Hearing Research 348 (2017) 98–111

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research Paper

Contributions of sensory tuning to auditory-vocal interactions in marmoset auditory cortex

Steven J. Eliades [a, *], Xiaoqin Wang [b]

[a] Department of Otorhinolaryngology: Head and Neck Surgery, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA
[b] Laboratory of Auditory Neurophysiology, Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD, USA

ARTICLE INFO

Article history:
Received 9 December 2016
Received in revised form
27 February 2017
Accepted 2 March 2017
Available online 9 March 2017

Keywords:
Auditory cortex
Auditory-vocal interaction
Marmoset
Vocalization
Sensory-motor

ABSTRACT

During speech, humans continuously listen to their own vocal output to ensure accurate communication. Such self-monitoring is thought to require the integration of information about the feedback of vocal acoustics with internal motor control signals. The neural mechanism of this auditory-vocal interaction remains largely unknown at the cellular level. Previous studies in naturally vocalizing marmosets have demonstrated diverse neural activities in auditory cortex during vocalization, dominated by a vocalization-induced suppression of neural firing. How underlying auditory tuning properties of these neurons might contribute to this sensory-motor processing is unknown. In the present study, we quantitatively compared marmoset auditory cortex neural activities during vocal production with those during passive listening. We found that neurons excited during vocalization were readily driven by passive playback of vocalizations and other acoustic stimuli. In contrast, neurons suppressed during vocalization exhibited more diverse playback responses, including responses that were not predictable by auditory tuning properties. These results suggest that vocalization-related excitation in auditory cortex is largely a sensory-driven response. In contrast, vocalization-induced suppression is not well predicted by a neuron's auditory responses, supporting the prevailing theory that internal motor-related signals contribute to the auditory-vocal interaction observed in auditory cortex.

1. Introduction

Recent investigations in both humans and non-human primates have begun to reveal the role of the central auditory system, and in particular the auditory cortex, in representing the sound of an animal's own vocalizations during vocal production. During vocal communication, vocalized sounds are heard by both the intended recipients and the individual producing them (Bekesy, 1949). Neural encoding of this vocal feedback is thought to be crucial for monitoring one's own voice (Hickok et al., 2011; Houde and Nagarajan 2011; Levelt, 1983), and may play a role in feedback-dependent control of vocalization in both animals (Brumm et al., 2004; Leonardo and Konishi, 1999; Schuller et al., 1974; Sinnott et al., 1975) and humans (Burnett et al., 1998; Houde and Jordan, 1998; Lane and Tranel, 1971; Lee, 1950).

Single neuron recordings in the auditory cortex of the marmoset (*Callithrix jacchus*), a highly vocal New World primate, have demonstrated the presence of two types of responses during vocal production, vocalization-induced suppression and vocalization-related excitation (Eliades and Wang, 2003). Vocalization-induced suppression affects approximately 70% of neurons in marmoset auditory cortex (Eliades and Wang, 2013), is observed across different types of vocalizations, and is thought to be caused by inhibitory signals originating from brain regions that initiate and control vocal production. Moreover, neurons showing vocalization-induced suppression exhibit an increased sensitivity to alterations in auditory feedback during vocalization and may play a role in self-monitoring (Eliades and Wang, 2008a, 2012). In contrast, neurons showing vocalization-related excitation, representing a small proportion of auditory cortex neurons, tend to respond during a more limited set of vocalization types (Eliades and Wang, 2013) and are less sensitive to altered auditory feedback (Eliades and Wang, 2008a). The origin of the differences between these two groups of neurons is not clear.

Several recent parallel human investigations have addressed the

[*] Corresponding author. Department of Otorhinolaryngology: Head and Neck Surgery, University of Pennsylvania Perelman School of Medicine, 3400 Spruce St, 5 Ravdin, Philadelphia, PA, 19104, USA.
E-mail address: seliades@med.upenn.edu (S.J. Eliades).

http://dx.doi.org/10.1016/j.heares.2017.03.001
0378-5955/© 2017 Elsevier B.V. All rights reserved.

suppression of human auditory cortex during speech (Crone et al., 2001; Curio et al., 2000; Christoffels et al., 2007; Flinker et al., 2010; Greenlee et al., 2011; Heinks-Moldonado et al., 2005; Houde et al., 2002). These studies demonstrated that auditory cortex is activated during both speech production and perception, with reduced responses during speaking, termed speech-induced suppression. Human studies have also demonstrated vocal feedback sensitivity similar to that observed in marmosets (Behroozmand and Larson, 2011; 2016; Chang et al., 2013). However, a lack of spatial resolution has prevented a more accurate characterizations of the auditory component of speech production-related activity in human auditory cortex.

More recent work in rodents has begun to reveal possible neural circuits underlying such suppression. These experiments have revealed a direct suppression of auditory cortex from connections originating in M2, a putative equivalent of pre-motor cortex (Nelson et al., 2013; Schneider et al., 2014; Schneider and Mooney, 2015a). When paired with a predictable motor-triggered tone, there is a suppression of the tone-evoked sensory response in auditory cortex (Schneider and Mooney, 2015b), similar to what has been described in human subjects (Martikainen et al., 2005; Agnew et al., 2013). Although this suppression of self-generated sensory responses is thought to have a generally similar mechanism to vocalization- and speech-induced suppression, the extent of the mechanistic overlap remains an open question.

A better understanding of these auditory-vocal interactions and their underlying mechanisms requires a more thorough characterization of the contributions of sensory inputs. However, our previous efforts to examine these integration mechanisms has not revealed meaningful differences in auditory tuning between vocalization suppressed and excited auditory cortex neurons (Eliades and Wang, 2003, 2008a). Here we conducted further analyses of single neuron recordings obtained from auditory cortex of two naturally vocalizing marmosets (Eliades and Wang, 2013) in order to more specifically compare auditory and vocal responses of each neuron. We expand on our previous results by examining responses to previously un-analyzed auditory control stimuli. In contrast to our previous findings in which auditory tuning of suppressed and excited neurons were found to be similar, this new analysis demonstrates that vocalization-related excitation is highly predictable based on a neurons passive auditory responses, whereas neurons exhibiting vocalization-induced suppression exhibit more diverse auditory tuning properties, including vocal responses that could not be predicted based upon passive auditory responses. Given the scarcity of single neuron data obtained from naturally vocalizing monkeys, these results add valuable contributions to our understanding of auditory-vocal interaction mechanisms in the primate brain.

2. Materials methods

All experiments were conducted under the guidelines and protocols approved by the Johns Hopkins University Animal Care and Use Committee. The neural data analyzed in this report were obtained from the same animals studied in our previous work (Eliades and Wang, 2013). In these chronic recording experiments, we typically collected a large amount of data under multiple experimental conditions from each neuron. In the Eliades and Wang (2013) study, we focused on comparing vocal responses in auditory cortex of marmosets between different classes of marmoset vocalization. This previous publication, however, only included responses from a limited subset of the auditory control stimuli tested. The present study includes analyses of previously unpublished neural responses to auditory control stimuli and additional analyses including modeling of vocal responses, described further below. Details of the neural recording experiments can be found in our previous publication (Eliades and Wang, 2013) and are only briefly described below.

2.1. Electrophysiological recordings

Two marmoset monkeys (*Callithrix jacchus*) were each implanted bilaterally with Warp-16 multielectrode arrays (Neuralynx, Bozeman, MT). Each array contained 16 individually moveable microelectrodes (2–4 MOhm impedances). Details on the electrode arrays and recordings, as well as spike sorting procedures, have been previously described (Eliades and Wang, 2008a,b). Auditory cortex was located with standard single-electrode methods prior to array placement (Lu et al., 2001). Both hemispheres were recorded for each animal, starting first in the left hemisphere and subsequently in both simultaneously. Histological examination showed arrays spanning primary auditory cortex, lateral belt and possibly a portion of parabelt fields (Eliades and Wang, 2008b).

2.2. Auditory response characterization

Auditory responses were measured within a soundproof chamber (Industrial Acoustics, Bronx, NY), with the animal seated and head restrained in a custom primate chair. Auditory stimuli were presented free-field by a speaker (B&W DM601) located 1 m in front of the animal. Stimuli included both tone- and noise-based sounds to assess frequency tuning and rate-level responses. Tone-based stimuli consisted of randomly ordered 100 ms pips at 500 ms inter-stimulus intervals, with frequencies spanning 1–32 kHz (5 octaves) at 1/10th octave steps. During most sessions, frequency tuning was measured at 3 sound pressure levels (30, 50, 70 dB SPL); a subset of sessions used a more extensive SPL range (−10 to90 dB in 10 or 20 dB intervals) to measure the full frequency response area (FRA) map. Band-pass noise stimuli were presented similarly to tones, but were 200 ms in duration, 0.5 octave in bandwidth, and the center frequency varied at 1/5th octave steps. Selected tone and bandpass frequencies were tested more extensively at multiple SPLs (−10 to 90 dB in 10 dB intervals) to assess rate-level tuning. Rate-level functions using wideband (white) noise stimuli were also collected from all units.

In addition, multiple examples of recorded vocalizations were played at different sound levels ("playback"). These include samples of the animal's own vocalizations (previously recorded from that animal) and conspecific vocalization samples (from other animals living in the marmoset colony). These included multiple exemplars (6–10) from each of the four major marmoset vocalization classes: phee, trillphee, trill, and twitter (Agamaite et al., 2015; Epple, 1968). Based upon the responses to these vocalization stimuli, one or two samples of each call type were selected and presented at multiple SPLs (0–90 dB in 10 dB steps) to measure vocal rate-level tuning. All vocalization samples were previously recorded at 50 kHz sampling rate, filtered to exclude low-frequency (<1 kHz) background noise, and normalized to have equal stimulus power. A subset of vocalization stimuli were also presented with a parametrically varying mean frequency, computed using a hetrodyning technique (Schuller et al., 1974). This technique involves serial convolution of a vocal signal with cosines of different frequencies and results in a linear frequency shift of a desired magnitude. Samples were first up-sampled (3×), scaled in frequency by convolution with a 25 kHz cosine, high-pass filtered to remove the aliased signal, convolved with a second cosine of 25-*f* kHz (where *f* is the desired frequency shift), low-pass filtered, and finally down-sampled back to the original sample rate. The responses from these additional vocalization stimuli, including parametric changes in loudness and mean frequency, were not included in previous analyses.

