Vocal Generalization Depends on Gesture Identity and Sequence

Lukas A. Hoffmann¹ and Samuel J. Sober²

¹Neuroscience Doctoral Program and ²Department of Biology, Emory University, Atlanta, Georgia 30322

Generalization, the brain's ability to transfer motor learning from one context to another, occurs in a wide range of complex behaviors. However, the rules of generalization in vocal behavior are poorly understood, and it is unknown how vocal learning generalizes across an animal's entire repertoire of natural vocalizations and sequences. Here, we asked whether generalization occurs in a nonhuman vocal learner and quantified its properties. We hypothesized that adaptive error correction of a vocal gesture produced in one sequence would generalize to the same gesture produced in other sequences. To test our hypothesis, we manipulated the fundamental frequency (pitch) of auditory feedback in Bengalese finches (*Lonchura striata* var. *domestica*) to create sensory errors during vocal gestures (song syllables) produced in particular sequences. As hypothesized, error-corrective learning on pitch-shifted vocal gestures generalized to the same gestures produced in other sequential contexts. Surprisingly, generalization magnitude depended strongly on sequential distance from the pitch-shifted syllables, with greater adaptation for gestures produced near to the pitch-shifted syllable. A further unexpected result was that nonshifted syllables changed their pitch in the direction opposite from the shifted syllables. This apparently antiadaptive pattern of generalization could not be explained by correlations between generalization and the acoustic similarity to the pitch-shifted syllable. These findings therefore suggest that generalization depends on the type of vocal gesture and its sequential context relative to other gestures and may reflect an advantageous strategy for vocal learning and maintenance.

Key words: Bengalese finch; generalization; motor error correction; motor learning; vocal error correction; vocal learning

Introduction

Generalization, the ability to transfer motor adaptation to a new context, is crucial for learning and maintaining complex behaviors. Generalization is especially important in vocal behavior, during which vocal muscles must be precisely activated to reach time-varying acoustic targets. Because complex vocal behaviors involve producing the same vocal gesture within many different sequences, generalization would allow adaptive modifications of a gesture to transfer to the same gesture produced in other sequences, improving performance.

What are the properties of vocal generalization? In humans, learned changes to one vowel can transfer to the same vowel produced in other words and to other vowels (Houde and Jordan, 1998; Villacorta et al., 2007; Cai et al., 2010; Rochet-Capellan et al., 2012). Furthermore, generalization lessens with increasing acoustic distance between training and transfer utterances (Cai et al., 2010; Rochet-Capellan et al., 2012). Such results parallel findings in limb movement studies, where generalization depends on the similarity between training and transfer movements (Shad-

mehr and Mussa-Ivaldi, 1994; Krakauer et al., 2000). However, other speech studies have suggested that learning is instance-specific with no generalization (Pile and Dajani, 2007; Tremblay et al., 2008; Rochet-Capellan and Ostry, 2011). Thus, the rules of vocal generalization are not well understood. One reason for this could be that speech studies typically use a small set of training and transfer utterances. It is therefore also unclear how generalization occurs across the full natural range of vocalizations and sequences.

Songbirds have provided insight into the neural basis of vocal behavior. Speech and birdsong share numerous parallels (Doupe and Kuhl, 1999; Bolhuis et al., 2010), which include using auditory feedback to correct vocal errors (Houde and Jordan, 1998; Jones and Munhall, 2000; Sober and Brainard, 2009). Physiological studies in songbirds have proposed models of how neural circuits shape the sequencing and acoustic structure of vocal gestures (Hahnloser et al., 2002; Kao et al., 2005; Leonardo and Fee, 2005; Sober et al., 2008; Wohlgemuth et al., 2010). However, although such models can suggest how adaptive vocal changes might be implemented in neural circuits, our poor understanding of the behavioral structure of vocal learning leaves such models badly underconstrained. In this study, we therefore investigated a fundamental question about the computations underlying vocal error correction by asking whether generalization occurs in songbirds or instead might be a unique feature of human speech.

We used Bengalese finches to ask whether generalization occurs in a nonhuman vocal learner and to quantify the properties

Received Dec. 9, 2013; revised Feb. 25, 2014; accepted March 12, 2014.

Author contributions: L.A.H. and S.J.S. designed research; L.A.H. and S.J.S. performed research; L.A.H. and S.J.S. analyzed data; L.A.H. and S.J.S. wrote the paper.

This work was supported by National Institutes of Health Grants P30NS069250 and R01NS084844. We thank Diala Chehayeb and Harshila Ballal for animal care and Reid Schwartz for technical assistance.

The authors declare no competing financial interests.

Correspondence should be addressed to Lukas A. Hoffmann, Predoctoral Student, Neuroscience Doctoral Program, Emory University, Room 2158, 1510 Clifton Road NE, Atlanta, GA 30322. E-mail: lahoffm@emory.edu.

DOI:10.1523/JNEUROSCI.5169-13.2014

Copyright © 2014 the authors 0270-6474/14/345564-11\$15.00/0

Figure 1. Technique for manipulating auditory feedback during individual vocal gestures. **a**, Experimental apparatus. Song is collected by a microphone, pitch-shifted by an online sound processor, and immediately relayed to head-mounted speakers. **b**, Example of pitch-shift during a five-syllable motif. The targeted syllable (B₂, red) was artificially pitch-shifted by ~ 100 cents, whereas nearby same-type (B₁ and B₃, green) and different-type (A and C, blue) syllables were unaltered. Left spectrogram, the bird's vocal output ("sung"). Right spectrogram, the auditory feedback signal played through the headphones ("heard"). Inset at right, sung and heard pitches are the same for B₁ and B₃ but not B₂. **c**, Top, spectrogram of a stereotyped motif from a different bird. Bottom, syllables are color-coded by their assigned category. Only the targeted syllable (C₂, red) was pitch-shifted. All syllables in the motif were assigned a sequential distance from the target syllable.

of generalization in a natural vocal repertoire. We hypothesized that, when feedback errors are experienced for a song syllable appearing in a particular sequential context, error correction on the perturbed syllable would generalize to the same syllable produced in other contexts. Our hypothesis was based on similarities between other forms of vocal learning in humans and songbirds (Houde and Jordan, 1998; Jones and Munhall, 2000; Sober and Brainard, 2009) as well as data showing that neural activity is strongly associated with song syllable identity across contexts (Yu and Margoliash, 1996; Leonardo and Fee, 2005; Wohlgemuth et al., 2010). We fitted songbirds with miniaturized headphones (Hoffmann et al., 2012) and shifted the pitch of auditory feedback from single syllables in particular sequential contexts (Fig. 1). We then quantified how birds adapted to the shifts and how this error correction generalized across other vocal gestures and contexts.

Materials and Methods

Six adult (>135 -d-old) male Bengalese finches (*Lonchura striata* var. *domestica*) were used. Throughout the experiment, birds were isolated in sound-attenuating chambers and maintained on a 14 h:10 h light/dark cycle, with lights on from 7 A.M. to 9 P.M. All recordings analyzed here are from undirected song (i.e., no other bird was present). All procedures were approved by the Emory University Institutional Animal Care and Use Committee.

