean), we argued, prompting us to introduce an additional bias parameter ϵ_0 in model fits (ϵ_0 is simply a constant bias added to ϵ). Indeed, including this additional parameter, model fits looked excellent, Fig 3a.

Humans and birds that produce more pitch baseline variability also compensate more to pitch-shifted feedback [18,23], an effect that we probed in our model. To investigate the precise dependence of compensation ϵ on pitch variability σ_m^2 , we digitized the pitch-shift compensation curves as a function of pitch variability reported in [23] and fitted these curves using model parameters k and σ_f . We found rough numerical agreement only in the range in which the parameter k was negligibly small, (Fig 3) implying that the model imposes an unconditional self-source interpretation, $P(s|p_f) \simeq 1$. Perhaps not surprisingly, a self-source interpretation might have been enforced in [23] by telling participants to match a target note despite possible feedback alterations, which implies to reject the external-source interpretation, which in our model means to clamp the self-source posterior probability to one. Indeed, when we hard-coded $P(s|p_f) = 1$ into the model, the equal fit resulted, Fig 3b (red dashed line).

Interestingly, the model fits to [11] and [23], despite both applying to human data, differed in terms of their inferred sensory variability ($\sigma_f = 0$ cents vs $\sigma_f = 7.5$ cents). These numbers, however, were close enough and almost equally good fits resulted when we fitted [11] with $\sigma_f = 7.5$ cents inferred from [23]. Note that by contrast, the converse was not true and $\sigma_f = 0$ cents provided a poor account of the data [23]. Thus, overall we estimate σ_f in the range of

(a) Liu and Larson (2007)



(b) Scheerer and Jones (2012)