2.3. Vocal recordings

Simultaneous vocal and neural recordings were performed following auditory testing. These were performed either in the marmoset colony (Eliades and Wang, 2008a,b), allowing the subject to vocally interact with other animals, or in the laboratory, where the animal engaged in antiphonal calling with computer-controlled playback of conspecific vocalizations (Miller and Wang, 2006). Vocal production was recorded using a directional microphone (AKG C1000S) placed ~20 cm in front of the animal and digitized at a 50 kHz sampling rate (National Instruments PCI-6052E) and synchronized with neural recordings. Individual vocalizations were extracted from the microphone recording and manually classified into established marmoset vocalization types (Agamaite et al., 2015) based upon visual inspection of their spectrograms.

2.4. Data analysis

Neural responses to individual vocalizations were calculated by comparing the firing rate during vocalization to spontaneous activity before vocal onset. Individual vocalization responses were quantified with a normalized metric, the response modulation index (RMI) to correct for firing rate differences between units (Eliades and Wang, 2003).

$$\mathrm{RMI} = \left(\mathrm{R_{vocal}} - \mathrm{R_{prevocal}}\right) \Big/ \left(\mathrm{R_{vocal}} + \mathrm{R_{prevocal}}\right)$$

where $\mathrm{R_{vocal}}$ is the firing rate during vocalization and $\mathrm{R_{prevocal}}$ is the average rate before vocalization. Negative RMIs indicate suppression during vocalization and positive values indicate strongly driven activity. The overall response of a neuron to a given vocalization type was measured by averaging the RMI from multiple individual vocalizations. Only units with sufficient samples of a given vocalization type (≥ 4) were included for analysis. Responses to playback of vocalization stimuli were similarly quantified. To differentiate, RMI measured during vocal production are referred to as 'Vocal RMI' and measurements during playback of vocal stimuli as 'Auditory RMI'. Only those vocal stimuli presented at similar SPLs to vocal production (determined *post-hoc* separately for each class of vocalization and for each session) were included in the Auditory RMI calculation.

Auditory tuning properties, include center frequency (CF) and rate-level tuning, were measured from responses to tone, bandpass, and wideband noise stimuli. CF was defined as the frequency evoking the highest firing rate response across all SPLs tested. A separate measurement of CF was performed using those tones matching the loudness of the vocalizations actually produced by the animal (typically $\geq$70 dB SPL) for secondary analyses. In cases where there was a response to both tone and bandpass stimuli, the CF was chosen from the stimulus with the strongest response. Rate-level responses were measured for both simple stimuli and vocal playback stimuli; however, the two correlated highly and therefore rate-level analysis is presented only for vocal playback responses. A Monotonicity Index (MI) was measured for each rate-level response, defined as the firing rate to the loudest stimulus divided by the strongest response (Sadagopan and Wang, 2008). An MI > 0.5 indicates a monotonically increasing or saturating rate-level function, while an MI < 0.5 indicates a non-monotonic (peaked) function.

Statistical significance of differences between vocalization and playback responses (RMIs) was determined for individual units using *Wilcoxon signed-rank* testing. Trends across neural populations were tested using correlation coefficients and *Kruskal-Wallis non-parametric ANOVAs*. P values < 0.05 were considered statistically significant.

2.5. Vocal response model

In order to better characterize the contribution of auditory tuning to vocal responses, a simple linear model was created, similar to that of Bar-Yosef et al. (2002). Similar models have also been used successfully to explain responses to complex stimuli in sub-cortical auditory brain areas (Bauer et al., 2002; Holmstrom et al., 2010). First, the acoustic frequency spectrum of each vocalization was measured using a power-spectral density function. Because only the four major marmoset vocalization types were used, none of which contain low frequency spectral information, frequencies below 2 kHz were discarded. The power-spectral density function was then used to select matching frequency-level bins from the tone FRA, and the firing rate of these bins averaged according to:

$$R_{vocal} = \frac{1}{N} \sum_{f=1}^{N} R_{tone}(f, A\{f\})$$

where R_{tone} is the tone-based FRA firing rate, and $A\{f\}$ is the power spectrum of a given vocal sample. Because of the higher sampling density of the vocal power-spectral function, the FRA was spline-interpolated to increase density by 10×. Only units with full FRAs (those with at least 5 sound levels tested) were included. This process was repeated for each vocalization produced, and for all vocal playback samples. Model prediction results were measured at the population level by the correlation coefficient between predicted and measured unit mean firing rates for vocalization and for playback. Unsurprisingly, predictions within individual units (i.e. predictions based upon responses to different vocalizations/samples) were found to be weak. Therefore only the prediction of the unit average response was used, and prediction accuracy was calculated at the population level (i.e. predicting which units would be more and which units less responsive to vocal production and playback). All calculations were performed separately for each class of vocalization.

3. Results

We recorded neural activities from 1603 single-units in the bilateral auditory cortices of two marmoset monkeys (Eliades and Wang, 2013). Of these units, 66% were collected from the first marmoset, the remaining 34% from the second marmoset which was recorded over a shorter time period due to other constraints. All units were studied both during self-initiated vocal production and during auditory testing (passive playback) to measure receptive field properties and responses to previously recorded vocal stimuli. Based on our previous observations (Eliades and Wang, 2003), we broadly classified responses during vocal production as either “suppressed” (RMI ≤ -0.2) or “excited” (RMI $\geq$0.2), but also examined vocal responses along a continuous axis from strongly suppressed (RMI -1) to strongly driven (RMI $+1$).

3.1. Comparison of responses during vocal production and playback

Each unit recorded during vocal production was also tested to determine its responses to passive playback of a library of vocalizations previously recorded from the same animal. The neural activities for a given type of marmoset vocalization were then compared for each unit to determine what components of vocalization-related modulation (suppression or excitation) might be explained by the passive auditory responses to vocal playback stimuli.

Fig. 1 illustrates one example unit's responses to trill vocalizations. This unit was excited during vocal production (mean vocal RMI 0.52 ± 0.36) with a strong onset response followed by sustained activity for the duration of the trill vocalizations (Fig. 1B, D). Playback of previously recorded trill vocalizations also resulted in strongly driven auditory responses (mean auditory RMI 0.84 ± 0.2), but with considerable variability between different exemplars tested (Fig. 1C and D). This pattern of responses, excited during both vocal production and playback, was characteristic for excited units.

In contrast to excited units, units suppressed during vocal production had more variable responses to playback vocalizations (Fig. 2). Some suppressed units also exhibited suppression during playback, such as the example unit in Fig. 2A–D. This unit was suppressed during trillphee vocalizations (Fig. 2B, RMI -0.79 ± 0.16) and during playback of trillphees (Fig. 2C, RMI -0.38 ± 0.37). Interestingly, the suppression during playback did not develop until later in the stimuli (Fig. 2D). Another suppressed unit (Fig. 2E–H) was suppressed during trill vocalizations (Fig. 2F, −0.46 ± 0.53), but strongly driven by playback of recorded trills (Fig. 2G, 0.46 ± 0.40).

Examining the relative prevalence of these unit populations reveals that vocalization excited units account for only 8.7% of the total samples (Table 1). Of these excited units, however, <5% were suppressed by playback vocalizations, suggesting that vocalization-related excitation is primarily an auditory response. In contrast, vocalization suppressed units made up 55% of all neurons recorded. Of these suppressed units, only 10.7% were also suppressed by playback vocalizations. Only in this small set of units might vocal suppression be a direct product of auditory tuning. About 45% of the suppressed units were driven by playback vocalizations, suggesting that vocalization-related suppression was likely induced by sources other than the ascending auditory inputs.

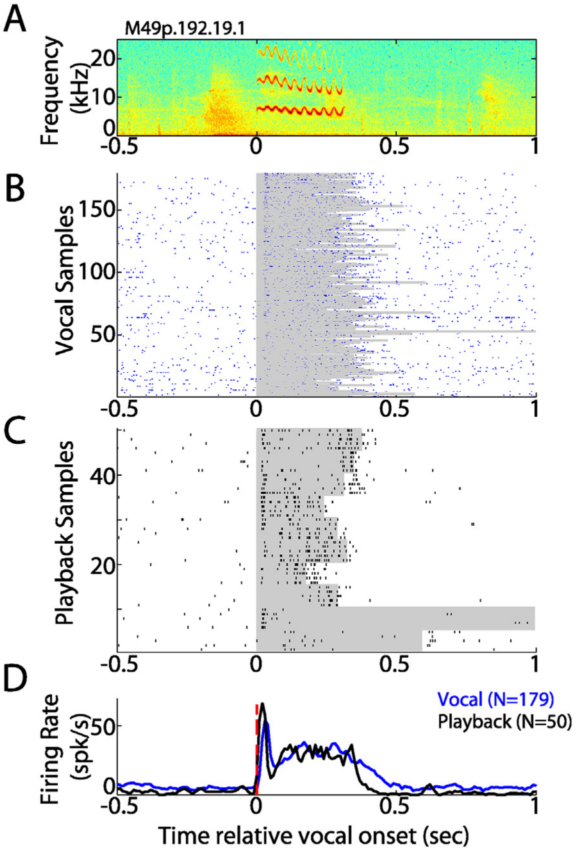

Fig. 1. Sample unit with excitatory responses during both vocal production and playback. *A*: Spectrogram of sample trill vocalization. *B*: Raster plot of unit response to produced trill vocalizations, aligned by vocal onset. *Shaded*: duration of vocalization. *C*: Raster plot of unit response to playback of trills, including phase locking to vocal oscillations for some samples. *D*: Peri-stimulus time histograms (PSTHs) for trill vocalization production (*blue*) and playback (*black*) aligned by vocal onset. This unit showed similar response to both production and playback, including onset and sustained responses. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Population comparison of vocal production and playback responses

Fig. 3 compares vocal and auditory RMIs for all tested units and each marmoset vocalization type. The results show that excited units were consistently excited by both vocal production and playback vocalizations (Fig. 3). Units with vocal RMI values near 0 had more diverse responses to playback vocalizations, but were generally biased towards driven responses. Suppressed units exhibited a greater variety of playback responses, including both driven and suppressed response. The majority of playback responses had positive auditory RMI values, indicating driven activities, regardless of the corresponding vocal responses. Overall, only 7.3% units showed suppression during playback vocalizations. There was a weak, but statistically significant correlation between vocal and auditory RMIs (phee $r = 0.13$, $p < 0.001$; trillphee $r = 0.17$, $p < 0.001$; trill $r = 0.16$, $p < 0.001$; twitter $r = 0.10$, $p < 0.05$).