Experimental procedure. Online, real-time manipulations of auditory feedback were used to induce adaptive changes in song pitch. As described previously (Hoffmann et al., 2012), custom-built headphones were attached to each bird's head. Sound-processing hardware shifted the pitch of acoustic signals and immediately relayed them to the headphone speakers. There was a mean delay of 8.5 ms from the time the bird sang a syllable to the time that syllable was played through the speakers. Vocal pitch in Bengalese finches remains stable when headphones are

used to deliver auditory feedback that has not been pitch-shifted (Sober and Brainard, 2009), so changes in vocal pitch reflect responses to pitch manipulation.

The experiment began with a baseline period (5–7 d) of singing with headphone speakers relaying the songs online at zero pitch shift. After this, birds were exposed to 14 d (2 birds) or 20 d (4 birds) of altered auditory feedback ("shift days"). All analyses of the shift epoch are restricted to the 14 d in which all 6 birds were exposed to shifts; however, similar patterns were seen in days 15–20 in the birds undergoing longer pitch shifts (data not shown).

During the shift days, all songs were relayed online through the headphone speakers. Target syllables within a stereotyped motif were pitch-shifted upward or downward by 100 cents, whereas all other syllables remained at zero pitch shift (Fig. 1). A ± 100 cents pitch shift has been shown to robustly drive learning when applied to all song syllables (Sober and Brainard, 2009, 2012). Custom-written LabView software (Turner and Brainard, 2007) was used to detect a syllable early in the motif. After syllable detection, a square wave was sent to an analog audio switch (Intersil, ISL54405), which switched from a zero-shift to shift channel for the desired time interval. As a result, targeted syllables were played through headphone speakers at 100 cents (3 targeted syllables) or ~ 100 cents (5 targeted syllables). Averaged across all syllables in our analysis, the mean hit-rate (percentage of all targeted syllables correctly pitch-shifted) was 92.1% and in no case was $<90\%$ and the false-positive rate was 0.9%. Three syllables (of a total of 85) were excluded from analysis because false-positive pitch shifts occurred on $>20\%$ of those syllable iterations, making them neither target nor nontarget syllables.

Two birds had two targeted syllables. For both birds, these syllables were sung consecutively within the motif with a very small temporal gap (for example, ABCD). During each iteration, the targeting algorithm detected the preceding syllable (A) at a slightly different time. This caused the pitch-shift to frequently overlap the second syllable. Therefore, it was decided to lengthen the shift time period to shift both syllables. Further-

more, Bengalese finch song consists of several motifs, which are frequently sung in different order. Some motifs are nonstereotyped; in these motifs, syllables are added, omitted, or swapped. However, all targeted syllables occurred in a stereotyped motif where syllable order was always the same.

Syllable categories and sequential distance. Song syllables were divided into three groups (Fig. 1c), designated “targeted” ($n = 8$), “same-type” ($n = 19$), and “different-type” ($n = 55$). Targeted syllables were artificially pitch-shifted, as described above. Some of the nontargeted syllables were visually indistinguishable from targeted syllables when viewed in a spectrogram (Fig. 1, green syllables) but occurred in different sequential positions within the motif. These were considered “same-type” syllables. The rest of the pitch-quantifiable syllables were assigned to the “different-type” group (Fig. 1, blue syllables). Same-type and different-type syllables were not pitch-shifted, except during rare accidental false-positive syllable detections.

Syllable labels were assigned using their acoustic structure as well as their sequential context. For example, Figure 1c shows a spectrogram of a stereotyped motif above the labels for each syllable. In this example, syllables are labeled A–G based on their spectral structure. Additionally, syllables are given numerical subscripts (A_1 , A_2 , etc.) to identify when a syllable is produced at different sequential positions within a motif. In the example shown, only syllable C_2 is pitch-shifted (Fig. 1c, red box). Syllables C_1 and C_3 are therefore same-type syllables, and all other syllables are categorized as different-type, as described above. Figure 1c (bottom) also shows the sequential distance of each syllable from the shifted syllable. Syllables produced at different sequential distances from the target syllable were analyzed separately to quantify how changes to a shifted syllable generalized to nearby syllables as a function of sequential distance. Bengalese finches tend to have relatively short stereotyped motifs, so there were more data points for smaller sequential distances than for larger ones. Syllables produced outside of stereotyped motifs had variable sequential distances to the nearest target syllable and were not included in the sequential distance analysis, although they were included in all other analyses.

Pitch quantification. Song pitch changes were quantified by measuring pitch at specific times within individual syllables as previously described (Sober and Brainard, 2009). Although individual song syllables can be made up of multiple vocal gestures in some songbird species, the majority of Bengalese finch syllables contain only a single gesture with a reliably quantifiable pitch. Therefore, although measurements of “syllable pitch” refer to the pitch of individual vocal gestures, our analysis includes every gesture with quantifiable pitch. Each bird produced 5–8 (median 7) syllable types with quantifiable pitch. Pitch was quantified for all songs sung from 10:00 A.M. to 12:00 P.M. each day. Birds sang a median of 945 (range, 0–4328) pitch-quantifiable syllable iterations in each 2 h window. Whenever a bird did not sing within this window (3.5% of all experimental days), the bird did not contribute to pitch data for that day.

Each individual syllable iteration’s measured pitch (in Hz) was converted to the fractional change from that syllable’s baseline pitch (in cents) as follows:

$$C = 1200 \log_2 \frac{H}{B}$$

where C is the syllable’s pitch change from baseline (in cents), H is the syllable’s measured pitch (in Hz), and B is the mean pitch (in Hz) of all iterations of that syllable over the last three baseline days. A shift of 100 cents corresponds to one semitone, which is a $\sim 6\%$ change in absolute frequency.

The “mean pitch change for one syllable” is the average of individual syllable iterations over the specified time period (one day or multiple days). The mean pitch change across multiple syllables was calculated in two ways (which yielded similar results):

$$M_{syls} = \frac{M_{S1} + M_{S2} + \dots + M_{SJ}}{J}$$

where M_{syls} = mean pitch change across multiple syllable types, M_S = mean pitch change of syllable S , J = total number of syllables used, or:

$$M_{iterations} = \frac{\sum_j \sum_d \sum_{i=1}^{N_{d,j}} S_{i,d,j}}{\sum_j \sum_d N_{d,j}}$$

where $M_{iterations}$ = mean pitch change across multiple syllables’ iterations, j = syllables used, d = days used, $N_{d,j}$ = total number of iterations of syllable S_j on day d within the 2 h window, $S_{i,d,j}$ = the pitch change relative to baseline for individual syllable S_j iteration i on day d . A frequently-sung syllable will contribute proportionally more to $M_{iterations}$, where each syllable iteration is one data point, than M_{syls} , where each type of syllable is one data point.