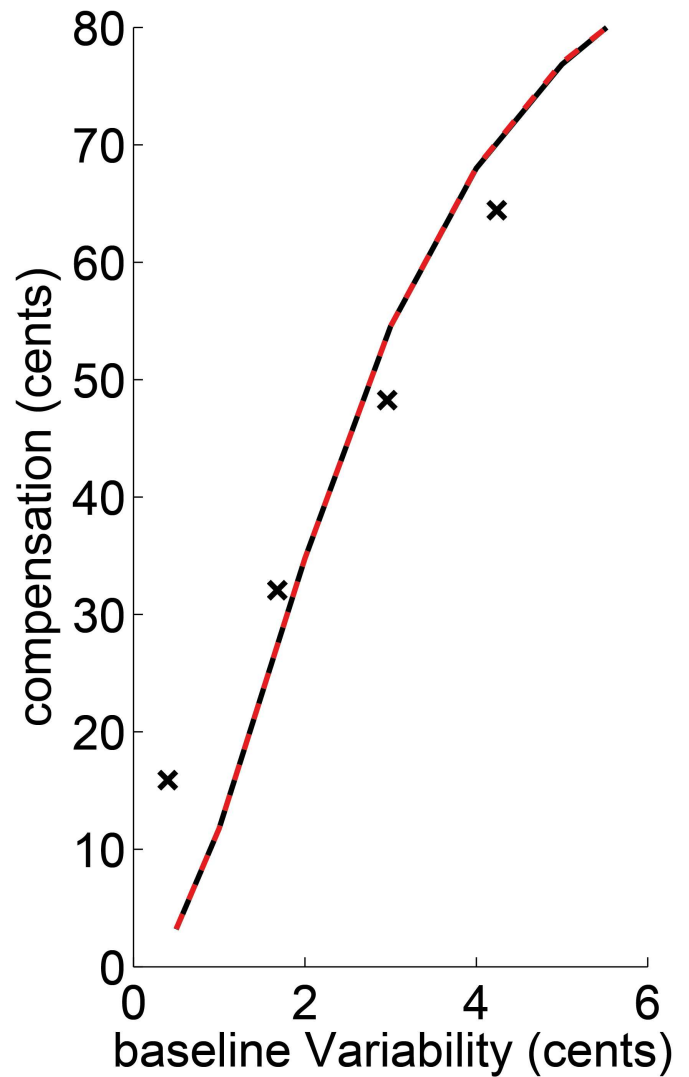


Fig 3. Model fits (lines) to human pitch compensation data (black crosses) digitized from [11]. (a) The model fit (black line) reveals only qualitative agreement but no precise match; $k = 5.2 * 10^{-4}$, $\sigma_f = 0$ cents, $\sigma_m = 32$ cents. After introducing an additional offset parameter ϵ_0 to account for a read-out bias, the model fit (red line) becomes excellent; $k = 1.4 * 10^{-3}$, $\sigma_f = 0$ cents, $\sigma_m = 14$ cents, $\epsilon_0 = 31$ cents. (b) Fits (black line) through data points (crosses) extracted from the linear regression in [23]. $k = 10^{-320}$ (essentially $k = 0$), $\sigma_f = 7.5$ cents. The same fit results (red dashed line) when enforcing a self-source interpretation, $P(s|\rho) = 1$. $\sigma_f = 7.5$ cents.

doi:10.1371/journal.pone.0169795.g003

5–10 cents to provide a decent estimate of human perceptual pitch noise. A nontrivial prediction derived from our model is a non-monotonic dependence of the corrective motor bias as a function of sensory noise σ_f , shown in Fig 4. Intuitively, the decreased motor compensation for large sensory noise σ_f arises from the disregard of noisy sensors by causal inference. By contrast, the decreased compensation for small noise σ_f is a result of trial-to-trial motor variability: Syllable renditions with pitch closer to the expected value μ_m have a larger posterior source probability than renditions with pitch further away from the expected value. Hence, when