In Fig. 4, we plot further analysis comparing vocal production and playback responses. For all types of vocalizations, the vocal-auditory RMI difference was biased towards negative values, indicating more suppression during vocal production compared to playback (Fig. 4A, shaded bars indicating statistically significant units). Difference values of zero indicate units with identical responses to vocal production and playback. For phee vocalizations, the average RMI difference was −0.46 ± 0.43 ($p < 0.001$, *signed-rank*). For trillphees, trills, and twitters, this difference was −0.52 ± 0.43, −0.45 ± 0.42, and −0.50 ± 0.45, respectively ($p < 0.001$, for all). Units with positive differences, indicating stronger excitation during vocalization than playback, were uncommon, particularly units with statistically significant increases (Fig. 4A, *shaded*).

Analysis of overall population vocal-auditory response differences as a function of the vocal RMI shows that the largest difference were for the units that were suppressed most, and decreasing differences in less suppressed units (Fig. 4B). This trend was present for all vocalization types independently ($p < 0.001$, *Kruskal-Wallis ANOVA*) and collectively as a group. Interestingly, excited units were the ones in which vocal and auditory responses matched most closely (difference close to zero). Another important observation was that neurons unresponsive during vocal production (vocal RMI ~0) also had negative vocal-auditory differences, ($p < 0.01$, *signed-rank*), indicating decreased vocal production responses when compared to vocal playback (relative suppression).

3.3. Vocalization responses and sound level tuning

One possible explanation for the differences between vocalization-suppressed and excited units is lower-level auditory tuning properties that are not fully captured by the responses to the playback of recorded vocalization stimuli. We therefore also examined basic auditory response properties of these units and compared results to vocalization-related activity. We first measured rate-level functions for multiple classes of stimuli, including tones, bandpass noise, wideband noise, and

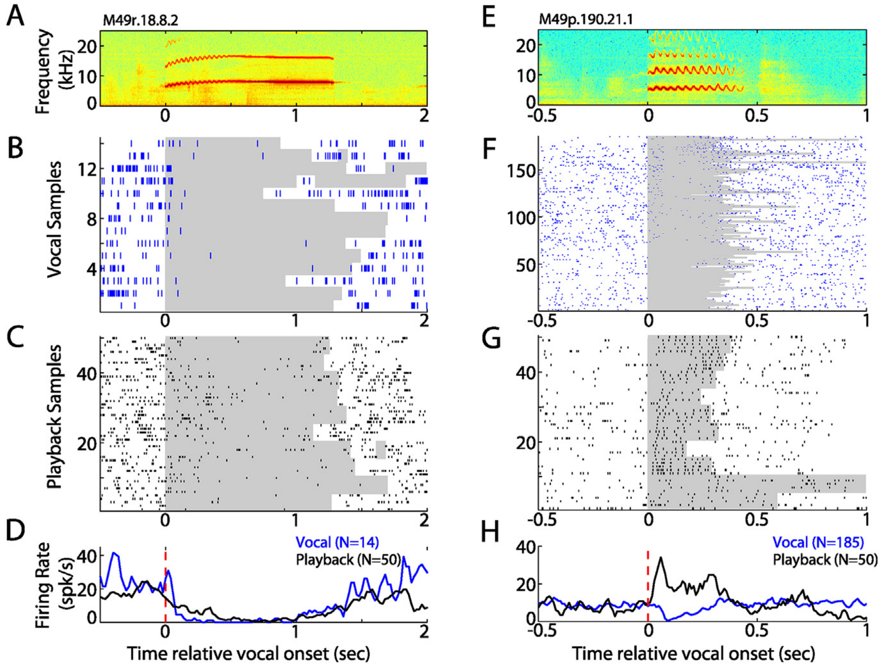

Fig. 2. Sample units suppressed during vocalization, but with different playback responses. One unit (*A-D*) was suppressed during trillphee vocal production as well as during playback (though with some delay). The second unit (*E-H*) was suppressed during trill production, but strongly driven during playback. The second type of unit was more commonly encountered than the first.

Table 1
Distribution of playback responses in suppressed and excited units.

	Suppressed (RMI ≤ −0.2)	(−0.2 to 0.2)	Excited (RMI ≥ 0.2)	*Total*
Auditory				
RMI ≥ 0.2	24.8%	18.7%	5.9%	*49.4%*
−0.2 to 0.2	24.3%	16.6%	2.4%	*43.3%*
≤−0.2	5.9%	1.1%	0.4%	*7.3%*
Total	*55.0%*	*36.3%*	*8.7%*	

vocalizations. To illustrate the dependency of vocal and playback responses on sound level, we examined the relationship between vocal and auditory RMIs on the degree of rate-level monotonicity (Fig. 5). A monotonicity index (MI, see *Methods*) was calculated for each unit based on the response to vocal playback stimuli of varying SPL. Since playback vocalization stimuli were presented at sound levels matched to those of vocal production (generally >60 dB SPL), it was not surprising that units excited by playback stimuli (auditory RMI > 0) tended to be monotonic (MI > 0.5), whereas those units suppressed by the playback stimuli (auditory RMI <0) tended to be non-monotonic (MI < 0.5), and therefore less responsive to the loud vocal playback stimuli (Fig. 5C).

An examination of the relationship between vocal production responses (vocal RMI) and monotonicity revealed a more complex relationship (Fig. 5A). As with auditory responses, units with positive vocal RMIs were biased towards monotonic units. The units with negative vocal RMIs, vocal production suppressed units, exhibited more variable MIs with both monotonic and non-monotonic playback responses. Further analysis of the interactions between vocal production and playback responses revealed that the variability of MI with vocal RMI strength was highly dependent upon the auditory response (Vocal: F = 3.06, df = 5, p < 0.05; Auditory: F = 8.69, df = 5, p < 0.001; Interaction: F = 1.81, df = 68, p < 0.001, *Kruskal-Wallis*). Specifically, units with suppressed vocal responses (especially those vocal RMI near −1), tended to be monotonic if they had excitatory vocal playback responses (positive auditory RMI), and tended to be non-monotonic if they had suppressed playback responses (Fig. 5B). These observations are consistent with the hypothesis that, while vocalization-related excitation is a product of auditory sensory tuning, vocalization-induced suppression is not due not purely due to auditory response properties of the neurons, but rather related to the act of vocal production.

3.4. Vocalization responses and frequency tuning

We next examined auditory frequency tuning, measured with either tones or bandpass noise, to determine if frequency selectivity might account for differences in vocalization responses. A few units were found with clear correlations between vocal and auditory responses, such as the unit illustrated in Fig. 6, a multi-peaked unit as has been previously described (Kadia and Wang, 2003). One frequency peak overlaps the vocalization frequency range (Fig. 6B), and another overlaps the first harmonic of vocalization frequency. We tested this unit with two trill vocalization exemplars that were shifted in mean frequency using a heterodying technique (Schuller et al., 1974). This unit's responses to these playback stimuli showed a similar spectral sensitivity profile (Fig. 6C) as the frequency tuning measured with tones (Fig. 6B) (r = −0.74, p < 0.001, between 5 and 8.5 kHz). The unit's responses during self-produced trill vocalizations (Fig. 6D) also exhibited a similar frequency-dependence as the tone tuning (r = −0.65, p < 0.001). Such units would appear to have vocalization preferences arising from their auditory tuning; however units with such clear correlations were uncommon in our sampled neural population.