Pitch contrast. To determine whether pitch changes in different-type syllables acted to restore preexisting pitch relationships between syllables, three “pitch contrasts” were calculated for each target-syllable/different-type syllable pair. We define this pitch contrast at three different times during the experiment. First, for each pair of targeted and different-type syllables within each experiment, we define the “baseline” pitch contrast $C_{Baseline}$ as the difference (in cents) between the mean baseline pitch of a target syllable and a different-type syllable:

$$C_{Baseline} = 1200 \log_2 \frac{T_{Baseline}}{D_{Baseline}}$$

Where $T_{Baseline}$ and $D_{Baseline}$ are the mean pitches of the targeted and different-type syllables, respectively, before shift onset. We then define the “shift start” contrast as the pitch difference between those two syllables when the shift on the target syllable first began (as heard by the bird, i.e., with the target syllable pitch-shifted ± 100 cents relative to baseline):

$$C_{Shift\ start} = 1200 \log_2 \frac{T_{Shift\ start}}{D_{Shift\ start}}$$

Finally, we define the “shift end” contrast as the difference during days 12–14 (as heard by the bird, i.e., with the target syllable pitch-shifted ± 100 cents relative to the sung pitch on those days):

$$C_{Shift\ end} = 1200 \log_2 \frac{T_{Shift\ end}}{D_{Shift\ end}}$$

Thus, $C_{Baseline}$ quantifies the preexisting pitch relationship between two syllables, $C_{Shift\ start}$ quantifies the suddenly changed pitch relationship when pitch-shifted auditory feedback began, and $C_{Shift\ end}$ quantifies the pitch relationship after the bird has responded to the perturbation by altering syllable pitches.

These contrasts were used to calculate percentage “restoration of pitch contrast” (R_C) for each syllable pair:

$$R_C = \frac{(C_{Shift\ end} - C_{Shift\ start})}{(C_{Baseline} - C_{Shift\ start})} * 100\%$$

Therefore, if by the end of the shift period the bird has restored the pitch contrast between syllables to its baseline value (i.e., if $C_{Shift\ end} = C_{Baseline}$), then R_C will equal 100%. On the other hand, if the pitch contrast at the end of the shift epoch is unchanged from the beginning of the shift epoch ($C_{Shift\ end} = C_{Shift\ start}$), then restoration will be 0%. Importantly, the restoration of pitch contrast can be achieved by changing the pitch of the targeted syllable (T_x), the pitch of different-type syllables (D_x), or both.

To determine whether different-type syllable pitch changes significantly contributed to restoring pitch contrast, for each pair of targeted and different-type syllables, percentage restoration was calculated once when pitch changes in different-type syllable were included and once when they were excluded. We excluded the effects of different-type syllable changes by computing

$$C'_{Shift\ end} = 1200 \log_2 \frac{T_{Shift\ end}}{D_{Shift\ start}}$$

which quantifies the pitch contrast at the end of the shift epoch but does not include the effects of changes in different-type syllable (because it uses $D_{Shift\ start}$ rather than $D_{Shift\ end}$). We then computed R'_C , a measure of

pitch restoration that excludes the effect of changes in different-type syllables

$$R'_C = \frac{(C'_{\text{Shift end}} - C'_{\text{Shift start}})}{(C_{\text{Baseline}} - C'_{\text{Shift start}})} * 100\%$$

If different-type syllable pitch changes contribute to pitch contrast restoration, then R_C will be greater than R'_C . A pairwise comparison of these two quantities was performed on the set of target-syllable/different-type-syllable pairs across birds. Finally, we quantified the fraction of total restoration provided by different-type syllables as $\frac{R_C - R'_C}{R_C}$.

Acoustic distance to target syllable. To determine whether there was a relationship between adaptive changes in a syllable and its similarity to the target syllable(s), an acoustic distance metric was calculated for each syllable. The algorithm is similar to a previously described metric (Wohlgemuth et al., 2010). First, three acoustic features were calculated for each syllable iteration that was sung during baseline: fundamental frequency (pitch), amplitude, and spectral entropy (Wohlgemuth et al., 2010). These three features were chosen because they capture a large fraction of the total trial-by-trial acoustic variation and are correlated with neural activity in the vocal motor system (Sober et al., 2008). Each individual feature was then transformed to a z-score using the global mean and SD of that feature across all syllables within each bird. Next, the 3D center of mass (COM) was calculated for each syllable: the mean z-score for pitch, amplitude, and spectral entropy across syllable iterations. Finally, the acoustic distance was defined as the Euclidean distance between a syllable's COM to the bird's target syllable's COM. For the two birds that had two target syllables, the acoustic distance is the mean distance to each target syllable. Additionally, we calculated COM distances using only one acoustic feature (pitch, amplitude, or entropy) at a time using the same procedure. In that case, the distance between two syllables was obtained by subtracting the mean z-score of those syllables' pitches (or amplitudes, or entropies).

Mean spectrograms. Mean spectrograms were computed by aligning individual syllable spectrograms and taking the natural log of the mean spectral power. The spectrograms are used for display only, and all reported analyses use acoustic data from single trials, not mean spectrograms.

Results

We investigated how a sensory error on one syllable in a sequence caused adaptive changes on that syllable (targeted syllable) and examined generalization by quantifying the concurrent pitch changes of other syllables in the song (same-type and different-type syllables). A representative experiment in which the pitch of the targeted syllable was shifted by -100 cents is shown in Figure 2.

Adaptive pitch changes in the same syllable produced in other sequential contexts

Spectrograms in Figure 2*a* show the targeted syllable (A_2) and surrounding syllables on the last baseline day (left) and the same sequence on the last day of a 100 cent downwards pitch shift (right). Figure 2*b* (magenta and blue symbols) shows the pitch of the shifted auditory feedback on the first and last shift days, respectively. In response to the downward shift, the pitch of the target syllable changed in the adaptive direction (increased) between the baseline period (Fig. 2*b*, white symbol) and the end of the shift period (Fig. 2*b*, black symbol). The spectrograms also show surrounding same-type syllables (A_1 , A_3 , A_4 , A_5) in the motif (i.e., the same vocal gesture as the target syllable but produced in other sequential contexts). The bird changed the pitch of syllables produced after the targeted syllable, although they were not artificially pitch-shifted (A_3 , A_4 , A_5 ; white and black symbols, Fig. 2*b*). By the end of the shift period, the mean target syllable pitch had increased 30.9 cents above baseline, indicating

$\sim 30\%$ compensation in the adaptive direction for the -100 cents shift (Fig. 2*c*, rightmost red symbol). The bird also changed its same-type syllable pitches in the adaptive direction over shift days 12–14 (Fig. 2*c*, rightmost green symbol). Surprisingly, different-type syllable pitch was changed significantly in the antiadaptive direction (Fig. 2*c*, rightmost blue symbol). Data from this representative experiment therefore indicate that, although only the targeted syllable was artificially pitch-shifted, pitch changed in the adaptive direction for the same vocal gesture produced in other sequential contexts and in the antiadaptive direction for different vocal gestures.