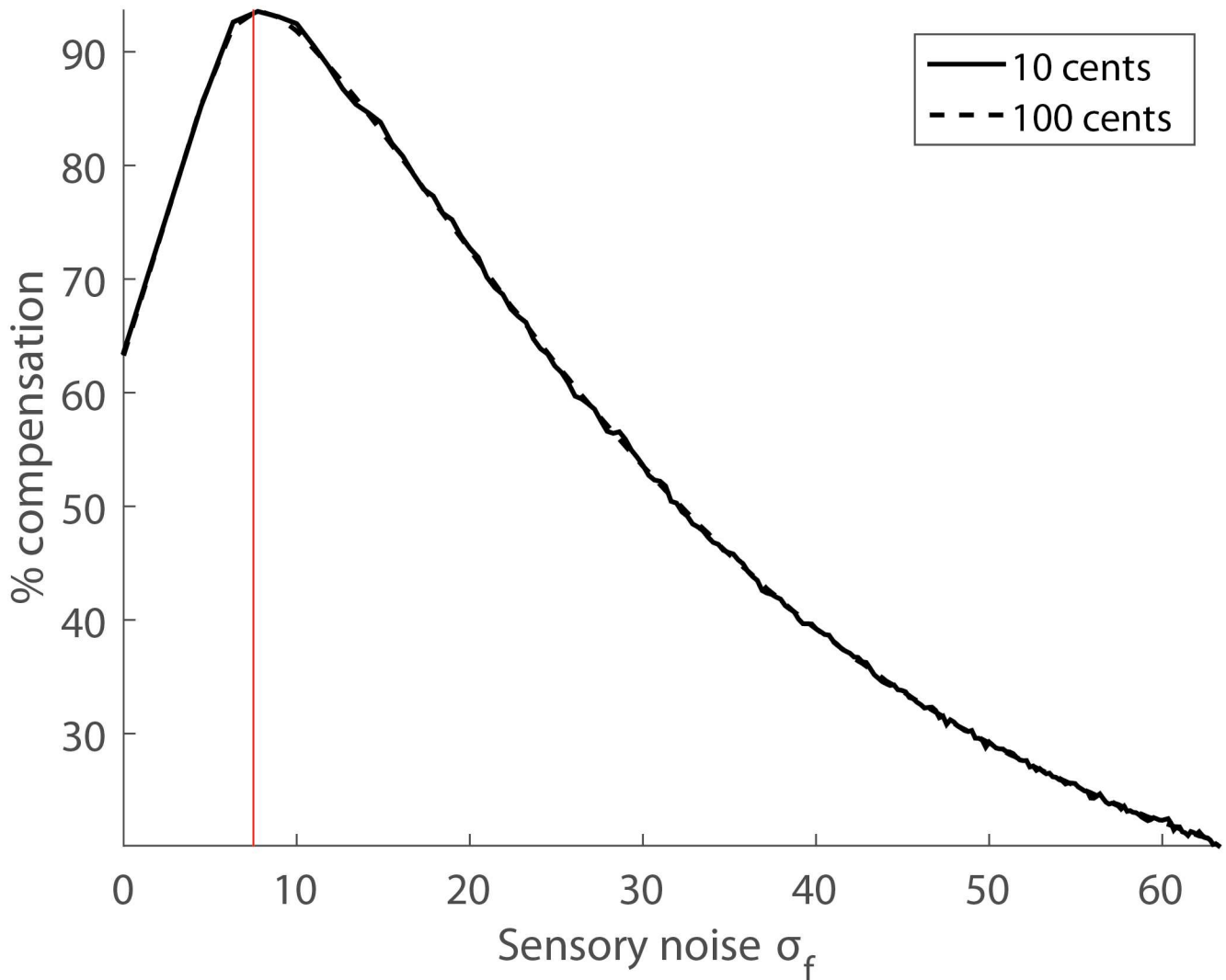


Fig 4. Non-monotonic dependence of percent compensation as a function of sensory noise. For both small and large pitch shifts p_Δ (superimposed full and dashed lines), the percent pitch compensation is a non-monotonic function that peaks at an intermediate level of sensory noise. Model simulations were performed with best-fit parameters for the human data in Fig 3: $\sigma_m = 32$ cents, $k = 0$. The red line marks the upper limit of our inferred pitch variability in humans ($\sigma_f = 7.5$ cents).

doi:10.1371/journal.pone.0169795.g004

averaging over many syllable renditions, deviant renditions are more often ‘rejected’ from being self-caused than close ones; this rejection of outliers by causal inference is very prominent when sensory feedback is reliable.

Discussion

We presented a normative model of vocal adaptation in birds and humans. The model contains only a small set of parameters, yet it reproduces the widely found decrease in motor compensation with increasing distortion of sensory feedback, and it also reproduces the recently demonstrated positive correlation between motor variability and motor compensation. We did not identify any serious disagreement with published data; any possible disagreement would be interesting, as it would introduce a requirement for model revisions.

Computationally, our model can be seen as a generalization of forgetting models of incomplete adaptation [24]: Namely, by replacing the term $P\left(s|p_f\right) \frac{\sigma_m^2}{\sigma_f^2 + \sigma_m^2}$ in Eq (4) by a constant term α , we find that the level of adaptation is set by two terms among which the target pitch μ^* has the role of driving a forgetting term. Our nonlinear model is more general than simple forgetting-retention models in the sense that it correctly explains zero adaptation for highly magnified errors (Fig 2a) and that adaptation time constants are not fixed but depend on posterior evidence.

One of our aims was to include a model component for sensory processing, to support the possibility that motor adaptation might be constrained by the ability to perceive sensory feedback. No sensory component exists in overlap models of adaptive vocal behavior [18]. The model we studied includes sensory noise through the parameter σ_f . The predicted compensation decreases with increasing sensory noise, and the model correctly produces zero adaptation ($\epsilon = 0$ in Eq 4) in deaf birds ($\sigma_f = \infty$).

The pitch perception noise inferred in our model ($\sigma_f = 1$ to 7 cents) agrees well with reports in the literature. Human pitch perception studies report just-noticeable differences of 5–10 Hz for fundamental frequencies of complex harmonics [25] and of tonal speech [26]. If we assume a cumulative Gaussian function to underlie the measured psychometric functions, we obtain an estimated standard deviation σ_f of sensory noise of about 3–7 cents, agreeing well with our fit of the Scheerer and Jones (2012) data.

One way forward of using our modeling approach in further studies is to separately estimate perceptual pitch noise σ_f to further constrain the model down to essentially a single degree of freedom represented by the parameter k . In birds, for example, it would be interesting to probe the pitch discrimination ability using a go/no-go auditory discrimination paradigm [27] and to relate the perceptual precision of birds to their compensation magnitudes.