We compared vocal response and frequency tuning, measured by center frequency (CF), over the whole population of tested units (Fig. 7). There was no clear relationship between CF and vocal RMI. Both vocalization-suppressed and excited units were found at CFs near or distant from the frequency range typically occupied by the

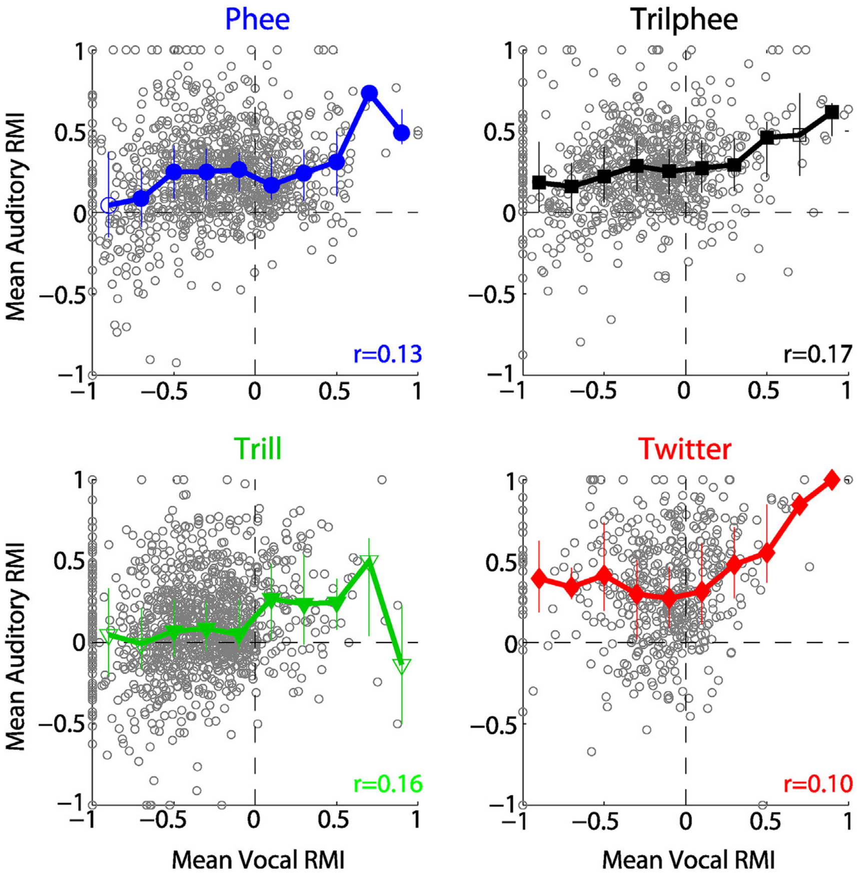

Fig. 3. Population comparison of vocalization and playback responses. Vocalization (“Vocal RMI”) and playback (“Auditory RMI”) responses were quantified by a normalized RMI measure and averaged for each unit. Comparisons are plotted individually for the four most common marmoset vocalization classes: phee (*blue, top left*), trillphee (*black, top right*), trill (*green, bottom left*), and twitter (*red, bottom right*). Plotted curves indicate mean auditory RMI for units binned by their vocal RMI. Vocal RMIs <0 indicate suppression during vocalization, and RMIs>0 indicate excitation. Auditory responses were distributed around zero for suppressed units and increased with vocal responses. *Error bars*: bootstrapped 95% confidence intervals. *Filled symbols*: statistically significant deviations from 0 ($p < 0.01$. *signed-rank*). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

first two spectral components of marmoset vocalizations (marked by grey bars).

It is likely, however, that using CF alone would under-represent the complexity of spectral tuning of auditory cortex units. Units varied widely in the tuning-width around the CF peak, and often had multiple frequency peaks (i.e. Fig. 6), both of which could affect the response to vocalizations. Marmoset vocalizations also typically consist of at least 2 harmonics of the fundamental frequency, any one of which could interact with a unit's frequency receptive field, possibly contributing to these inconsistencies. To explore these factors, we computed the fraction of units with firing rates ≥80% of the maximum frequency tuning peak within 1/2 octave of vocalization mean frequency or one of its harmonics (Fig. 8). Such units would be expected to have significant overlap of vocal spectral energy and the tone/noise tuning curve. In general, at least half of auditory cortex units met this criterion. Such overlaps were more prevalent amongst units excited by vocal production (Fig. 8A–B), although, again, the relationship to vocal playback was less clear (Fig. 8C). Even amongst units suppressed during vocalization, about half of the units had this proximity between frequency tuning peaks and vocal mean frequency, suggesting that frequency tuning also cannot fully account for vocalization-induced suppression, except in a subset of units.

3.5. Model prediction of vocal responses

Because marmoset vocalizations contain multiple acoustic components, and receptive fields can be quite complex, explanation of vocal responses based upon sound level or frequency tuning in isolation are likely poor estimates. We therefore constructed a simple linear model (further details in *Methods*) to predict both vocalization and playback responses based on the approach from a previous study (Bar-Yosef et al., 2002). We measured tone-based frequency response areas (FRA) for each of 334 units. The spectral content for each vocal sample was determined using power-spectral density calculations and projected onto the FRA response (Fig. 9A). The firing rates of congruent frequency-level bins were then averaged to estimate the mean rate response to the vocalization. This was repeated for all recorded self-produced vocalizations and playback vocal stimuli (the same samples and matched loudness used above to estimate the auditory RMI). Comparison of mean unit responses and model predictions provides an estimate of the model's ability to predict the average degree of vocal suppression or excitation for the unit.

Overall, this model provided a reasonable population-level prediction ($r = 0.55$, $p < 0.001$) of mean unit responses to vocal playback (Fig. 9B). The prediction for vocal production was much poorer ($r = 0.18$, $p < 0.001$). Further examination of model

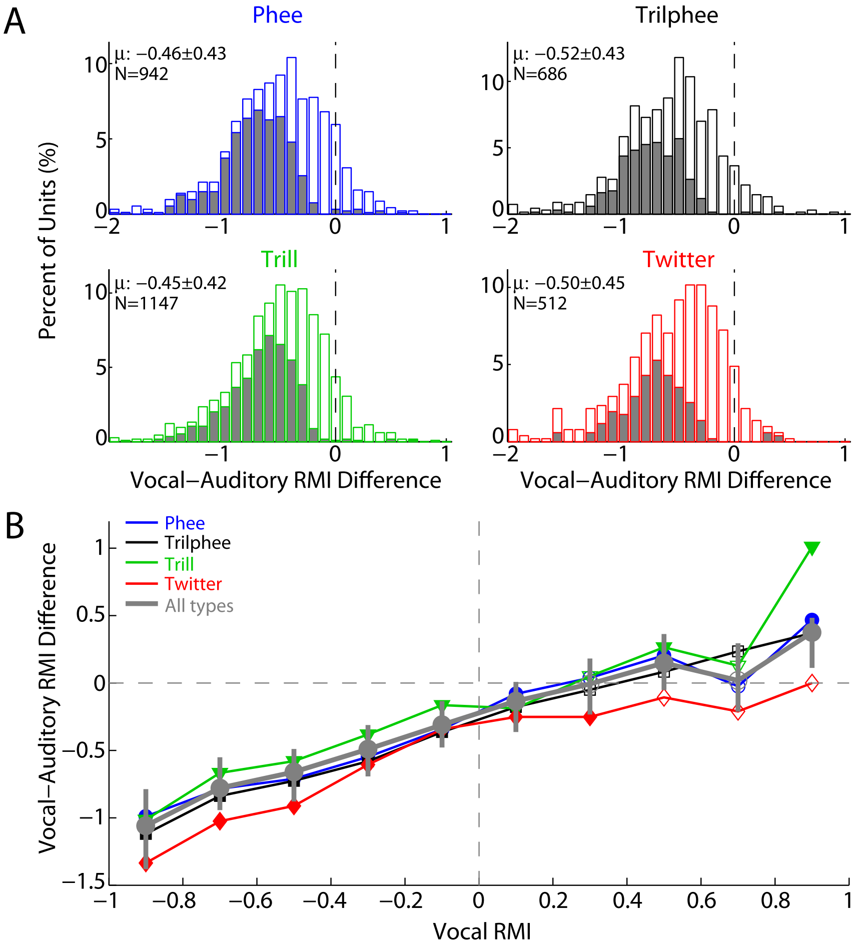

Fig. 4. Distribution of vocal-auditory differences. *A*: Histograms are plotted showing the distribution of RMI differences between vocalization and playback (vocal - auditory) for each unit. Most units showed large shifts towards negative values, indicating suppression. *Shaded bars*: units with statistically significant differences between vocal production and playback (p < 0.05, *ranksum*). *B*: Plot of mean vocal-auditory differences for units binned by vocal RMI. Differences were nearly zero for excited units, indicating matched vocal and auditory responses. Differences were negative for units with vocal RMI near zero, indicating that these vocal "unresponsive" units were actually suppressed compared to playback Colors indicate vocalization types as in *A*. *Grey*: average response including all vocalization types. Error bars: bootstrapped 95% confidence intervals. *Filled symbols*: statistically significant deviations from 0 (p < 0.01. *signed-rank*). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

predictions for different units showed that the model rarely predicted strong suppression or inhibition, either for vocalization or playback, a likely source of the poor prediction of vocal production results. The model tended to underestimate the playback responses to trills and twitters (linear regression slopes 0.35 [95% CI: 0.31,0.38] and 0.31 [0.27 0.34]) more so than phees and trillphees (0.56 [0.45 0.67] and 0.43 [0.32 0.53]), particularly for strongly driven activity. Overall, however, the accuracy of the prediction was surprisingly good for playback of individual vocal types (phee r = 0.49, trillphee r = 0.41, trill r = 0.74, twitter r = 0.75; all p < 0.001). Predictions were poorer during vocal production, as expected given the rarity of predicted suppressed responses (phee r = −0.08, trillphee r = −0.08, p > 0.05; trill r = 0.31, twitter r = 0.47, p < 0.001).

We further measured model predictions by grouping units according to their vocal and playback responses and examining the predictions for each sub-group (Fig. 10). Model playback predictions were strongest for units with strongly driven responses (auditory RMI near +1), and weaker for units with less driven auditory responses (Fig. 10A). Interestingly, the model did a reasonable job for some suppressed units, particularly those with matched auditory and vocal RMIs, suggesting the model could account for some of sensory-related inhibition during vocal playback. Multiple linear regression confirmed increased predictions with excitation (r = 0.57, F = 3.88, p < 0.05), with a stronger dependence on playback response strength (coefficient 0.67, [95% CI: 0.15 1.20]) than for production (−0.26, [CI -0.69 0.18]).

Model predictions of vocal production responses (Fig. 10B) showed poor (negative) correlation for suppressed units, but good predictions (r > 0.5) for most excited units. Linear regression again confirms improved prediction with excitation (r = 0.63, F = 5.38, p < 0.05), with a similar dependence on playback responses (0.67, [0.18 1.17]), and modest improvement in dependence upon vocal responses (0.16, [-0.24 0.57]). This close match between predictions for both playback and production in excited units again suggests that such responses were a result of tuned ascending auditory inputs. In contrast, significantly poorer and even inverse predictions for suppressed units were present.

