

Dissociation of Unit Activity and Gamma Oscillations during Vocalization in Primate Auditory Cortex

Joji Tsunada^{1,2} and Steven J. Eliades¹

¹Auditory and Communication Systems Laboratory, Department of Otorhinolaryngology, Head and Neck Surgery, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, 19104, and ²Department of Veterinary Medicine, Faculty of Agriculture, Iwate University, Morioka, Iwate Japan, 020-8550

Vocal production is a sensory-motor process in which auditory self-monitoring is used to ensure accurate communication. During vocal production, the auditory cortex of both humans and animals is suppressed, a phenomenon that plays an important role in self-monitoring and vocal motor control. However, the underlying neural mechanisms of this vocalization-induced suppression are unknown. γ -band oscillations (>25 Hz) have been implicated a variety of cortical functions and are thought to arise from activity of local inhibitory interneurons, but have not been studied during vocal production. We therefore examined γ -band activity in the auditory cortex of vocalizing marmoset monkeys, of either sex, and found that γ responses increased during vocal production. This increase in γ contrasts with simultaneously recorded suppression of single-unit and multiunit responses. Recorded vocal γ oscillations exhibited two separable components: a vocalization-specific non-synchronized (“induced”) response correlating with vocal suppression, and a synchronized (“evoked”) response that was also present during passive sound playback. These results provide evidence for the role of cortical γ oscillations during inhibitory processing. Furthermore, the two distinct components of the γ response suggest possible mechanisms for vocalization-induced suppression, and may correspond to the sensory-motor integration of top-down and bottom-up inputs to the auditory cortex during vocal production.

Key words: auditory cortex; gamma oscillations; marmoset; sensory-motor; vocal production; vocalization

Significance Statement

Vocal communication is important to both humans and animals. In order to ensure accurate information transmission, we must monitor our own vocal output. Surprisingly, spiking activity in the auditory cortex is suppressed during vocal production yet maintains sensitivity to the sound of our own voice (“feedback”). The mechanisms of this vocalization-induced suppression are unknown. Here we show that auditory cortical γ oscillations, which reflect interneuron activity, are actually increased during vocal production, the opposite response of that seen in spiking units. We discuss these results with proposed functions of γ activity during inhibitory sensory processing and coordination of different brain regions, suggesting a role in sensory-motor integration.

Received Nov. 19, 2019; revised Feb. 10, 2020; accepted Feb. 26, 2020.

Author contributions: J.T. and S.J.E. designed research; J.T. and S.J.E. performed research; J.T. and S.J.E. analyzed data; J.T. and S.J.E. edited the paper; J.T. and S.J.E. wrote the paper; S.J.E. wrote the first draft of the paper.

The authors declare no competing financial interests.

This work was supported by National Institutes of Health Grant DC014299 to S.J.E., Triological Society Clinician-Scientist Development Award to S.J.E., the Ministry of Education, Culture, Sports, Science and Technology of Japan Leading Initiative for Excellent Young Researchers Grant 1071421 to J.T. and Grant A19K237690 to J.T., Ichiro Kanehara Foundation to J.T., and Daiichi Sankyo Foundation of Life Science to J.T. We thank T. Coleman and P. Sayde for assistance with animal training and care; and D. Contreras for comments on this manuscript.

Correspondence should be addressed to Steven J. Eliades at seliades@pennmedicine.upenn.edu.

<https://doi.org/10.1523/JNEUROSCI.2749-19.2020>

Copyright © 2020 Tsunada and Eliades

Introduction

Vocal communication plays an important behavioral role for both humans and many animal species. Accurate communication, however, requires auditory self-monitoring to correct any errors and ensure the accuracy of produced vocal sounds. Recent evidence has demonstrated that activity in the auditory cortex is suppressed during vocal production, a phenomenon observed for both human speech (Paus et al., 1996; Numminen et al., 1999; Crone et al., 2001b; Ford et al., 2001; Flinker et al., 2010; Greenlee et al., 2011; Whitford, 2019) and nonhuman primate vocalization (Muller-Preuss and Ploog, 1981; Eliades and Wang, 2003, 2019). This vocalization-induced suppression has been suggested to result from top-down predictions of expected auditory feedback (Niziolek et al., 2013; Houde and Chang, 2015) and appears to play a role in both vocal self-monitoring during

altered feedback (Eliades and Wang, 2008b; Greenlee et al., 2013) and feedback-dependent vocal control (Chang et al., 2013; Behroozmand et al., 2016; Eliades and Tsunada, 2018). However, despite this potentially important role for vocal suppression of auditory cortex, the underlying neural mechanisms and whether this is primarily a result of cortical inhibition remain unclear.