In principle, the dependence of motor compensation on sensory noise is non-trivial, Fig 4. Our model predicts that the compensation first increases as a function of sensory noise σ_f until it reaches a local maximum after which it decreases because the internal model becomes more reliable than the noisy auditory system. Unfortunately, this predicted non-monotonic relationship will not be simple to test in experiments, mainly for two reasons. First, it is not straightforward to model pitch noise. Second, even if it were possible to model pitch noise, our estimate of sensory variability $\sigma_f = 7.5$ cents in humans (Fig 3) coincides with the peak compensation in Fig 4. Because it is impossible to cancel out noise in the brains' auditory processing streams, the low-noise region of the model seems to be inaccessible to experiments. Nevertheless, provided that our modeling approach applies to adaptation in other sensory systems such as vision, it may be possible to find behavioral paradigms in which highly precise sensors reside on the left of the non-monotonic compensation curve in Fig 4.

Regarding the speed of compensation there seems to be a discrepancy between humans and birds: in humans larger pitch shifts led generally to shorter adaptation latencies [11], which contrasts with birds in which large shifts are compensated more slowly than small shifts [14]. In humans, the speed of learning depends on many factors including environmental consistency [28]. Overall, given these discrepancies there is currently little hope to identify a unified principle underlying the speed of adaptation. For this reason we have excluded the time constant of compensation (Fig 2c) in the fitting procedure used to identify model parameters σ_f^2 and k .

It may be possible to further specify the role of model parameters from diverse reports in the literature. These are that singers compensate more than speakers [29] but good singers compensate less than untrained singers [30]. In the context of our model, singers compensating less than untrained singers might be explainable through σ_m being smaller in singers. By

contrast, the same explanation via σ_m is unlikely to explain differences between singing and speaking. Rather, we would predict that the model parameter k is smaller during singing than during speaking, $k_{sing} < k_{speak}$, possibly arising from increased attention to auditory feedback during singing (increase in $P(s)$). In summary, as suggested by our findings in Fig 3, we believe the model parameter k can account for instruction effects [31] according to which the compensation magnitude depends on the precise instructions given.

With regards to the parameter k , we have here assumed it to be a fixed constant combining self and external source priors $P(s)$ and $P(e)$, and the uniform distribution of other-produced pitch, $P(p_f|e)$. Had the fits presented in Fig 2 not looked convincing, we would have assumed that birds estimate the source priors as well as the pitch likelihoods and posteriors, for example via equalities $P(s) = \int P(s|p_f)dp_f$ and $P(e) = 1 - P(s)$. When we played around with this possibility and iteratively estimated $P(s)$ and $P(e)$ in this way while estimating the fixed likelihood $P(p_f|e)$ as a free model parameter or even as a Gaussian function, we found that compensation curves (Fig 2a) looked much steeper than in reality, providing little preliminary support for such more extended models.

Motor adaptation is sensitive to the characteristics of the input modality. In humans, high fundamental frequencies of vocalizations are associated with lesser pitch compensation [11]. Whether these observations can be related deterministically to frequency-dependence of motor variability and sensory noise remains to be seen. Our approach could be used to model compensatory responses also with respect to other modalities such as sound amplitude. For example, [32] reported that for loudness-altered voice feedback, a compensatory gain close to 1 was only reported when subjects attempted to maintain a relatively quiet voice, but not when they produced louder vocalizations. In general, the Bayesian framework is powerful enough to be extended through the same normative principles to additional behavioral and sensory modalities.

Acknowledgments

We acknowledge Alex Pouget, Homare Yamahachi, and Ryosuke Tachibana for helpful discussions and feedback on the manuscript and the support by the Swiss National Science Foundation, Project 31003A_156976/1.

Author Contributions

Conceptualization: RH.

Data curation: RH.

Formal analysis: RH GN.

Funding acquisition: RH.

Investigation: RH.

Methodology: RH GN.

Project administration: RH.