We also examined the performance of the model in predicting responses to individual vocalization samples for each unit. For playback responses, prediction correlation coefficients varied widely from −1 to 1, but the average correlation across units was weak (0.06 ± 0.34; p < 0.01 *signed-rank*). The predictions for vocal production responses were even poorer (0 ± 0.4; p > 0.05). These

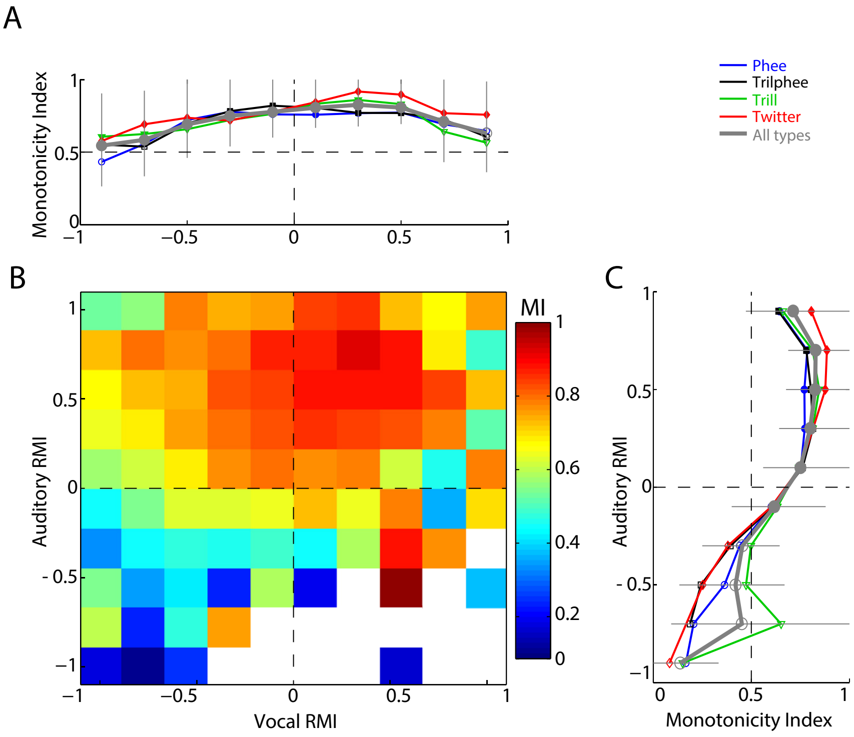

Fig. 5. Effect of rate-level tuning on vocalization and playback responses. *A*: Average monotonicity index (MI) for units grouped by vocal RMI, showing equal monotonic and non-monotonic units for suppression and increasing MI with vocal excitation. Error bars: 95% confidence intervals. B: Two-dimensional plot of mean MI grouped by both vocal and auditory RMI, showing the largest MIs for units excited by both vocalization and playback and smallest MIs for units suppressed by both. Color bar (*right*) indicates the MI scale. C: Average MI grouped by auditory RMI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

results suggest that, while a simple model can reasonably predict which units will be excited or suppressed by playback vocalizations, and to a lesser extent vocal production, the model cannot predict how a unit will respond to the varying acoustics of individual vocalizations.

3.6. *Subcortical contribution to model prediction of playback-production differences*

Previous work has suggested that subcortical attenuation, in particular a combination of middle ear reflexes (Carmel and Starr, 1963; Henson, 1965; Salomon and Starr, 1963; Suga and Jen, 1975) and brainstem-level attenuation (Papanicolaou et al., 1986; Suga and Schlegel, 1972; Suga and Shimozawa, 1974) may also be present during vocal production, but not during passive listening. Such attenuation has been estimated between 20 and 40 dB SPL (Suga and Shimozawa, 1974) and might bias model estimates of vocalization responses. We therefore repeated model calculations, factoring in varying degrees of sub-cortical loudness attenuation (0–60 dB). Overall model performance decreased with increasing attenuation, from $r = 0.18$ for un-attenuated, to 0.06 and −0.07 for 20 and 40 dB, respectively. Specific examination of excited units ($RMI \geq 0.2$), those with the best model accuracy, also showed reduced performance from $r = 0.71$ to 0.52 and 0.12. Suppressed units did not exhibit any changes in performance ($r = -0.41, -0.40$, and -0.34). These results suggest that presumed sub-cortical sources of attenuation do not account for differences between vocal production-related activity and model estimates from sensory receptive fields.

3.7. *Influence of cortical location on vocal responses*

We further examined the effects of recording location on neural responses during vocal production. The recording arrays contained four rows of electrodes, with the medial two rows generally falling within primary auditory cortex (A1), the third row on lateral belt (LB), and the fourth row on parabelt (PB) areas. We first compared the prevalence of suppression and excitation by electrode row, and found generally similar proportions of suppressed (medial to lateral: 65.0%, 59.1%, 46.2%, and 52.5%; overall: 55%) and excited (8.6%, 6.6%, 11.4% 10.1%; overall: 8.7%) units. There was a general trend towards more suppressed units in medial (A1) electrodes, and more excited units in lateral (LB/PB) electrodes. We further examined the overall magnitude of the vocal RMI by electrode row, these distributions were highly overlapping, but with pattern of increased suppression in medial over lateral rows (mean RMI: −0.38 ± 0.38, −0.32 ± 0.29, −0.21 ± 0.27, −0.26 ± 0.30) that was statistically significant ($p < 0.001$, *Kruskall-Wallis*).

Examination of all units' playback responses also revealed stronger auditory RMIs in more medial electrodes (medial to lateral: 0.25 ± 0.34, 0.15 ± 0.26, 0.15 ± 0.24. 0.17 ± 26; $p < 0.001$, *Kruskall-Wallis*). However, when only vocal suppressed units (vocal RMI < -0.2) were examined, average auditory RMIs were not different between electrode rows (0.16 ± 0.37, 0.11 ± 0.30, 0.15 ± 0.27, 0.15 ± 0.30; $p = 0.36$, *Kruskall-Wallis*). These results raise the possibility that some of the apparent differences in vocal responses between electrodes may have been due to differences in their passive auditory responses. Given that different cortical areas were not sampled at matched positions along the tonotopic axis in these experiments, it is difficult to disambiguate this confound or make strong claims about the role of different auditory cortical

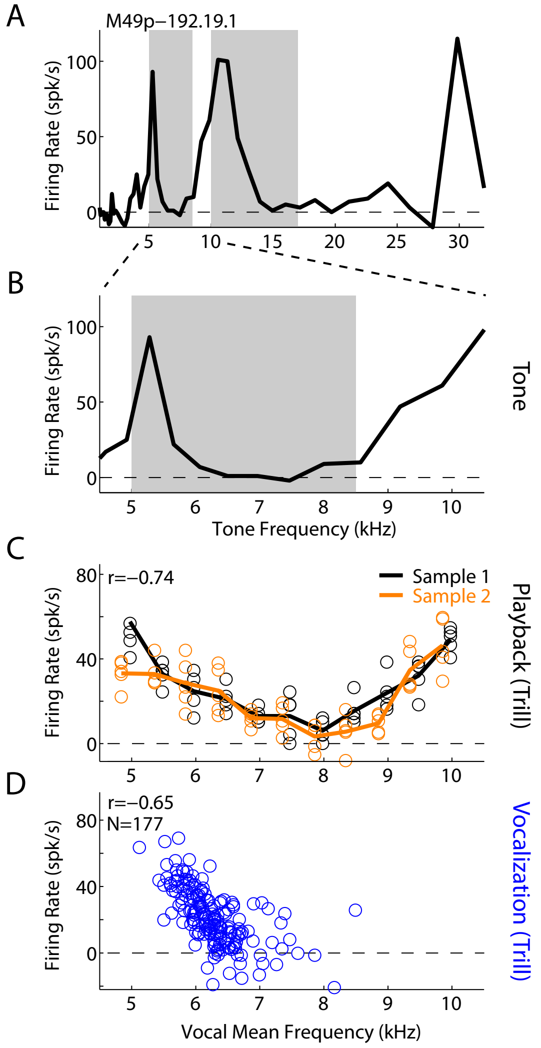

Fig. 6. Sample unit frequency tuning and vocal responses. *A*: Tone frequency tuning curve exhibiting multi-peaked frequency tuning. *Shaded*: range of fundamental frequency and first harmonic for trill vocalizations recorded during the same testing session. *B*: Expansion of the frequency tuning curve focusing on the vocal range. *C*: Responses to individual samples (*circles*) and mean response (*line*) to the playback of two trill exemplars (*orange, black*) acoustically modified to sample a range of mean frequencies. Trill frequency tuning qualitatively reflects tone-based tuning in *B*. *D*: Firing rates are plotted against the mean frequency of self-produced trill vocalizations. Correlation coefficients of responses with mean vocal frequency are indicated. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

areas in vocal responses with the present data set.

4. Discussion

We examined the activities of a large number of single-units in the marmoset auditory cortex and compared the responses during self-produced vocalizations, basic auditory tuning, and responses to playback of recorded vocal sounds. We found that (1) neurons excited during vocal production were almost always excited by playback vocalizations, while (2) neurons suppressed by vocal production had more diverse playback responses, though generally favoring playback excitation. (3) Neurons excited by either playback or vocal production tended to have monotonic rate-level functions, while neurons suppressed by both tended to be non-monotonic. (4) Frequency tuning and frequency-based models predict playback responses more accurately than responses during vocal production, but generally fail to explain vocalization-induced suppression. These findings further our understanding of auditory-vocal mechanisms in the auditory cortex, and begin to explain some of the diverse neural activities that have been observed during vocal production.