There has been considerable recent interest in the role of high-frequency brain oscillations in sensory processing. In particular, γ -band activity (>25 Hz) has been implicated in a variety of neural processes, including sensory binding of stimulus features (Eckhorn et al., 1988; Singer, 1999), object representation (Tallon-Baudry and Bertrand, 1999), attention and arousal (Tiitinen et al., 1993; Fries et al., 2001; Womelsdorf and Fries, 2007), associative learning (Miltner et al., 1999; Jeschke et al., 2008), and sensory-motor integration (Murthy and Fetz, 1992; Sanes and Donoghue, 1993; Schoffelen et al., 2011). In the auditory cortex, γ oscillations have been demonstrated during sound presentation in a variety of species, including bats (Medvedev and Kanwal, 2008), rats (Sukov and Barth, 2001; Vianney-Rodrigues et al., 2011), gerbils (Jeschke et al., 2008), cats (Lakatos et al., 2004), and nonhuman primates (Brosch et al., 2002; Steinschneider et al., 2008). The neural mechanisms of these cortical γ oscillations remain an area of active study, although it is generally accepted that they reflect local cortical activity (Barth and MacDonald, 1996; Welle and Contreras, 2016) and, in particular, the interactions of local inhibitory interneurons with nearby pyramidal cells (Whittington and Traub, 2003; Bartos et al., 2007; Cardin et al., 2009; Sohal et al., 2009; Buzsaki and Wang, 2012; Chen et al., 2017). However, the study of γ oscillations in sensory cortex has largely been limited to excitatory sensory stimuli, rather than cortical processes with large inhibitory components (Ray and Maunsell, 2011). In contrast, the suppression of auditory cortex neurons during vocalization has been well described, but the role of local inhibitory processes is unclear.

Given the apparent role of γ oscillations in local inhibitory processing and other top-down sensory-motor processes, and the uncertain role of cortical inhibition during vocalization-induced suppression, we sought to determine whether γ oscillations were modulated during vocal production. We simultaneously recorded spiking activity and cortical oscillations in the auditory cortex of vocalizing marmoset monkeys and found robust γ -band activity during vocal production that correlated with the magnitude and timing of vocal suppression. This relationship between unit and γ activities was distinct from that seen during passive listening, implicating γ oscillations as a potential marker of local inhibitory activity during vocalization.

Materials and Methods

We recorded neural activities from 3 adult marmoset monkeys (*Callithrix jacchus*), 1 female and 2 males, while the animals produced self-initiated vocalizations. Neural activity from auditory cortex was recorded using implanted multielectrode arrays, including both spiking and local field potential (LFP) activity, and compared with simultaneously recorded vocal behavior. All experiments were conducted under the guidelines and protocols approved by the University of Pennsylvania Institutional Animal Care and Use Committee.

Vocal recordings

Using previous methods (Eliades and Wang, 2008a; Eliades and Tsunada, 2018), we recorded marmosets vocalizing while in their home colony. Subjects were placed in a small cage within a custom three-walled sound-attenuating booth allowing free visual and vocal interaction with the remainder of the marmoset colony. During recordings, marmosets were

tethered within a small cage, to allow neural recording, but were otherwise unrestrained. Vocalizations were recorded using a directional microphone (Sennheiser ME66) placed ~20 cm in front of the animal, amplified (Focusrite OctoPre MkII), and digitized at 48.8 kHz sampling rate (RX-8, Tucker-Davis Technologies). Vocalizations were extracted from the recordings and spectrographically classified into established marmoset call types (Agamaite et al., 2015) using a semiautomated system. All major call types were produced in this context (phee, trillphee, trills, twitters); however, we excluded the multiphrase twitter calls to instead focus on the other three types, which exhibit more continuous rather than phasic vocal production.

Neural recordings

All marmosets were implanted bilaterally with multielectrode arrays (Warp 16, Neuralynx), one in each auditory cortex. Details of the array design and recording technique have been previously published (Eliades and Wang, 2008a). These arrays consist of a 4×4 grid of individually moveable sharp microelectrodes (4 M Ω tungsten; FHC). Consistent with our previous methods, we first localized the center of primary auditory cortex using single-electrode methods, and placed arrays to cover the full range of the tonotopic axis, verified by frequency tuning. Based on relative responses to tone and noise stimuli, electrodes were judged to likely span both primary (A1) and nonprimary (belt, parabelt) auditory cortex (Rauschecker and Tian, 2004). Because of variability in electrode array placement and cortical anatomy, it is unclear whether any potential differences based on cortical field or hemisphere were due to sampling bias, and we therefore did not perform systematic location comparisons.

During recordings, neural signals were passed through a unitary-gain headstage (Tucker-Davis Technologies, RA16CH) and digitized (Tucker-Davis Technologies, System III PZ2 and RZ2). A single electrode on each array was used as a reference and subtracted from the remaining electrodes to reduce muscle potentials and other movement artifacts. Neural signals were observed online to guide electrode and optimize signal quality.