Resources: RH.

Software: RH GN.

Supervision: RH.

Validation: RH GN.

Visualization: RH GN.

Writing – original draft: RH.

Writing – review & editing: RH GN.

References

1. Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. Causal inference in multisensory perception. *PLoS One*. 2007; 2.
2. Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*. 2002; 415: 429–33. doi: [10.1038/415429a](https://doi.org/10.1038/415429a) PMID: [11807554](https://pubmed.ncbi.nlm.nih.gov/11807554/)
3. Körding KP, Wolpert DM. Bayesian integration in sensorimotor learning. *Nature*. 2004; 427: 244–247. doi: [10.1038/nature02169](https://doi.org/10.1038/nature02169) PMID: [14724638](https://pubmed.ncbi.nlm.nih.gov/14724638/)
4. Marko MK, Haith AM, Harran MD, Shadmehr R. Sensitivity to prediction error in reach adaptation. *J Neurophysiol*. 2012; 108: 1752–1763. doi: [10.1152/jn.00177.2012](https://doi.org/10.1152/jn.00177.2012) PMID: [22773782](https://pubmed.ncbi.nlm.nih.gov/22773782/)
5. Shadmehr R, Mussa-Ivaldi FA. Adaptive representation of dynamics during learning of a motor task. *J Neurosci*. 1994; 14: 3208–3224. PMID: [8182467](https://pubmed.ncbi.nlm.nih.gov/8182467/)
6. Fine MS, Thoroughman K. Motor adaptation to single force pulses: sensitive to direction but insensitive to within-movement pulse placement and magnitude. *J Neurophysiol*. 2006; 96: 710–720. doi: [10.1152/jn.00215.2006](https://doi.org/10.1152/jn.00215.2006) PMID: [16707722](https://pubmed.ncbi.nlm.nih.gov/16707722/)
7. Gutfreund Y, Zheng W, Knudsen EI. Gated Visual Input to the Central Auditory System. *Science (80-)*. 2002; 297: 1556–1559.
8. Wei K, Körding K. Relevance of error: what drives motor adaptation? *J Neurophysiol*. 2009; 101: 655–64. doi: [10.1152/jn.90545.2008](https://doi.org/10.1152/jn.90545.2008) PMID: [19019979](https://pubmed.ncbi.nlm.nih.gov/19019979/)
9. Houde JF, Jordan MI. Sensorimotor adaptation in speech production. *Science*. 1998; 279: 1213–1216. PMID: [9469813](https://pubmed.ncbi.nlm.nih.gov/9469813/)
10. Jones JA, Munhall KG. Perceptual calibration of F0 production: evidence from feedback perturbation. *J Acoust Soc Am*. 2000; 108: 1246–51. PMID: [11008824](https://pubmed.ncbi.nlm.nih.gov/11008824/)
11. Liu H, Larson CR. Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex. *J Acoust Soc Am*. 2007; 122: 3671–7. doi: [10.1121/1.2800254](https://doi.org/10.1121/1.2800254) PMID: [18247774](https://pubmed.ncbi.nlm.nih.gov/18247774/)
12. Burnett TA, Freedland MB, Larson CR, Hain TC. Voice F0 responses to manipulations in pitch feedback. *J Acoust Soc Am*. 1998; 103: 3153–3161. PMID: [9637026](https://pubmed.ncbi.nlm.nih.gov/9637026/)
13. Bauer JJ, Larson CR. Audio-vocal responses to repetitive pitch-shift stimulation during a sustained vocalization: improvements in methodology for the pitch-shifting technique. *J Acoust Soc Am*. 2003; 114: 1048–1054. doi: [10.1121/1.1592161](https://doi.org/10.1121/1.1592161) PMID: [12942983](https://pubmed.ncbi.nlm.nih.gov/12942983/)
14. Sober SJ, Brainard MS. Vocal learning is constrained by the statistics of sensorimotor experience. *Proc Natl Acad Sci*. 2012; 1–5.
15. Wolpert DM, Ghahramani Z. An internal model for sensorimotor integration. *Science (80-)*. 1995; 269: 1880–1882.
16. van Beers RJ. How Does Our Motor System Determine Its Learning Rate? *PLoS One*. 2012; 7.
17. van Beers RJ. Motor Learning Is Optimally Tuned to the Properties of Motor Noise. *Neuron*. Elsevier Ltd; 2009; 63: 406–417. doi: [10.1016/j.neuron.2009.06.025](https://doi.org/10.1016/j.neuron.2009.06.025) PMID: [19679079](https://pubmed.ncbi.nlm.nih.gov/19679079/)
18. Kelly CW, Sober SJ. A simple computational principle predicts vocal adaptation dynamics across age and error size. *Front Integr Neurosci*. 2014; 8: 75. doi: [10.3389/fnint.2014.00075](https://doi.org/10.3389/fnint.2014.00075) PMID: [25324740](https://pubmed.ncbi.nlm.nih.gov/25324740/)
19. Eliades SJ, Wang X. Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*. Nature Publishing Group; 2008; 453: 1102–1106. Available: <http://www.ncbi.nlm.nih.gov/pubmed/18454135>
20. Keller GB, Hahnloser RHR. Neural processing of auditory feedback during vocal practice in a songbird. *Nature*. Nature Publishing Group; 2009; 457: 187–90.
21. Marler P, Tamura M. Culturally transmitted patterns of vocal behavior in sparrows. *Science (80-)*. 1964; 146: 1483–1486.
22. Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. Causal inference in multisensory perception. *PLoS One*. 2007; 2: e943. doi: [10.1371/journal.pone.0000943](https://doi.org/10.1371/journal.pone.0000943) PMID: [17895984](https://pubmed.ncbi.nlm.nih.gov/17895984/)
23. Scheerer NE, Jones JA. The relationship between vocal accuracy and variability to the level of compensation to altered auditory feedback. *Neurosci Lett*. Elsevier Ireland Ltd; 2012; 529: 128–132.