When data were combined across birds, we found an overall pattern of significant adaptive changes in target syllables. Figure 3 shows pitch changes for all target syllables (red symbols). We quantified mean pitch changes two different ways (Fig. 3*a,b*, M_{syls} ; Fig. 3*c,d*, $M_{\text{iterations}}$; see Materials and Methods) and obtained similar results. M_{syls} uses the mean pitch of each syllable as one data point, whereas $M_{\text{iterations}}$ uses each individual syllable iteration as one data point. The calculated fraction of sensory error compensated for by the end of the shift period in Figure 3*a* (27.1% on shift day 14, red line) and Figure 3*c* (35.2% on shift day 14, red line) is similar to the 36% compensation fraction reported earlier when the entire song was shifted (Sober and Brainard, 2009). Thus, the birds changed pitch by a similar amount despite performing a syllable-specific learning task. Furthermore, as reported previously in experiments where the entire song was shifted (Sober and Brainard, 2009), adaptive pitch changes on target syllables fell within the range of baseline pitch variability. The SD of baseline variation was 38.3 cents (averaged across birds), so birds changed target syllable pitch by ~ 1 SD.

To investigate whether error correction generalizes to the same vocalization produced in other sequential contexts ("same-type syllables"), we combined data across birds. Average pitch time courses for same-type syllables are shown in Figure 3*a, c* (green lines). Similar to the example experiment shown in Figure 2, the pitch of same-type syllables changed in the adaptive direction. Figure 3*b, d* displays the average pitch over shift days 12–14. Same-type syllables changed significantly in the adaptive direction (green asterisks), as did target syllables (red asterisks). Thus, although the magnitude of vocal changes depended somewhat on the averaging technique used ($M_{\text{iterations}}$ vs M_{syls}), both methods show that there were adaptive pitch changes for the same vocal gestures produced in other contexts.

Antiadaptive pitch changes in different vocal gestures

We also combined data across birds to ask whether error-corrective learning on one vocal gesture generalizes to different gestures. Figure 3*a, c* shows that different-type syllables (blue lines) gradually separated from target and same-type syllables. Surprisingly, we found that, on average, different-type syllables significantly changed pitch in the antiadaptive direction (Fig. 3*b,d*; blue asterisks). That is, whereas shifted song syllables (Fig. 3, red) change their pitch in the direction opposite the applied pitch shift, different-type syllables (Fig. 3, blue) exhibit pitch changes in the same direction as the applied pitch shift. We refer to the latter as "negative generalization" in the Discussion to indicate that nonperturbed syllables are changing in the opposite, or antiadaptive, direction as shifted syllables. Compared with the adaptive pitch changes observed in same-type syllables (Fig. 3*b,d*; green), the overall antiadaptive pitch change found in different-type syllables is smaller in magnitude (Fig. 3*b,d*; blue) and somewhat more variable over the duration of the experiment (Fig. 3*a,c*; blue). Nevertheless, these antiadaptive changes are both statisti-

Figure 2. Example of pitch-shift learning on a targeted vocal gesture and generalization to other contexts. **a**, Mean spectrograms showing vocal pitch before (left) and after (right) a bird experienced a -100 cent (downward) pitch shift to syllable A₂. Red box represents portion of the vocal sequence to which pitch shift was applied; syllables A₁, A₃, A₄, and A₅ were not pitch-shifted. **b**, Mean $\pm$ SEM pitches for baseline and last shift day, same motif as in **a**. The bird increased pitch of the targeted syllable and subsequent syllables in the motif, even though the latter were not artificially pitch-shifted. The magenta and blue symbols represent the pitch heard by the bird on the first and last shift days, respectively. Syllables A₁, A₃, A₄, and A₅ were not pitch-shifted, so no colored points are shown. **c**, Pitch changes for the same bird during shift period. Colored lines represent mean $\pm$ SEM pitch, measured in cents relative to baseline. Each syllable iteration is one data point ($M_{iterations}$; see Materials and Methods). The bird changed pitch of the targeted syllable in the adaptive direction, partially compensating for the sensory error. Same-type and different-type syllables were not artificially pitch-shifted but changed pitch in the adaptive and antiadaptive direction, respectively. *Significant pitch changes for syllables on shift days 12–14 ($p < 10^{-15}$, two-tailed t test).

cally significant and insensitive to the choice of how changes in vocal pitch are measured ($M_{iterations}$ vs M_{sylls} ; see Materials and Methods).

In conjunction with the adaptive changes observed in shifted syllables, antiadaptive changes in different-type syllables act to partially restore the pitch differences between song syllables. As shown in Figure 4a, introduction of a pitch shift (Fig. 4a, “pre-learning”) will perturb the relative pitches of hypothetical syllables

“A” and “Z” when syllable “A” is pitch-shifted. Whereas an adaptive change in the pitch of syllable “A” (Fig. 4a, “post-learning,” upward arrow) will partially restore this acoustic relationship (or “pitch contrast”), a concomitant antiadaptive change in the pitch of syllable “Z” can further restore the relative pitch.

As described in Materials and Methods, we determined whether pitch changes in different-type syllables contributed sig-

Figure 3. Pitch-shift learning on targeted vocal gestures generalizes to other gestures. *a*, Pitch changes during the first 14 shift days, combined across $n = 6$ birds. Birds were exposed to either 100 or -100 cents shift on the targeted syllable(s), but same-type and different-type syllables were not artificially pitch-shifted. Colored lines indicate mean $\pm$ SEM pitch, measured in cents relative to baseline. Each syllable is one data point (see Materials and Methods, M_{syll}). The pitches for birds exposed to 100 cents shift were multiplied by -1 . Thus, positive values signify that the birds changed pitch in a direction that is opposite to the artificial pitch shift (adaptive direction). *b*, Mean $\pm$ SEM pitch over shift days 12–14 for each syllable category, combined across birds. Each syllable is one data point (circles). Birds changed the pitch of targeted syllables to partially compensate for the error. They also changed same-type syllable pitch in the adaptive direction and different-type syllable pitch in the antiadaptive direction. Colored asterisks indicate significant pitch changes ($p < 0.05$, two-tailed t test). Black asterisks indicate that the indicated syllable categories have different pitch change distributions ($p < 0.05$, two-tailed two-sample t test). *c*, Same as *a*, except using each syllable iteration as one data point (see Materials and Methods, $M_{\text{iterations}}$). *d*, Same as *b*, except using each syllable iteration as one data point. Target syllables and same-type syllables changed pitch in the adaptive direction, whereas different-type syllables changed in the antiadaptive direction. Asterisks are defined as in *b*.

nificantly to this restoration of pitch contrast by quantifying restoration for each pair of target syllables and different-type syllables. A pairwise analysis isolated the contributions of pitch changes in different-type syllables and revealed that these changes contributed significantly to pitch contrast restoration across the learning epoch ($p < 0.05$, one-tailed paired-sample t test) and on average accounted for 19.6% of the total restoration (see Materials and Methods), with changes in the targeted syllables contributing the remainder of the restoration. These antiadaptive changes may therefore reflect a corrective mechanism that preserves the pitch relationships across syllables (see Discussion).