Digitized signals were sorted offline using custom MATLAB software and a principle component-based clustering method, and then classified as either single-unit or multiunit as previously described (Eliades and Wang, 2008a). Only recording sites with spiking units (single or multi) were included for subsequent analyses to allow direct comparisons between γ oscillations and spiking. No significant vocal γ responses were seen at sites in which spiking was not observed, perhaps due to greater distance from local neurons. Targeted sites were further analyzed to extract multiunit activity (MUA), reflecting the summed responses of local neural populations (Super and Roelfsema, 2005). The MUA analysis was chosen because individual recording sites often contain a mix of both vocally suppressed and excited single units (Eliades and Wang, 2005). Recorded LFPs, including γ -band activity, likely reflect summed local inputs simultaneously projecting onto this heterogeneous unit population but cannot be distinguished to determine which inputs affect which unit. Therefore, we took the conservative approach to compare this field activity to the averaged local activity, reflected by the MUA, of a recording site. We generated MUAs by first subtracting the reference electrode signal from the raw neural signal, then bandpass filtering (300–5000 Hz), full-wave rectifying, and finally low-pass filtering (500 Hz) before downsampling.

Auditory stimuli

Before each neural recording in the colony, we first characterized tuning properties of the auditory cortex neurons by the presentation of auditory stimuli. Marmosets were seated in a custom primate chair within a soundproof chamber (Industrial Acoustics). Auditory stimuli were digitally generated at 97.6 kHz sampling rate and delivered using Tucker-Davis Technologies hardware (System III) in free-field through a speaker (B&W 686 S2) located ~1 m in front of the animal. Stimuli included tones (1–32 kHz, 10/octave; –10–80 dB SPL by 10 dB) and bandpass noise (1–32 kHz, 5/octave, 1 octave bandwidth), as well as wide-band noise stimuli. The center frequency (CF) of a neuron's frequency responses area was determined by the strongest MUA response to either tone or bandpass stimuli. We further presented multiple recorded

vocalizations at different sound levels, including samples of animals' own vocalizations (previously recorded from those animals) and conspecific vocalization samples (from other animals in the marmoset colony). Vocal stimuli were presented at multiple sound levels, but only those samples overlapping produced vocalization loudness were used for comparisons between vocal production and auditory playback.

Data analysis

LFPs were calculated from the referenced electrode signals by filtering (0.1–300 Hz) and then downsampling. We focus here on the γ -band activity due to its suggested role in cortical inhibition. γ -Band activity was calculated as the log-power of the bandpass-filtered LFP (25–60 Hz) from each trial aligned by vocal onset and then averaged across trials. Adjacent frequency bands were also examined as controls to determine the specificity of γ -band changes. γ responses during vocalization were calculated as the log-power change relative to the baseline period (–1000 to –500 ms). The γ power that was not phase-locked to the onset of vocalization (induced γ) was quantified from single trials after subtraction of the vocal onset-aligned average LFP (Galambos, 1992; Crone et al., 2001a; Brosch et al., 2002). The γ power that was phase-locked to the vocal onset (evoked γ) was calculated as the difference between the total and induced responses. Because of trial-to-trial variability in vocal duration, we focused our analysis on a window of 0–300 ms after vocal onset (vocal period) for our analyses, unless otherwise specified. Time-frequency plots of the vocal response were calculated, for display purposes, using a 6-point Morlet wavelet transform of individual trial LFPs. The time course of γ responses was calculated from γ power, binned to match unit peristimulus time histograms (PSTHs, 20 ms bins), and smoothed (5-point) before plotting. Analyses were performed similarly for responses to playback of vocal stimuli. For population average PSTHs, we subtracted mean baseline activity before population averaging. We did not attempt to directly compare the timing of spiking activity relative to γ oscillation phase due to the paucity of action potentials in our vocally suppressed units.

An additional control analysis was performed to evaluate the possibility that part of the γ -band oscillatory activity was due to phase synchronization between neural activity and the sinusoidal oscillation seen in marmoset trill vocalizations. A cross-coherence analysis was performed between single-trial LFP responses and the frequency contour of vocalization. Vocal frequency contours were calculated from time-binned peak frequency in the acoustic spectrogram. The cross-coherence was calculated from the squared product of the complex Fourier transforms of the vocalization and the LFP response, normalized by the magnitude of the individual Fourier transforms, and then averaged across trials. This coherence shows the phase alignment between the two signals across both frequency and time, and is bounded between 0 (no phase alignment) and 1 (perfect alignment). The cross-coherence analysis was also performed for the MUA and the LFP power, both of which lack the specific phase information seen in raw LFP trials.