24. Van Der Kooij K, Brenner E, Van Beers RJ, Smeets JBJ. Visuomotor adaptation: How forgetting keeps us conservative. *PLoS One*. 2015; 10.
25. Bernstein JGW, Oxenham AJ, Ives DT, Patterson RD, Kaernbach C, Bering C, et al. The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss. *J Acoust Soc Am*. 2006; 120: 3929. PMID: [17225420](#)
26. Liu C. Just noticeable difference of tone pitch contour change for English- and Chinese-native listeners. *J Acoust Soc Am*. 2013; 134: 3011–20. doi: [10.1121/1.4820887](#) PMID: [24116436](#)
27. Tokarev K, Tchernichovski O. A novel paradigm for auditory discrimination training with social reinforcement in songbirds [Internet]. *bioRxiv*. Cold Spring Harbor Labs Journals; 2014 Apr. Available: [http://
biorxiv.org/content/early/2014/04/12/004176.abstract](http://biorxiv.org/content/early/2014/04/12/004176.abstract)
28. Gonzalez Castro LN, Hadjiosif AM, Hemphill MA, Smith MA. Environmental consistency determines the rate of motor adaptation. *Curr Biol*. Elsevier Ltd; 2014; 24: 1050–61.
29. Natke U, Donath TM, Kalveram KT. Control of voice fundamental frequency in speaking versus singing. *J Acoust Soc Am*. 2003; 113: 1587–1593. PMID: [12656393](#)
30. Jones JA, Keough D. Auditory-motor mapping for pitch control in singers and nonsingers. *Exp Brain Res*. 2008; 190: 279–287. doi: [10.1007/s00221-008-1473-y](#) PMID: [18592224](#)
31. Hain TC, Burnett TA, Kiran S, Larson CR, Singh S, Kenney MK. Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Exp Brain Res*. 2000; 130: 133–141. PMID: [10672466](#)
32. Bauer JJ, Mittal J, Larson CR, Hain TC. Vocal responses to unanticipated perturbations in voice loudness feedback: an automatic mechanism for stabilizing voice amplitude. *J Acoust Soc Am*. 2006; 119: 2363–71. doi: [10.1121/1.2173513](#) PMID: [16642849](#)