4.1. Comparison with previous results

In our previous investigations, we failed to find a relationship between vocal production-related neural responses and sensory tuning of auditory cortex neurons. In particular, we noted that CF, threshold, and monotonicity did not predict the behavior of an auditory cortex neuron during vocalization and that many suppressed neurons would respond to playback of previously recorded (conspecific) vocalizations (Eliades and Wang, 2003, 2013). We also noted that the variation of vocalization responses with vocal acoustics (Eliades and Wang, 2005) or altered feedback (Eliades and Wang, 2008a, 2012) was seemingly unrelated to auditory tuning. These previous studies were limited, however, by examining only simple frequency tuning parameters such as CF and the responses to a limited set of vocal playback stimuli.

The new analyses conducted in the present work, as well as the inclusion of additional auditory responses in our analyses, provide new insights beyond our previous work. In contrast to our previous findings, here we demonstrate that neurons with vocalization-related excitation are nearly universally responsive to vocal playback (Figs. 1 and 3) and have mostly monotonic rate-level tuning to vocal playback (Fig. 5). Additionally, we examined frequency tuning properties besides CF, which allowed us to take into account tuning bandwidth, multiple frequency peaks, and vocal harmonics, and a large overlap between vocal acoustics and frequency-tuning (Fig. 8) as well as high degree of predictability of vocal responses for both production and playback based upon a frequency-response area model (Figs. 9 and 10). These results suggest that vocalization-related excitation during vocal production is largely, if not entirely, a sensory phenomenon. Since such neurons do not appear to be biased by vocal production, they may provide a mechanism for encoding outside sounds during vocalization.

In contrast, the results of the present study confirm our earlier finding that vocalization-induced suppression cannot be predicted based upon a neuron's auditory responses (Eliades and Wang, 2003) (Figs. 2–3). This finding is consistent with the prevailing theory of vocal suppression arising from internal modulatory signals, as we further discuss below. One interesting exception is the presence of a small subset of suppressed neurons that were also suppressed by playback stimuli (~10%), suggesting that the suppression was not entirely caused by motor signals in these neurons (Fig. 3). Such neurons may be a significant contaminant of our previous analyses. For example, our previous results showing sensitivity to altered vocal feedback in suppressed neurons also found decreased sensitivity for the most strongly suppressed neurons (Eliades and Wang, 2008a). If many of these maximally suppressed neurons were driven by sensory instead of sensory-motor processes, it can explain why our previous work did not observe a relationship between vocal production-related neural responses and sensory tuning of auditory cortex neurons.

Another novel finding in these analyses were a significant number of neurons whose vocal responses were reduced compared to vocal playback responses, but not significantly below spontaneous activity (Fig. 4). Under our previous definition of vocalization-induced suppression, these neurons were not classified as suppressed neurons. However, previous human studies have used a similar production-playback comparison measure to

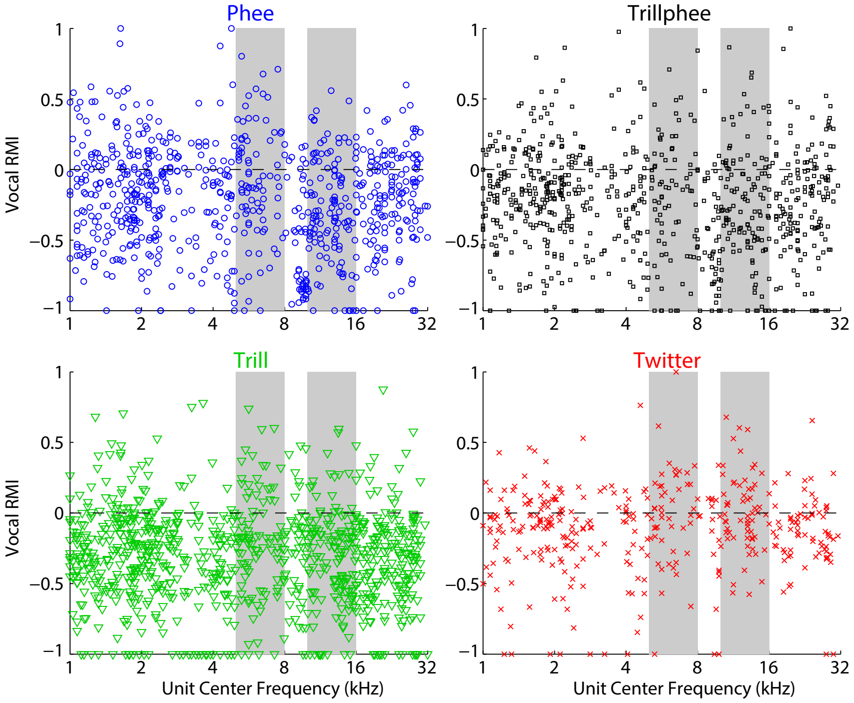

Fig. 7. Comparison of vocal response and unit center frequency. Scatter plots show unit mean vocal production response (RMI) against CF measured from either tone or bandpass noise tuning. No clear relationship is evident. *Shaded*: range of vocalization mean fundamental frequency and first harmonic.

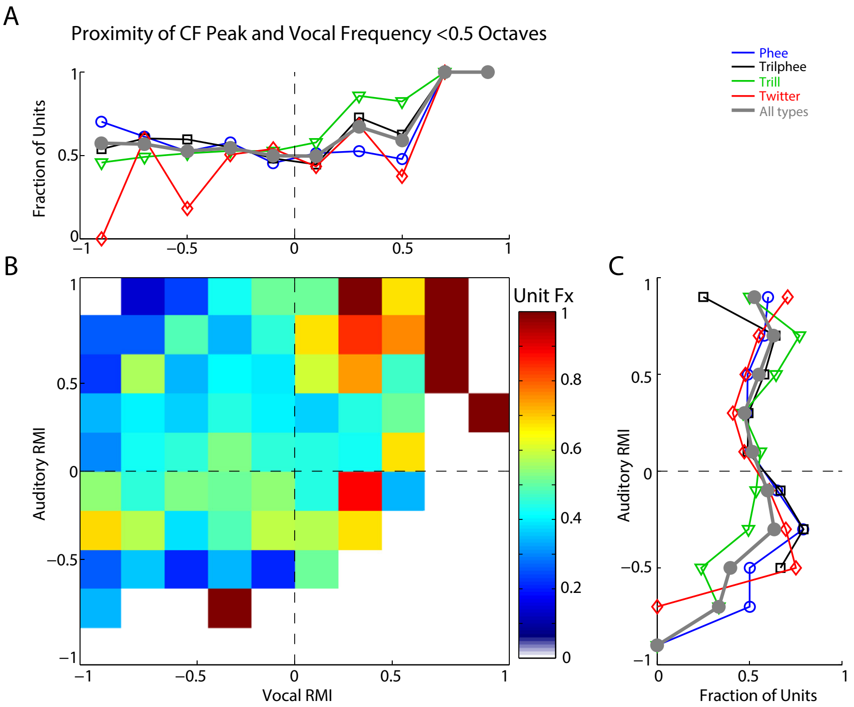

Fig. 8. Effect of frequency tuning on vocalization and playback responses. The distance between vocal frequency (either fundamental or any harmonic) and the nearest frequency peak was measured. A: Fraction of units with a CF peak-vocal frequency difference <0.5 octaves is plotted against vocal RMI. Unit fraction with close proximity was relatively constant except for units excited by vocalization, where it was increased. B: Two-dimensional histogram of unit fraction grouped by both vocal and auditory RMI, showing the largest overlapping fraction for units excited by both vocalization and playback. Colorbar (*right*) indicates the unit fraction scale. D: Unit fraction grouped by auditory RMI. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

establish speech-induced suppression (i.e. Houde et al., 2002; Chang et al., 2013). Some differences in results (i.e. the relationship between suppression and altered feedback effects) between marmoset and human experiments may be in part attributable to these differing definitions of vocalization-induced suppression. Further work will need to take into account both possible