The observed pattern of vocal changes is not an artifact of targeting errors

As described in Materials and Methods, online targeting of pitch shifts to selected syllables was very accurate, with mean hit and false-positive rates of 92.1% and 0.9%, respectively. However,

because infrequent targeting errors sometimes resulted either in pitch shifts being applied to nontargeted syllables or a lack of pitch shift to targeted vocal gestures, it was important to assess whether our findings might have resulted from these targeting errors. This is very unlikely to have been the case. First, false-positive rates did not differ significantly for same-type and different-type syllables (1.8% and 0.6%, respectively, $p = 0.09$, two-tailed two-sample t test), suggesting that the difference in pitch changes shown in Figure 3 does not reflect a difference in the frequency with which pitch shifts were accidentally applied to same-type and different-type syllables. Second, we performed an alternate analysis in which we excluded the two same-type syllables with the highest false-positive rates (13% and 15%; remaining same-type syllables' false-positive rates did not exceed 2%). In this alternate analysis, the false-positive rates were nearly identical (0.4% and 0.6% for same-type and different-type syllables, respectively; $p = 0.55$), and the results of all other analyses were

Figure 4. Two patterns of generalization in vocal learning. *a*, Antiadaptive shifts preserve pitch relationships across syllables. Schematic represents changes in the pitch sung by the bird (black) and heard through auditory feedback (magenta) of two song syllables, “A” and “Z”. The dashed line indicates the relative pitch of the two syllables. Just after the onset of a downward pitch shift applied to syllable A (“Pre-learning”), the relative pitch of auditory feedback between the two syllables is altered. After error correction (“Post-learning”), syllable A exhibits an adaptive pitch change (i.e., a change in vocal pitch opposite the imposed sensory error), indicated by the upward-pointing black arrow. This pitch change does not completely correct the imposed pitch shift. However, syllable Z exhibits an “antiadaptive” pitch change (downward arrow). This contributes to partially restoring the relative pitch between auditory feedback from the two syllables (dashed line). *b*, Adaptive changes generalize to nearby syllables in a sequence. Here, the “credit” for a pitch error during syllable A is generalized to nearby syllables, resulting in “adaptive” pitch changes in multiple syllables (black arrows at right). Vocal plasticity in response to single-syllable pitch shifts may reflect an interaction between the patterns shown in *a* and *b*.

qualitatively identical to the original analysis. Third, if false-positive shifts caused pitch changes, we might also expect the reverse to be true, where higher false negative percentages on targeted syllables would cause smaller pitch changes. However, there was no correlation between number of false-negative shifts on target syllables and the amount they changed pitch ($p = 0.96$). Thus, although we cannot exclude the possibility that rare false-positive or false-negative pitch shifts may have affected the birds’ behavior, the above evidence suggests that this alone cannot account for our results.

Vocal changes are not predicted by acoustic similarity

The above analyses compare how errors during a particular syllable affect both different syllables and the same syllable produced in different sequences. However, the bird’s repertoire exists on a continuum, where some syllables are more acoustically similar to the targeted syllable than others. Many studies of generalization have found that the amount of transfer declined as movements became more dissimilar from the learned movement (Shadmehr and Mussa-Ivaldi, 1994; Roby-Brami and Burnod, 1995; Ghahramani et al., 1996; Field et al., 1999; Krakauer et al., 2000; Shadmehr, 2004). This has also been found in some studies of human speech, where learning transferred less as vocalizations became less similar to the training utterances (Villacorta et al., 2007; Cai et al., 2010; Rochet-Capellan et al., 2012). We therefore asked whether the amount that birds changed pitch of different-type vocal gestures was correlated with acoustic similarity to the targeted vocal gesture.

For each different-type syllable, we calculated the acoustic distance between it and the target syllable in a 3D space consisting

Figure 5. A vocal gesture’s acoustic similarity to targeted gesture does not predict learning transfer. Each symbol shows the mean pitch change for one different-type syllable over shift days 12–14 as a function of its acoustic distance to the targeted syllables (see Materials and Methods). Data from individual birds are shown with different symbol shapes. Two birds had two targeted syllables; thus, their syllables’ acoustic distance is the mean distance to both targeted syllables ($\blacktriangle$ and $\blacktriangledown$). The regression is not significant ($p = 0.7$), suggesting that amount of pitch change does not vary with the degree of similarity to the trained gesture.

of pitch, amplitude, and spectral entropy, acoustic features that have previously been shown to account for a substantial portion of acoustic variability in songbirds (Sober et al., 2008). We compared this acoustic distance with the amount the syllable pitch changed during the pitch shift period. We found no significant correlation between different-type syllable pitch changes and acoustic distance (Fig. 5). Furthermore, we performed three additional analyses in which acoustic distances were computed using only a single acoustic parameter (i.e., three separate one-dimensional analyses; see Materials and Methods). The additional analyses also failed to yield any significant correlation between acoustic similarity and adaptive vocal change ($p > 0.4$ in all cases). Therefore, in contrast to several studies of human speech (see Discussion), we find no significant relationship between change in pitch and acoustic similarity to the perturbed vocal gesture, suggesting that the pitch change observed in non-trained song syllables is not strongly related to their acoustic similarity to the trained gesture.

Sequential context affects vocal changes

After analyzing how birds changed the pitch of other vocal gestures of varying acoustical similarity, we investigated whether generalization of error correction was sequence-dependent. Specifically, we asked whether the amount of pitch change in a vocal gesture depended on how closely it was produced to the targeted vocal gesture in the motif. For example, Figure 2*b* shows that the first syllable sung after the target syllable (A_3) changed pitch by a greater amount than the third syllable sung after the target syllable (A_5), suggesting a correlation between sequential distance and the amount of pitch change.

Across all birds, we found that, on average, nontargeted syllables changed in the adaptive direction when they were produced in close proximity to the pitch-shifted syllable. Figure 6 shows that both same-type and different-type syllables produced immediately adjacent to the targeted syllable changed pitch in the adaptive direction. Linear regression analyses revealed significant relationships between sequential distance and changes in vocal pitch in all cases (Fig. 6, blue and green lines), indicating that the amount of pitch change decreased with increasing sequential distance from the targeted syllable. This sequence-dependent pat-

Figure 6. Transfer of learning across vocal gestures in a stereotyped sequence depends on sequential distance. *a*, Each point shows the mean $\pm$ SEM pitch changes for same-type syllables that were produced at a stereotyped distance from the nearest targeted syllable. Data are combined across $n = 6$ birds and use syllables from shift days 12–14. Each syllable iteration is one data point (see Materials and Methods, $M_{\text{iterations}}$). Adaptive pitch change lessened and turned into antiadaptive pitch change as the sequential distance increased ($p < 0.0001$ for both lines). The slopes differed for syllables produced before versus after target syllables: $*p < 10^{-10}$ (F test). Results were qualitatively similar when using M_{syll} , but did not reach significance, possibly because there were too few data points at each distance. *b*, Same as *a*, except showing different-type syllables. There was increasing antiadaptive pitch change as sequential distance increased, and different slopes for syllables produced before versus after target syllables: $*p < 10^{-10}$ (F test). Only syllables that were produced within stereotyped sequences were included in this analysis (see Materials and Methods).

tern of adaptive vocal changes suggests that error information from one syllable is used to generate adaptive vocal changes in nearby syllables as well, as schematized in Figure 4*b*.