Comparisons of vocal responses and unit CF were performed by calculating a smoothed moving average. MUA or γ vocal responses were ordered by CF and then smoothed with a 21 point moving average. In order to determine significance of this association, we performed a shuffle-resampling procedure, randomizing the association between vocal response and CF and then smoothing the result. This was repeated 1000 times, and then confidence intervals (CIs) calculated that included 95% of these shuffled moving averages, which correspond to the expected range for a null hypothesis of no association between responses and CF.

Experimental design and statistical analyses

We recorded neural and behavioral (vocal) activity from 3 marmosets. All vocalizations produced in the three most common long-call categories (trill, trillphee, phee) were included in analyses, unless otherwise noted in the text. Because we wanted to evaluate the relationship between unit activity and γ responses and because some implanted electrodes may have been outside auditory cortex or not auditory-responsive, only sites with significant unit or MUA responses to either vocal production or sound presentation (either tones, bandpass noise, or vocal playback; $p < 0.05$, Wilcoxon rank-sum test) were included in further

analysis. A total of 550 sites were included for the first animal, the longest recorded, and 336 and 370 for the remaining two.

With the exception of correlation analyses, all statistical tests were performed using nonparametric methods. Wilcoxon rank-sum and signed-rank tests (two-sided) were used to test the differences between unmatched and matched distribution medians, respectively. Kruskal-Wallis ANOVAs were used when comparing more than two conditions. Correlation values and linear regressions within individual unit, and between unit parameters, were calculated with Pearson correlation coefficients, with p values and CIs calculated from the t distribution. For moving-window methods of measuring time changes in MUA or γ responses, p values were first calculated for individual time bins relative to prevocal baseline bins, and then false discovery rate (FDR)-corrected for multiple time point comparisons. Shuffle-resampling methods were used to calculate CIs of CF associations, as discussed above. p values < 0.05 were considered statistically significant throughout.

Data availability

The data and computer code that support the findings of this study are available from the corresponding author on request.

Results

We recorded responses bilaterally from auditory cortex of marmoset monkeys during voluntary self-initiated vocalizations. In total, we report results from 1256 recording sites that exhibited significant single-unit or multiunit responses. We were interested in understanding the relationship between vocalization-induced suppression of spiking output and presence of γ oscillations during vocal production.

Auditory cortex exhibits γ -band responses during vocalization

Figure 1 shows a sample unit exhibiting strong vocalization-induced suppression. Consistent with previous results, single-unit responses significantly decreased (-3.52 ± 3.21 spk/s; $p < 0.001$) during vocalization (Fig. 1A,B). Our previous work has suggested that nearby neurons often show disparate responses during vocalization (Eliades and Wang, 2003) and, as a result, we also examined MUA (Fig. 1C,D) responses which reflect the sum of local neural activities and may serve as a better unit-level measure when comparing γ -band activity. MUA responses at this site also exhibited significant vocal suppression (-4.25 ± 3.24 μ V; $p < 0.001$), although with a prominent onset-aligned peak that may be a result of additional nearby neurons with excitatory responses to vocal onset. There was a strong correlation between the single-unit PSTH and MUA responses at this site ($r = 0.78$, $p < 0.001$). We next examined the frequency response of LFP activity at this recording site during vocalization. Time-frequency (Fig. 1E) and power-spectral density (Fig. 1F) analyses show a significant increase in spectral power during vocalization centered at 31 Hz ($p < 0.05$, range 29–33 Hz). We did not see any significant changes in nearby frequency bands, either in higher or lower frequencies, suggesting a narrow band-specific change in γ power. Overall γ -band power (25–60 Hz) during vocalization was increased with respect to baseline ($+1.49 \pm 1.59$ dB; $p < 0.001$), visible both at the single-trial (Fig. 1G) and average level (Fig. 1H). This increased γ power contrasts with the concurrently decreased spiking and MUA PSTHs ($r = -0.74$ and $r = -0.57$, respectively; $p < 0.001$ for both). This dissociation and robust γ -band response suggest that γ activity during vocalization does not reflect the output spiking activity of these auditory cortex neurons.