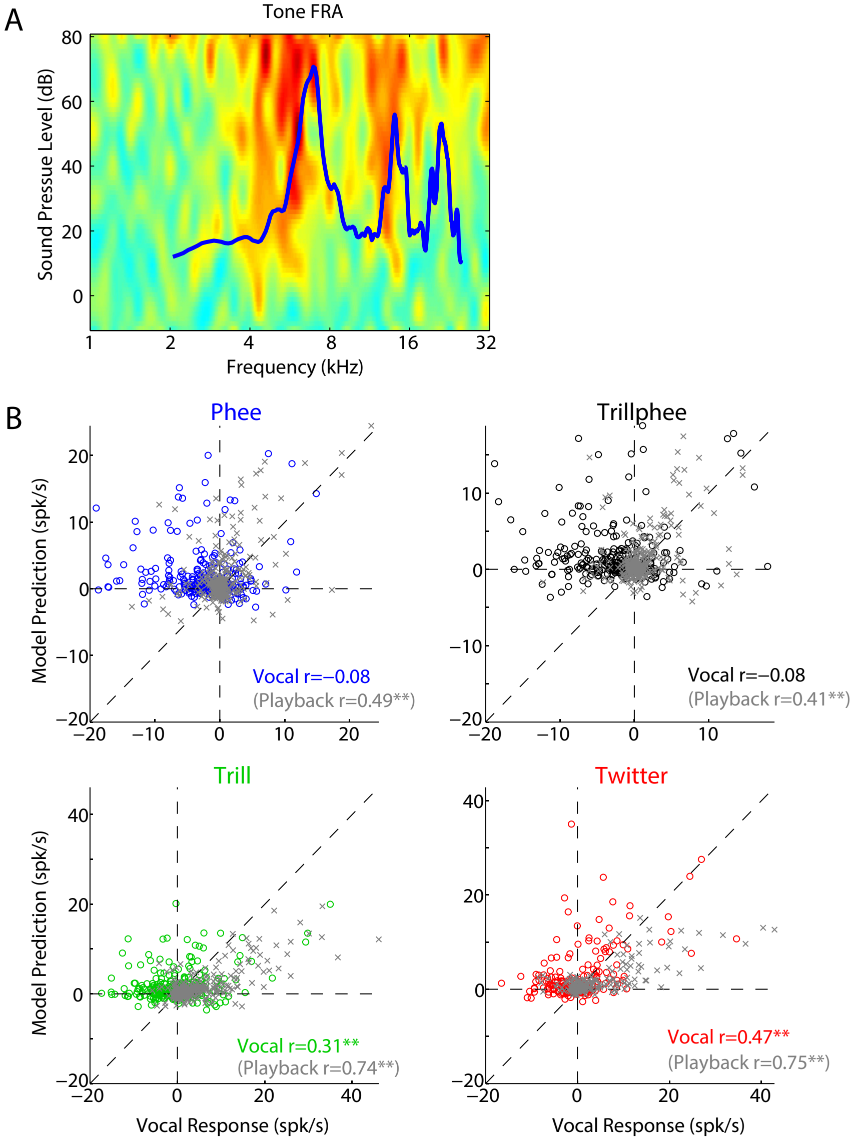

Fig. 9. Linear frequency response area model of mean vocalization responses. *A*: Illustration of the FRA-based model. A smoothed tone-measured FRA is shown overlaid with the power-spectral density function from a sample phee vocalization. The firing rate response from the overlapping bins was averaged to calculate the model prediction. *B*: Scatter plots comparing measured unit mean firing rates during vocal production and model predictions for all four major vocalization types (*colored*). Comparisons between units' mean playback responses and model predictions are shown in *grey*. Model predictions were better for playback than during vocal production, particularly for phee and trillphee vocalizations. (**p < 0.001). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

definitions of vocal suppression in order to better reconcile results based on single-neuron recording in animals with those based on surface potential recordings and imaging in humans.

4.2. Modeling results

As part of these analyses, we also constructed a simple linear model to predict vocal responses for both vocal production and auditory playback based upon pure-tone FRA responses. This simple model, based upon Bar-Yosef et al. (2002) has an appeal in its ability to simultaneously integrate for multiple aspects of a unit's receptive field (center frequency, bandwidth, multiple frequency peaks, amplitude tuning) as well as the multiple harmonic components of marmoset vocalizations. Given the simplicity of the model, it was surprising the degree to which it was able to predict auditory playback responses (correlation coefficients of 0.41–0.75), although it did not perform well in predicting vocal production responses. One of the limitations of this approach is that it is well known that such predictive models are highly dependent on the types of stimuli used to make the prediction, such as artificial vs. natural stimuli (Laudanski et al., 2012). Additionally, linear models often fail to fully capture important non-linear interactions between frequency components in auditory receptive fields (Young et al., 2005). Our model also fails to capture any sensitivity to temporal or spectro-temporal information which may also be important (Theunissen et al., 2000). Despite these limitations, the observation of significantly better model predictions of auditory playback responses than for responses during the production of *acoustically similar* vocalizations is consistent with the notion that non-auditory inputs contribute to vocalization-induced suppression.

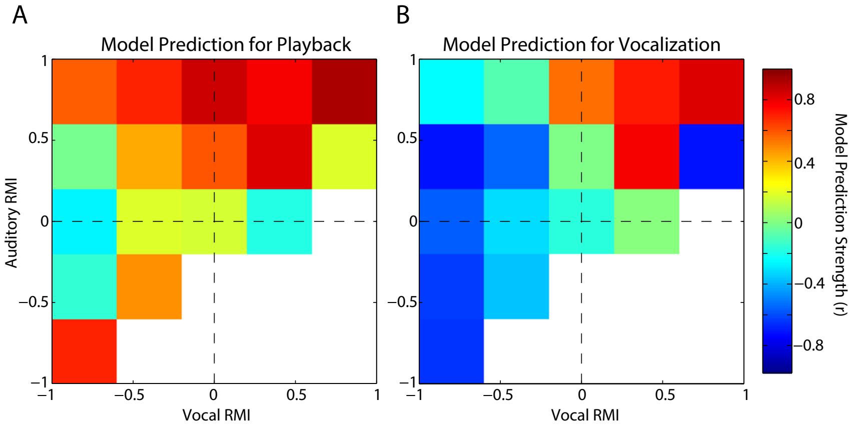

Fig. 10. Comparison of model prediction accuracy with vocalization and playback responses. Model predictions of unit mean responses were grouped by unit vocal and auditory RMIs and the prediction accuracy for each group measured by a correlation coefficient. Two-dimensional plots of the accuracy are shown separately for predictions of vocal playback (*A*) and production (*B*) responses. Colorbar (*right*) indicates the correlation scale. Predications were better for playback than vocalization. Predictions were also stronger for units excited by playback and/or vocal production, and weaker for vocalization-suppressed units. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.3. *Cortical location and auditory-vocal interaction*

We examined the strength and prevalence of vocalization-induced suppression in different auditory cortical fields. Marmoset auditory cortex is structured with a core-belt-parabelt organization common to non-human primates (de la Mothe et al., 2006; Hackett et al., 2001). Vocal response distributions were largely overlapping between more medial electrodes (A1), and more lateral ones (presumed lateral belt & parabelt areas). There were, however, statistically significant trends towards both stronger playback responses and stronger vocal suppression in more medial electrodes. However, these observations are possibly confounded by variations in the tonotopic locations of electrodes between different cortical areas. Further work with more extensive and matched spatial/spectral sampling of units will be needed in order to reveal the role of different auditory cortical areas in vocal suppression and auditory-vocal interaction.

4.4. *Mechanism of auditory-vocal integration*

One important question that remains unanswered is the mechanism by which a suppressive vocal motor signal is combined with auditory feedback signal at the level of an individual neuron in auditory cortex. The absence of correlation between passive auditory tuning and vocal responses for suppressed neurons shown in the present study suggests that vocalization-induced suppression is more complicated than a simple linear additive process (e.g., excitatory vocal feedback response being added to a static vocal motor inhibition). This is in contrast to the observed linear prediction of vocalization-related excitation based upon both vocal playback and pure tone responses. Several competing models can be posited to potentially explain the incongruent responses for vocal suppression.

The first is that vocalization-induced suppression represents an error signal (Behroozmand and Larson, 2011; Houde and Nagarajan 2011; Niziolek et al., 2013). In this model, suppressed vocal responses reflect a direct subtraction of expected (efferent copy) sensory input, with maximal suppression resulting from a perfect match between vocal feedback and the expected signal (i.e. no feedback error). The effects of such efferent signals, also termed corollary discharges, have long been studied in various model systems (Crapse and Sommer, 2008; Sperry, 1950; von Holst and Mittelstaedt, 1950). Recent results using human MEG studies are consistent with this model, where natural speech fluctuations in vowel formant frequencies were found to evoke increased auditory cortical activity compared to vowels closer to the mean (Niziolek et al., 2013).

Additionally, recent work on motor efference in non-vocalizing rodents may also be consistent with this model. Optogenetic techniques have demonstrated a direct neural pathway for motor-induced suppression of auditory cortex during locomotion (Nelson et al., 2013; Schneider et al., 2014; Schneider and Mooney, 2015a). This pathway appears to provide multiple inputs to the auditory cortex, including from both motor cortex and the basal forebrain. When locomotion is paired with an expected sound, there is a suppression of stimulus-evoked activity to similar tone frequencies, but not to tones of more distant frequencies, suggesting a subtractive error comparison (Schneider and Mooney, 2015b; Nelson and Mooney, 2016). Whether or not a similar mechanism is active during vocal production remains an open question.

A second possible explanatory model is that efferent copy signals bias the receptive fields of auditory cortex neurons to better encode vocalization feedback. A selective scaling model of sensory tuning, as has been described for attention (Fritz et al., 2007), is one possibility. Another is a wholesale shift in receptive fields as has been described in parietal cortex during saccades (Duhamel et al., 1992). Some recent evidence has emerged that auditory cortex receptive fields can change dynamically with behavioral tasks (Fritz et al., 2005), and such changes are likely under the control of frontal cortex (Fritz et al., 2010). Which of these models might best explain the auditory-vocal integration observed in primate auditory cortex remains unanswered. However it should also be noted that they are not necessarily mutually exclusive. Future work will more directly test these models to determine the functional mechanism of auditory-vocal interaction and integration.

Grants

This work was supported by National Institute on Deafness and Other Communication Disorders Grants DC-005808 (X.W.) and DC-014299 (S.J.E).

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

Author contributions

S.J.E. and X.W. shared conception and design of research, interpretation of results, manuscript editing and revision, and approval of the final version of this manuscript; S.J.E. performed experiments, analyzed data, prepared figures, and drafted the manuscript.

Acknowledgements

The authors thank A. Pistorio for assistance in animal care and training, and C. Miller for helpful feedback on this manuscript.

References

Agamaite, J.A., Chang, C.J., Osmanski, M.S., Wang, X., 2015. A quantitative acoustic analysis of the vocal repertoire of the common marmoset (*Callithrix jacchus*). J. Acoust. Soc. Am. 138, 2906–2928.

Agnew, Z.K., McGettigan, C., Banks, B., Scott, S.K., 2013. Articulatory movements modulate auditory responses to speech. Neuroimage 73, 191–199.

Bar-Yosef, O., Rotman, Y., Nelken, I., 2002. Responses of neurons in cat primary auditory cortex to bird chirps: effects of temporal and spectral context. J. Neurosci. 22, 8619–8632.

Bauer, E.E., Klug, A., Pollak, G.D., 2002. Spectral determination of responses to species-specific calls in the dorsal nucleus of the lateral lemniscus. J. Neurophysiol. 88, 1955–1967.

Behroozmand, R., Larson, C.R., 2011. Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. BMC Neurosci. 12, 54.

Behroozmand, R., Oya, H., Nourski, K.V., Kawasaki, H., Larson, C.R., Brugge, J.F., Howard 3rd, M.A., Greenlee, J.D., 2016. Neural correlates of vocal production and motor control in human Heschl's gyrus. J. Neurosci. 36, 2302–2315.