The overall pattern of pitch changes shown in Figure 6 may therefore reflect a combination of category-specific (same-type vs different-type) and sequence-dependent effects. That is, although on average same-type and different-type syllables changed in opposite directions (a “category-specific” effect, Fig. 3*b,d*), both types of syllables displayed adaptive changes in syllables immediately adjacent to the targeted syllable with smaller or negative changes at greater distances (a “sequence-dependent” effect). In the Discussion, we speculate that these two effects result from two distinct processes underlying vocal adaptation.

We also asked whether experiencing vocal errors at a single vocal gesture led to asymmetrical effects on the surrounding movements. For example, Figure 2*b* shows an example in which significant pitch changes occur in the syllables after, but not before, the targeted syllable. We therefore tested whether the magnitude of pitch change falls off more quickly at positive and negative sequential distances by comparing the absolute slopes of the regression lines shown in Figure 6*a, b*. We found a significant difference in slopes (Fig. 6*a*, asterisk; partial F test), indicating an asymmetrical pattern: the magnitude of pitch change decreased more rapidly (in terms of sequential distance) for same-type syllables produced before the targeted syllable. When we performed the same analysis for different-type syllables, we also found a significant difference in slopes (Fig. 6*b*, asterisk; partial F test) but with the opposite pattern: the amount of pitch change decreased more rapidly for syllables produced after the targeted syllable.

Finally, it is possible that the apparent effect of sequence on vocal learning (Fig. 6) and the lack of an effect of acoustic similarity (Fig. 5) result from a confound between sequential and acoustic distances. For example, if vocal changes depended only on acoustic similarity, and nearby syllables were more acoustically similar to the targeted syllable than distant syllables, we

could observe the same learning pattern as seen in Figure 6 and erroneously conclude that learning depends on sequential context when it actually depended on acoustic distance. However, we found that syllables’ sequential and acoustic distances were not correlated ($p = 0.36$ for different-type syllables, $p = 0.84$ for same-type syllables, $p = 0.32$ for combined syllables), suggesting that acoustic similarity cannot account for the patterns of vocal changes observed in Figure 6. The evidence therefore suggests that acoustic and sequential distances are largely independent in our dataset and make different contributions to the observed patterns of vocal plasticity.

Discussion

We found that songbirds use sensory errors to change the acoustics of individual vocal gestures occurring in particular sequential contexts. Songbirds also modified the pitches of unperturbed vocal gestures, which we suggest is evidence for transfer of learning, or generalization. As hypothesized, adaptive pitch changes generalized to the same syllables produced in other sequential contexts. We also found two surprising results. First, the average transfer of learning to different vocal gestures was in the antiadaptive direction, signifying negative generalization. Second, generalization magnitude was negatively correlated with the sequential distance from the pitch-shifted syllable.

Learning to alter specific vocal gestures

Although prior studies have used negative reinforcement to show that birds can alter specific syllables (Tumer and Brainard, 2007), our results demonstrate that birds adjust individual syllables in response to naturalistic pitch perturbations (Fig. 3, red), suggesting that birds generate internal error signals specific to particular gestures. Our results parallel similar findings in humans, who alter the acoustics of single phonemes in response to manipulated auditory feedback (Houde and Jordan, 1998; Jones and Munhall,

2000). Gesture-specific error correction may therefore be a general principle to help maintain learned vocal behaviors.

Learning generalizes to the same gesture in other contexts

We found that compensatory changes in a vocal gesture generalize to the same gesture produced in other contexts (Fig. 3, green). In humans, learning similarly generalizes to the same vowel in different words (Houde and Jordan, 1998; Villacorta et al., 2007). Our results show that this phenomenon occurs in songbirds during natural vocal behaviors, not only in the reduced “training utterance–test utterance” paradigms used in many human studies. This type of generalization may be advantageous because it reduces the need to relearn how to produce a vocal gesture correctly in each possible context.

Negative generalization to different gestures

Unexpectedly, learning generalized in the antiadaptive direction for different-type syllables (Fig. 3, blue). To our knowledge, negative generalization has not been reported in any sensorimotor learning studies. One study (Taylor et al., 2013) involving visual rotations showed that, when subjects received limited visual feedback and reached to one target, errors to targets in the opposite direction were in the same direction as the sensory perturbation. However, these effects disappeared when more visual feedback was provided and, as the authors noted, likely reflect normal adaptation in a Cartesian reference frame, not negative generalization.

Why would negative generalization occur? Intuitively, altering different-type syllables should cause sensory error signals and a subsequent return to baseline (Sober and Brainard, 2009). However, our data suggest that the brain does not treat vocal gestures as isolated units, each having a specific set of desired acoustic features. Rather, we propose that the brain seeks to maintain specific acoustic relationships among vocal gestures. The observed antiadaptive pitch changes act to partially reestablish the preexisting pitch contrast between syllables (Fig. 4a). Therefore, although songbirds generate syllable-specific vocal corrections, alterations of unperturbed syllables could be used to maintain the relationships between vocal gestures.

A recent study simulating reward-based learning (Darshan et al., 2014) proposes a possible computational mechanism for negative generalization. The authors showed that, when model neurons that transform sensory input into motor output respond to a wide range of sensory stimuli, negative generalization (or “destructive interference”) can result. Although future work is required to assess whether the songbird brain implements such mechanisms, these modeling results suggest that negative generalization might result from auditory responses in the song system. Auditory activity is influenced by multiple syllables in a sequence (Margoliash and Fortune, 1992; Lewicki and Arthur, 1996; Dave and Margoliash, 2000), suggesting that the effects on nontargeted syllables could reflect the relatively long integration time observed in auditory responses.

Generalization is not predicted by acoustic similarity

We found that learning transfer was uncorrelated with acoustic distance from the targeted syllable (Fig. 5). This contrasts with some human speech (Cai et al., 2010; Rochet-Capellan et al., 2012) and limb movement (Shadmehr and Mussa-Ivaldi, 1994; Roby-Brami and Burnod, 1995; Ghahramani et al., 1996; Field et al., 1999; Krakauer et al., 2000; Shadmehr, 2004) studies finding more generalization for gestures similar to the trained gesture. The differing results may arise because different-type syllables

represent entirely different movement categories to the bird, analogous to performing reaching versus ball-throwing movements. Although some studies have found that learning generalizes across categorically different movements (Conditt et al., 1997; Morton and Bastian, 2004; Krakauer et al., 2006; Alexander et al., 2011), others have not (Roby-Brami and Burnod, 1995; Field et al., 1999; Krakauer et al., 2000; Palmer and Meyer, 2000). Analogously, in speech studies, the training and test utterances may not have been internally represented as separate categories.