Induced and evoked γ responses during vocalization

γ -Band and other cortical oscillations have been shown to contain two different components (Galambos, 1992). The first are

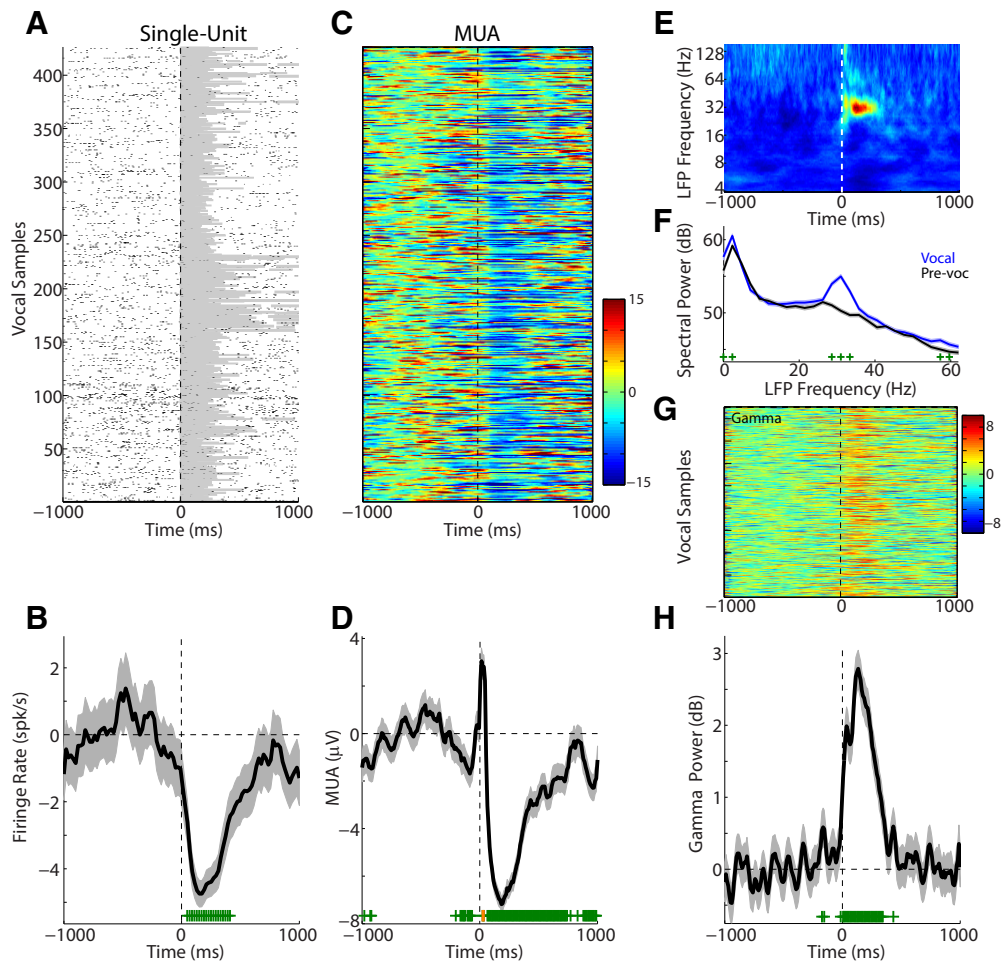


Figure 1. Representative vocalization responses from a single neuron and associated multiunit and field potential activity. **A**, Raster plot of single-unit spiking activity before, during (shaded area), and after vocalization. **B**, PSTH (20 ms bins) for the same single-unit showing decreased firing (suppression) compared with baseline. Error bars indicate SEM (shaded). Bins with significant change from baseline are indicated (green; $p < 0.05$, rank-sum test with FDR correction). Vertical dashed line indicates vocal onset. **C**, **D**, Single-trial and averaged MUA responses are shown, recorded simultaneously with the unit in **A** (orange represents time bins with significant increases). The MUA at this site was similar to the single unit, but with an early peak shortly after vocal onset. **E**, Time-frequency analysis of the LFP recorded simultaneously with the single and MUA responses shows an increase in mid-frequency power after vocal onset in the 30–32 Hz range. **F**, Power-spectral density comparisons of spectral power between prevocal and vocal intervals showed an increase in power centered at ~ 31 Hz ($p < 0.05$, green marks). **G**, **H**, Single-trial and average γ -band (25–60 Hz) power showing an increase during vocalization.

evoked oscillations, activity that is phase-locked to the onset of a stimulus, peaks early in the response, and may reflect stimulus inputs and feature binding (Galambos et al., 1981; Joliot et al., 1994; Steinschneider et al., 2008). The second are induced oscillations, which are poorly phase-locked to the stimulus onset, peak later in time, and are more task- and condition-dependent (Marshall et al., 1996; Crone et al., 2001a; Jeschke et al., 2008; Steinschneider et al., 2008). We therefore examined the relative contributions of evoked and induced γ activity during vocalization. Figure 2A, B shows two representative multiunit responses that were suppressed during vocalization, while exhibiting strong γ responses to vocalization. These γ responses were dominated by induced activity, particularly in the later times (Fig. 2A,B, bottom, blue curves). The largest difference between the total and induced γ responses, reflecting the evoked activity, was only notable as an early peak shortly after vocal onset.