Bekesy, G.v, 1949. The structure of the middle ear and the hearing of one's own voice by bone conduction. J. Accoust Soc. Am. 21, 217–232.

Brumm, H., Voss, K., Kollmer, I., Todt, D., 2004. Acoustic communication in noise: regulation of call characteristics in a New World monkey. J. Exp. Biol. 207, 443–448.

Burnett, T.A., Freedland, M.B., Larson, C.R., Hain, T.C., 1998. Voice F0 responses to manipulations in pitch feedback. J. Acoust. Soc. Am. 103, 3153–3161.

Carmel, P.W., Starr, A., 1963. Acoustic and nonacoustic factors modifying middle ear muscle activity in waking cats. J. Neurophysiol. 26, 598–616.

Chang, E.F., Niziolek, C.A., Knight, R.T., Nagarajan, S.S., Houde, J.F., 2013. Human cortical sensorimotor network underlying feedback control of vocal pitch. Proc. Natl. Acad. Sci. U. S. A. 110, 2653–2658.

Christoffels, I.K., Formisano, E., Schiller, N.O., 2007. Neural correlates of verbal feedback processing: an fMRI study employing overt speech. Hum. Brain Mapp. 28, 868–879.

Crapse, T.B., Sommer, M.A., 2008. Corollary discharge across the animal kingdom. Nat. Rev. Neurosci. 9, 587–600.

Crone, N.E., Hao, L., Hart Jr., J., Boatman, D., Lesser, R.P., Irizarry, R., Gordon, B., 2001. Electrocorticographic gamma activity during word production in spoken and sign language. Neurology 57, 2045–2053.

Curio, G., Neuloh, G., Numminen, J., Jousmaki, V., Hari, R., 2000. Speaking modifies voice-evoked activity in the human auditory cortex. Hum. Brain Mapp. 9, 183–191.

de la Mothe, L.A., Blumell, S., Kajikawa, Y., Hackett, T.A., 2006. Cortical connections of the auditory cortex in marmoset monkeys: core and medial belt regions. J. Comp. Neurol. 496, 27–71.

Duhamel, J.R., Colby, C.L., Goldberg, M.E., 1992. The updating of the representation of visual space in parietal cortex by intended eye movements. Science 255, 90–92.

Eliades, S.J., Wang, X., 2003. Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. J. Neurophysiol. 89, 2194–2207.

Eliades, S.J., Wang, X., 2005. Dynamics of auditory-vocal interaction in monkey auditory cortex. Cereb. Cortex 15, 1510–1523.

Eliades, S.J., Wang, X., 2008a. Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature 453, 1102–1106.

Eliades, S.J., Wang, X., 2008b. Chronic multi-electrode neural recording in free-roaming monkeys. J. Neurosci. Methods 172, 201–214.

Eliades, S.J., Wang, X., 2012. Neural correlates of the lombard effect in primate auditory cortex. J. Neurosci. 32, 10737–10748.

Eliades, S.J., Wang, X., 2013. Comparison of auditory-vocal interactions across multiple types of vocalizations in marmoset auditory cortex. J. Neurophysiol. 109, 1638–1657.

Epple, G., 1968. Comparative studies on vocalization in marmoset monkeys (*Hapalidae*). Folia Primatol. 8, 1–40.

Flinker, A., Chang, E.F., Kirsch, H.E., Barbaro, N.M., Crone, N.E., Knight, R.T., 2010. Single-trial speech suppression of auditory cortex activity in humans. J. Neurosci. 30, 16643–16650.

Fritz, J.B., Elhilali, M., Shamma, S.A., 2005. Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. J. Neurosci. 25, 7623–7635.

Fritz, J.B., Elhilali, M., David, S.V., Shamma, S.A., 2007. Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in A1? Hear Res. 229, 186–203.

Fritz, J.B., David, S.V., Radtke-Schuller, S., Yin, P., Shamma, S.A., 2010. Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. Nat. Neurosci. 13, 1011–1019.

Greenlee, J.D., Jackson, A.W., Chen, F., Larson, C.R., Oya, H., Kawasaki, H., Chen, H., Howard, M.A., 2011. Human auditory cortical activation during self-vocalization. PLoS One 6, e14744.

Hackett, T.A., Preuss, T.M., Kaas, J.H., 2001. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. J. Comp. Neurol. 441, 197–222.

Heinks-Maldonado, T.H., Mathalon, D.H., Gray, M., Ford, J.M., 2005. Fine-tuning of auditory cortex during speech production. Psychophysiology 42, 180–190.

Henson Jr., O.W., 1965. The activity and function of the middle-ear muscles in echo-locating bats. J. Physiol. 180, 871–887.

Hickok, G., Houde, J., Rong, F., 2011. Sensorimotor integration in speech processing: computational basis and neural organization. Neuron 69, 407–422.

Holmstrom, L.A., Eeuwes, L.B., Roberts, P.D., Portfors, C.V., 2010. Efficient encoding of vocalizations in the auditory midbrain. J. Neurosci. 30, 802–819.

Houde, J.F., Jordan, M.I., 1998. Sensorimotor adaptation in speech production. Science 279, 1213–1216.

Houde, J.F., Nagarajan, S.S., 2011. Speech production as state feedback control. Front. Hum. Neurosci. 5, 82.

Houde, J.F., Nagarajan, S.S., Sekihara, K., Merzenich, M.M., 2002. Modulation of the auditory cortex during speech: an MEG study. J. Cog. Neurosci. 14, 1125–1138.

Kadia, S.C., Wang, X., 2003. Spectral integration in A1 of awake primates: neurons with single- and multipeaked tuning characteristics. J. Neurophysiol. 89, 1603–1622.

Lane, H., Tranel, B., 1971. The Lombard sign and the role of hearing in speech. J. Speech Hear Res. 14, 677–709.

Laudanski, J., Edeline, J.M., Huetz, C., 2012. Differences between spectro-temporal receptive fields derived from artificial and natural stimuli in the auditory cortex. PLoS One 7, e50539.

Lee, B.S., 1950. Effects of delayed speech feedback. J. Acoust. Soc. Am. 22, 824–826.

Leonardo, A., Konishi, M., 1999. Decrystallization of adult birdsong by perturbation of auditory feedback. Nature 399, 466–470.

Levelt, W.J., 1983. Monitoring and self-repair in speech. Cognition 14, 41–104.

Lu, T., Liang, L., Wang, X., 2001. Neural representations of temporally asymmetric stimuli in the auditory cortex of awake primates. J. Neurophysiol. 85, 2364–2380.

Martikainen, M.H., Kaneko, K.I., Hari, R., 2005. Suppressed responses to self-triggered sounds in the human auditory cortex. Cereb. Cortex 15, 299–302.

Miller, C.T., Wang, X., 2006. Sensory-motor interactions modulate a primate vocal behavior: antiphonal calling in common marmosets. J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. 192, 27–38.

Nelson, A., Mooney, R., 2016. The basal forebrain and motor cortex provide convergent yet distinct movement-related inputs to the auditory cortex. Neuron 90, 635–648.

Nelson, A., Schneider, D.M., Takatoh, J., Sakurai, K., Wang, F., Mooney, R., 2013. A circuit for motor cortical modulation of auditory cortical activity. J. Neurosci. 33, 14342–14353.

Niziolek, C.A., Nagarajan, S.S., Houde, J.F., 2013. What does motor efference copy represent? Evidence from speech production. J. Neurosci. 33, 16110–16116.

Papanicolaou, A.C., Raz, N., Loring, D.W., Eisenberg, H.M., 1986. Brain stem evoked response suppression during speech production. Brain Lang. 27, 50–55.

Sadagopan, S., Wang, X., 2008. Level invariant representation of sounds by populations of neurons in primary auditory cortex. J. Neurosci. 28, 3415–3426.

Salomon, B., Starr, A., 1963. Electromyography of middle ear muscles in man during motor activities. Acta Neurol. Scand. 39, 161–168.

Schneider, D.M., Mooney, R., 2015a. Motor-related signals in the auditory system for listening and learning. Curr. Opin. Neurobiol. 33, 78–84.

Schneider, D.M., Mooney, R., 2015b. Neural Coding of Self-generated Sounds in Mouse Auditory Cortex. Society for Neuroscience Abstracts.

Schneider, D.M., Nelson, A., Mooney, R., 2014. A synaptic and circuit basis for corollary discharge in the auditory cortex. Nature 513, 189–194.

Schuller, G., Beuter, K., Schnitzler, H.-U., 1974. Response to frequency shifted articial echoes in the bat *Rhinolophus ferrumequinum*. J. Comp. Physiol. 89, 275–286.

Sinnott, J.M., Stebbins, W.C., Moody, D.B., 1975. Regulation of voice amplitude by the monkey. J. Acoust. Soc. Am. 58, 412–414.

Sperry, R.W., 1950. Neural basis of the spontaneous optokinetic responses produced by visual inversion. J. Comp. Physiol. Psych. 43, 482–489.

Suga, N., Schlegel, P., 1972. Neural attenuation of responses to emitted sounds in echolocating bats. Science 177, 82–84.

Suga, N., Shimozawa, T., 1974. Site of neural attenuation of responses to self-vocalized sounds in echolocating bats. Science 183, 1211–1213.

Suga, N., Jen, P.H., 1975. Peripheral control of acoustic signals in the auditory system of echolocating bats. J. Exp. Biol. 62, 277–311.

Theunissen, F.E., Sen, K., Doupe, A.J., 2000. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J. Neurosci. 20, 2315–2331.

von Holst, E., Mittelstaedt, H., 1950. Das Reafferenzprinzip: wechselwirkungen zwischen Zentralnervensystem und peripherie. Naturwissenschaften 37, 464–476.

Young, E.D., Yu, J.J., Reiss, L.A., 2005. Non-linearities and the representation of auditory spectra. Int. Rev. Neurobiol. 70, 135–168.