Differences between our results and those in humans may also reflect differences in the degeneracy of speech and birdsong. Generalization studies in humans have explored vowel formant changes (Houde and Jordan, 1998; Pile and Dajani, 2007; Villacorta et al., 2007; Cai et al., 2010; Rochet-Capellan et al., 2012). Vowel production involves precisely configuring vocal tract shape, and particular articulator configurations are thought to uniquely determine the formant frequencies in most cases (Hogden et al., 1996; Stevens, 1998; Mitra et al., 2010). In contrast, pitch control in songbirds appears to be more degenerate. Pitch is controlled by both air pressure and vocal fold tension (Goller and Suthers, 1996; Gardner et al., 2001; Laje et al., 2002), and different motor patterns can produce acoustically similar vocalizations (Suthers et al., 1996; Leonardo and Fee, 2005). Additionally, whereas voiced speech has one sound source, the songbird vocal organ (syrinx) contains two independently controlled sources, which might allow the same acoustic feature to be produced using either source (but see Secora et al., 2012). The discrepancy between our findings (Fig. 5) and those in humans might therefore arise because acoustically similar vocalizations necessarily reflect similar motor programs in humans but not in songbirds.

Importantly, our finding that acoustic similarity does not predict generalization may depend on the choice of acoustic parameters (pitch, amplitude, and spectral entropy; see Results) used to measure similarity. We used these parameters because they account for a significant fraction of behavioral variation (Sober et al., 2008). However, computing acoustic differences with more or other parameters might yield different results. Also, the degeneracy in pitch control described above may require measurements of muscle dynamics during song to obtain a more complete picture of gesture similarity. Nevertheless, the lack of a robust relationship between generalization and similarity may reflect either a significant difference between songbirds and humans or between vocal error correction across a natural vocal repertoire and experiments using a more limited set of vocalizations.

Sequence-dependent generalization

Unexpectedly, we found that, for same-type syllables, generalization decreased with increasing sequential distance from the pitch-shifted syllable (Fig. 6a). Furthermore, although on average they exhibited antiadaptive pitch changes (Fig. 3), different-type syllables also exhibited adaptive changes in syllables adjacent to the targeted syllable and antiadaptive changes at larger distances (Fig. 6b) and therefore exhibited similar sequence-dependent differences in generalization. To our knowledge, the finding that generalization varies with sequential context is novel in both the speech and limb movement literature. Importantly, prior studies suggest that this phenomenon does not reflect biomechanical constraints preventing syllable-specific pitch changes. Songbirds can be trained to modify the pitch of individual syllables without altering temporally adjacent syllables (Tumer and Brainard, 2007) and can modulate pitch within syllables with 10 ms resolution (Charlesworth et al., 2011). The spread of generalization to

adjacent syllables shown in Figure 6 is therefore highly unlikely to reflect a motor constraint preventing songbirds from generating syllable-specific changes.

What might account for this sequence-dependent generalization? We speculate that, when a sensory error occurs during one syllable in a sequence, the “error credit assignment” (Wolpert et al., 2011) is not localized in time to just that syllable but is partially assigned to nearby syllables in the sequence (Fig. 4*b*). Because trial-by-trial variations in consecutive syllables’ acoustics are typically correlated (Sober et al., 2008), if a bird makes an error on one syllable, it is likely to have made an error on the next and previous one as well. Thus, an advantageous strategy may be to extend the credit assignment function several syllables backward and forward in time from the syllable in which an error is detected. Why, then, are pitch changes not observed in syllables adjacent to those targeted by negative reinforcement (Tumer and Brainard, 2007)? One possible explanation is that the highly salient (and artificial) negative reinforcement signals provide additional information that birds use to localize vocal changes in time.

Interactions between sequence-dependent and category-specific effects

Although they show similarly sequence-dependent patterns of generalization (Fig. 6), same-type and different-type syllables differ significantly in the average magnitude and direction of generalization (Fig. 3). We therefore speculate that our results reflect a combination of sequence-dependent and category-specific effects. That is, the changes in generalization across sequential distance in both same- and different-type syllables (regression lines in Fig. 6) may reflect a sequence-dependent effect of error credit assignment (Fig. 4*b*), as discussed above. At the same time, the overall antiadaptive bias in different-type syllables may reflect an additional mechanism for reestablishing acoustic relationships between syllables (Fig. 4*a*) that affects different-type syllables but is weaker or absent in same-type syllables (and is thus syllable category-specific). Therefore, changes in different-type syllables (Fig. 6*b*) might reflect the combination of sequence-dependent (adaptive changes at small sequential distances) and category-specific (antiadaptive changes across all different-type syllables) effects. This combination of adaptive and antiadaptive effects might account for the relatively small average magnitude of changes in different-type syllables (Fig. 3*b*).

A key remaining question is whether generalization in same-type syllables (Fig. 6*a*) includes a global antiadaptive component in addition to the strongly sequence-dependent adaptive changes. The regression line shown in Figure 6*a* crosses the $y = 0$ line, suggesting that antiadaptive shifts might be present at long sequential distances, as is the case for different-type syllables (Fig. 6*b*). However, because our dataset contains very few examples of longer distances (and no examples of distances >7 syllables), we do not have sufficient statistical power to determine whether same-type syllables exhibit global antiadaptive changes in addition to local adaptive changes, as is the case for different-type syllables.

An alternate hypothesis is that, rather than resulting from the combined action of two processes (sequence-dependent error assignment and category-specific changes), the patterns in Figure 6 might reflect variations in how a sequence-dependent error signal affects different motor programs. In this scenario, a sensory error during one syllable would result in a modulatory signal affecting multiple syllables in a sequence, but this signal would interact differently with the motor programs underlying same-type and different-type syllable to bias vocal changes in the adap-

tive and antiadaptive directions, respectively. Future studies showing how auditory errors reshape premotor activity will provide mechanistic insight into how sensory signals interact with motor changes in sequenced behaviors.