We next examined the total γ power during vocalization across the population of recorded sites (Fig. 2C). Although a wide distribution of responses was seen, including some sites with decreased γ power, there was a significant bias toward increased total γ power during vocalization (mean \pm SD:

0.46 ± 0.88 dB; $p < 0.001$, $z = 16.3$, signed-rank). Overall, 42% of sites showed significant ($p < 0.05$) γ -band responses (Fig. 2C, shaded). We further measured the relative contributions of the induced and evoked components (Fig. 2D, blue represents induced; red represents evoked). There was a significant difference between the two, with increased induced activity across the population, but weaker average evoked γ responses for the same sites ($p < 0.001$, $z = 20.3$). This suggests that nonsynchronized induced oscillations dominate the γ -band response during vocalization.

As can be seen in the above examples, evoked activity showed a peak early after vocalization onset and induced activity began later. We measured the peak times of these two components and found that the evoked peaks averaged 21.6 ± 11.4 ms, whereas the induced peaks averaged 158.1 ± 61.2 (Fig. 2E). Based on the relative timing of these activities, we defined an onset period 0–50 ms following the start of vocalization (Fig. 2F). In this onset period, there were stronger evoked than induced γ responses ($p < 0.001$, $z = 11.69$). Directly comparing induced and evoked γ activities across the unit population revealed a wide distribution (Fig. 2G); there was a general correlation ($r = 0.64$, $p < 0.001$) between induced and onset-evoked responses, with many sites

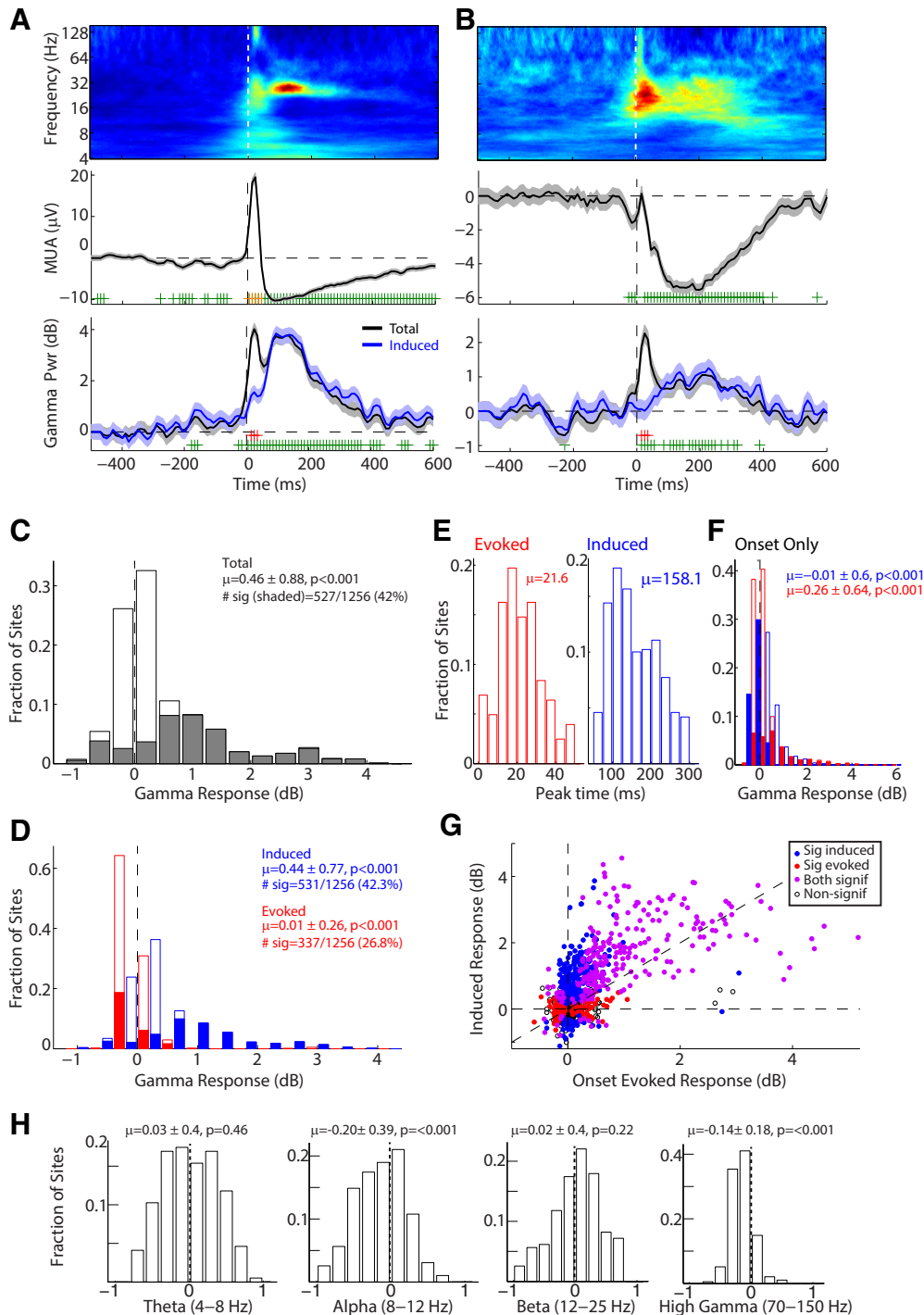


Figure 2. Induced and evoked γ responses during vocalization. **A, B**, Two example recording sites are shown, demonstrating two components of the γ -band response. Top, Time-frequency plots. Middle, MUA responses. Bottom, γ power. Both total (black) and nonsynchronized γ power (induced; blue) are plotted. Error bars indicate SEM. Significant changes in total γ are indicated (green; $p < 0.05$), as are time bins with significant differences between total and induced γ (red). **C**, Population histogram of total γ responses during vocalization showing large increases in vocal γ power. Sites with significant γ responses ($p < 0.05$) are filled, and accounted for 42% of all recorded sites. Mean and SD of the distribution are indicated, as is the number of significant sites (# sig). **D**, Distribution of vocal γ power divided into induced (blue) and evoked (red) responses. Filled bins represent sites with significant responses, as above. **E**, Distribution of peak response times, relative to vocal onset, for sites with significant evoked (left) and induced (blue) vocal γ responses. Evoked activity peaked early after vocal onset, with induced activity peaking later. **F**, Histogram of evoked and induced γ power immediately after vocal onset (averaged over 0–50 ms), showing larger evoked responses. **G**, Scatter plot comparing onset-evoked responses and sustained induced responses across the population. Colored markers represent sites with significant responses ($p < 0.05$): red represents evoked; blue represents induced; purple represents both. Non-significant sites are shown as open circles. **H**, Histograms of vocal oscillatory power in other frequency bands: theta (4–8 Hz), alpha (8–12 Hz), beta (12–25 Hz), and high γ (70–150 Hz). Reduced power was noted in alpha and high γ but not theta or beta-bands.

showing significant ($p < 0.05$) responses in both. Comparisons of the number of sites with significant evoked and induced activity reveal a significant bias toward induced (Table 1; $p < 0.001$, Fisher exact test). These results demonstrate that strong γ

activities are seen across the auditory cortex, but that these responses are dominated by a nonsynchronized induced component that is temporally distinct from a weaker and primarily onset-evoked response.

Table 1. Number of sites with significant induced and evoked γ activity

	Evoked significant	Evoked nonsignificant	Total
Induced significant	251	308	559 (45%)
Induced nonsignificant	115	569	684 (55%)
Total	366 (29.4%)	877 (70.6%)	

In order to determine whether these changes in cortical oscillations were specific to the γ -band, or whether they reflected a broad increase in oscillatory power, we further examined vocal responses in nearby frequency bands (Fig. 2H). We noted significant decreases in the higher-frequency high γ -band (70–150 Hz; -0.14 ± 0.18 dB; $p < 0.001$) and lower-frequency α -band (8–12 Hz; -0.20 ± 0.39 dB; $p < 0.001$). However, these modulations were weaker than those seen in the γ -band. The decrease in high γ is interesting as it likely reflects the decrease in spiking activity during vocalization-induced suppression. We did not see any power changes during vocalization in the adjacent β (12–25 Hz; 0.02 ± 0.40 dB; $p = 0.22$) or lower-frequency theta-band (4–8 Hz; 0.03 ± 0.40 dB; $p = 0.46$). These results suggest that the increase in γ power was a band-specific rather than a broad increase in cortical oscillations.