References

- Alexander MS, Flodin BW, Marigold DS (2011) Prism adaptation and generalization during visually guided locomotor tasks. *J Neurophysiol* 106:860–871. [CrossRef Medline](#)
- Bolhuis JJ, Okanoya K, Scharff C (2010) Twitter evolution: converging mechanisms in birdsong and human speech. *Nat Rev Neurosci* 11:747–759. [CrossRef Medline](#)
- Cai S, Ghosh SS, Guenther FH, Perkell JS (2010) Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization. *J Acoust Soc Am* 128:2033–2048. [CrossRef Medline](#)
- Charlesworth JD, Tumer EC, Warren TL, Brainard MS (2011) Learning the microstructure of successful behavior. *Nat Neurosci* 14:373–380. [CrossRef Medline](#)
- Condit M, Gandolfo F, Mussa-Ivaldi FA (1997) The motor system does not learn the dynamics of the arm by rote memorization of past experience. *J Neurophysiol* 78:554–560. [Medline](#)
- Darshan R, Leblois A, Hansel D (2014) Interference and shaping in sensorimotor adaptations with rewards. *PLoS Comput Biol* 10:e1003377. [CrossRef Medline](#)
- Dave AS, Margoliash D (2000) Song replay during sleep and computational rules for sensorimotor vocal learning. *Science* 290:812–816. [CrossRef Medline](#)
- Doupe AJ, Kuhl PK (1999) Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci* 22:567–631. [CrossRef Medline](#)
- Field DP, Shipley TF, Cunningham DW (1999) Prism adaptation to dynamic events. *Percept Psychophys* 61:161–176. [CrossRef Medline](#)
- Gardner T, Cecchi G, Magnasco M, Laje R, Mindlin G (2001) Simple motor gestures for birdsongs. *Phys Rev Lett* 87:1–4. [CrossRef Medline](#)
- Ghahramani Z, Wolpert DM, Jordan MI (1996) Generalization to local remappings of the visuomotor coordinate transformation. *J Neurosci* 16:7085–7096. [Medline](#)
- Goller F, Suthers RA (1996) Role of syringeal muscles in controlling the phonology of bird song. *J Neurophysiol* 76:287–300. [Medline](#)
- Hahnloser RH, Kozhevnikov AA, Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419:65–70. [CrossRef Medline](#)
- Hoffmann LA, Kelly CW, Nicholson DA, Sober SJ (2012) A lightweight, headphones-based system for manipulating auditory feedback in songbirds. *J Vis Exp JoVE*:e50027. [CrossRef Medline](#)
- Hogden J, Lofqvist A, Gracco V, Zlokarnik I, Rubin P, Saltzman E (1996) Accurate recovery of articulator positions from acoustics: new conclusions based on human data. *J Acoust Soc Am* 100:1819–1834. [CrossRef Medline](#)
- Houde JF, Jordan MI (1998) Sensorimotor adaptation in speech production. *Science* 279:1213–1216. [CrossRef Medline](#)
- Jones JA, Munhall KG (2000) Perceptual calibration of F0 production: evidence from feedback perturbation. *J Acoust Soc Am* 108:1246–1251. [CrossRef Medline](#)
- Kao MH, Doupe AJ, Brainard MS (2005) Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature* 433:638–643. [CrossRef Medline](#)
- Krakauer JW, Pine ZM, Ghilardi MF, Ghez C (2000) Learning of visuomotor transformations for vectorial planning of reaching trajectories. *J Neurosci* 20:8916–8924. [Medline](#)
- Krakauer JW, Mazzoni P, Ghazizadeh A, Ravindran R, Shadmehr R (2006) Generalization of motor learning depends on the history of prior action. *PLoS Biol* 4:e316. [CrossRef Medline](#)
- Laje R, Gardner T, Mindlin G (2002) Neuromuscular control of vocalizations in birdsong: a model. *Phys Rev* 65:1–8. [Medline](#)
- Leonardo A, Fee MS (2005) Ensemble coding of vocal control in birdsong. *J Neurosci* 25:652–661. [CrossRef Medline](#)
- Lewicki MS, Arthur BJ (1996) Hierarchical organization of auditory temporal context sensitivity. *J Neurosci* 16:6987–6998. [Medline](#)
- Margoliash D, Fortune ES (1992) Temporal and harmonic combination-sensitive neurons in the zebra finch’s HVC. *J Neurosci* 12:4309–4326. [Medline](#)

- Mitra V, Nam H, Espy-Wilson CY, Saltzman E, Goldstein L (2010) Retrieving tract variables from acoustics: a comparison of different machine learning strategies. *IEEE J Sel Top Signal Process* 4:1027–1045. [CrossRef Medline](#)
- Morton SM, Bastian AJ (2004) Prism adaptation during walking generalizes to reaching and requires the cerebellum. *J Neurophysiol* 92:2497–2509. [CrossRef Medline](#)
- Palmer C, Meyer RK (2000) Conceptual and motor learning in music performance. *Psychol Sci* 11:63–68. [CrossRef Medline](#)
- Pile E, Dajani H (2007) Talking under conditions of altered auditory feedback: does adaptation of one vowel generalize to other vowels. *Int Congr Phonet Sci* 645–648.
- Roby-Brami A, Burnod Y (1995) Learning a new visuomotor transformation: error correction and generalization. *Brain Res Cogn Brain Res* 2:229–242. [CrossRef Medline](#)
- Rochet-Capellan A, Ostry DJ (2011) Simultaneous acquisition of multiple auditory-motor transformations in speech. *J Neurosci* 31:2657–2662. [CrossRef Medline](#)
- Rochet-Capellan A, Richer L, Ostry DJ (2012) Nonhomogeneous transfer reveals specificity in speech motor learning. *J Neurophysiol* 107:1711–1717. [CrossRef Medline](#)
- Secora KR, Peterson JR, Urbano CM, Chung B, Okanoya K, Cooper BG (2012) Syringeal specialization of frequency control during song production in the bengalese finch (*Lonchura striata domestica*). *PloS One* 7:e34135. [CrossRef Medline](#)
- Shadmehr R (2004) Generalization as a behavioral window to the neural mechanisms of learning internal models. *Hum Mov Sci* 23:543–568. [CrossRef Medline](#)
- Shadmehr R, Mussa-Ivaldi FA (1994) Adaptive representation of dynamics during learning of a motor task. *J Neurosci* 14:3208–3224. [Medline](#)
- Sober SJ, Brainard MS (2009) Adult birdsong is actively maintained by error correction. *Nat Neurosci* 12:927–931. [CrossRef Medline](#)
- Sober SJ, Brainard MS (2012) Vocal learning is constrained by the statistics of sensorimotor experience. *Proc Natl Acad Sci U S A* 109:21099–21103. [CrossRef Medline](#)
- Sober SJ, Wohlgemuth MJ, Brainard MS (2008) Central contributions to acoustic variation in birdsong. *J Neurosci* 28:10370–10379. [CrossRef Medline](#)
- Stevens K (1998) *Acoustic phonetics*. Cambridge, MA: Massachusetts Institute of Technology.
- Suthers RA, Goller F, Hartley RS (1996) Motor stereotypy and diversity in songs of mimic thrushes. *J Neurobiol* 30:231–245. [CrossRef Medline](#)
- Taylor JA, Hieber LL, Ivry RB (2013) Feedback-dependent generalization. *J Neurophysiol* 109:202–215. [CrossRef Medline](#)
- Tremblay S, Houle G, Ostry DJ (2008) Specificity of speech motor learning. *J Neurosci* 28:2426–2434. [CrossRef Medline](#)
- Tumer EC, Brainard MS (2007) Performance variability enables adaptive plasticity of ‘crystallized’ adult birdsong. *Nature* 450:1240–1244. [CrossRef Medline](#)
- Villacorta VM, Perkell JS, Guenther FH (2007) Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *J Acoust Soc Am* 122:2306–2319. [CrossRef Medline](#)
- Wohlgemuth MJ, Sober SJ, Brainard MS (2010) Linked control of syllable sequence and phonology in birdsong. *J Neurosci* 30:12936–12949. [CrossRef Medline](#)
- Wolpert DM, Diedrichsen J, Flanagan JR (2011) Principles of sensorimotor learning. *Nat Rev Neurosci* 12:739–751. [CrossRef Medline](#)
- Yu AC, Margoliash D (1996) Temporal hierarchical control of singing in birds. *Science* 273:1871–1875. [CrossRef Medline](#)