Comparison of γ activity with multiunit suppression and excitation

Our previous work in marmoset auditory cortex identified two populations of neurons with distinct spiking activities during vocalization (Eliades and Wang, 2003, 2013). The first are neurons with vocalization-induced suppression, which account for $\sim 65\%$ of all units. The second are neurons with vocalization-related excitation, which is thought to reflect sensory responses to vocal acoustics in nonsuppressed auditory units (Eliades and Wang, 2017). Based on these previous findings, we next compared γ -band activity to the degree of vocal suppression or excitation. Consistent with previous single-unit results, we found most sites' MUA responses were suppressed (Fig. 3A), 781 sites with significant ($p < 0.05$) suppression. Population average responses at these sites showed strong increases in induced γ activity, and a smaller onset-evoked component, similar to our individual-site examples (Fig. 3B). We also found a smaller number of sites with significant MUA increases during vocalization ($N = 67$; Fig. 3C). Interestingly, these sites also showed significant MUA decreases both before and after vocalization, suggesting that these multiunit responses may reflect a mix of nearby individual suppressed and excited neurons. Total γ power in these sites also showed large increases, but with a more prominent onset component (Fig. 3D). We quantitatively compared the degree of vocal suppression and γ power across the recorded population. Total γ power showed a U-shaped distribution ($p < 0.001$, $df = 12$, $\chi^2 = 262.5$, Kruskal-Wallis ANOVA), with higher values in both strongly suppressed (Fig. 3E, left) and strongly excited (Fig. 3E, right) sites. We further found a strong correlation between induced γ power and vocal suppression (Fig. 3F; i.e., negative correlation with MUA; $r = -0.50$, $p < 0.001$; Kruskal-Wallis: $p < 0.001$, $\chi^2 = 318.3$), whereas sites with nonsuppressed vocal responses exhibited little induced γ power. A qualitatively similar correlation was seen for suppression in single units and induced γ power ($r = -0.20$, $p < 0.001$). In contrast, the evoked γ activity (measured in the onset period, as defined above) strongly correlated with the MUA responses also observed to peak during the onset period ($r = 0.56$, $p < 0.001$; Kruskal-Wallis: $p < 0.001$, $\chi^2 = 170.2$). These results suggest that the two components of the γ -band vocal response may contribute to different aspects of vocal unit responses, with early evoked activity corresponding to unit vocal excitation during the onset period and induced activity correlating with vocalization-induced suppression.

Prevocal suppression and γ activity

Previous work in marmoset auditory cortex has shown that vocalization-induced suppression of spiking activity often begins before the onset of vocalization, on the order of 250 ms, but can extend as early as a second (Eliades and Wang, 2003). Examination of the population average MUA responses in Figure 3 demonstrates similar prevocal decreases in unit activity. Interestingly, a prevocal increase in population γ -band power is also visible over similar time scales (Fig. 3B,D). We further examined prevocal γ and MUA changes for individual sites. Figure 4 shows an example multiunit with prevocal suppression (Fig. 4A) and a corresponding prevocal increase in γ power (Fig. 4B). This prevocal increase was primarily in the induced γ power and was significant as early as 200 ms before vocal onset. The timing of this increase was greater than could be explained based on the slow time scale of γ activity (30 Hz \sim 33 ms), time binning (20 ms), or smoothing (5 point moving average). We further compared evoked and induced γ -band activity in the prevocal period (-250 to -20 ms) across the population (Fig. 4C). Prevocal γ power was stronger for the induced than the evoked component ($p < 0.001$, $z = 21.3$). Only a small set of sites had significant prevocal-induced activity (15.8%), and many sites had significantly decreased evoked activity (34.2%), not surprising given that evoked activity is phase-locked to stimulus onset. Across different sites, the strength of the prevocal-induced γ power significantly correlated with that during vocalization ($r = 0.44$, $p < 0.0001$; linear regression slope 0.95, 95% CIs [0.84 1.06]), suggesting a common mechanism (Fig. 4D). Prevocal induced γ power also inversely correlated with the magnitude of prevocal MUA changes ($r = -0.28$, $p < 0.001$; Fig. 4E). These results demonstrate that the induced component of vocal γ activity, like vocalization-induced suppression of single units and multiunits, begins before the onset of vocalization. This prevocal timing and correlation further implicate induced γ 's involvement in vocal suppression, and suggest that induced γ may not be a result of ascending auditory inputs, but rather related to the act of vocalizing itself.

Comparison of γ -band activity during vocal production and passive listening

In order to determine how much of the observed γ activity during vocalization could be accounted for based on passive sensory responses and tuning, we next examined γ -band responses during the playback of vocal stimuli, previously recorded from the same animal, for the units tested during vocal production. Figure 5 shows a comparison of single-unit spiking (Fig. 5A,B), MUA (Fig. 5C), and γ power (Fig. 5D) during vocal production and playback. This unit exhibited strong responses to vocal playback, but suppression during vocal production. Simultaneous γ -band activity also showed an increase during playback. However, in contrast to vocal production, where induced γ activity showed a strong increase, playback-induced activity was decreased for this unit (Fig. 5D, blue). This induced γ playback decrease is not entirely unexpected; similar decreases have been previously seen in macaque auditory cortex during tone presentation (Steinschneider et al., 2008). Comparisons of total and induced γ across the recorded population showed similar results (Fig. 5E), with increased total γ power during playback (Fig. 5E, black), but an absent or reduced induced component (Fig. 5E, blue; $p < 0.001$, $z = 27.4$). As was seen for evoked γ during vocal production, there was an onset γ peak during playback at 29.6 ± 10 ms (Fig. 5F). This was slightly later than the peak seen during vocal production, but may be attributable to our playback stimuli, including a small time-lag between that start of the stimulus file and the actual vocal sound

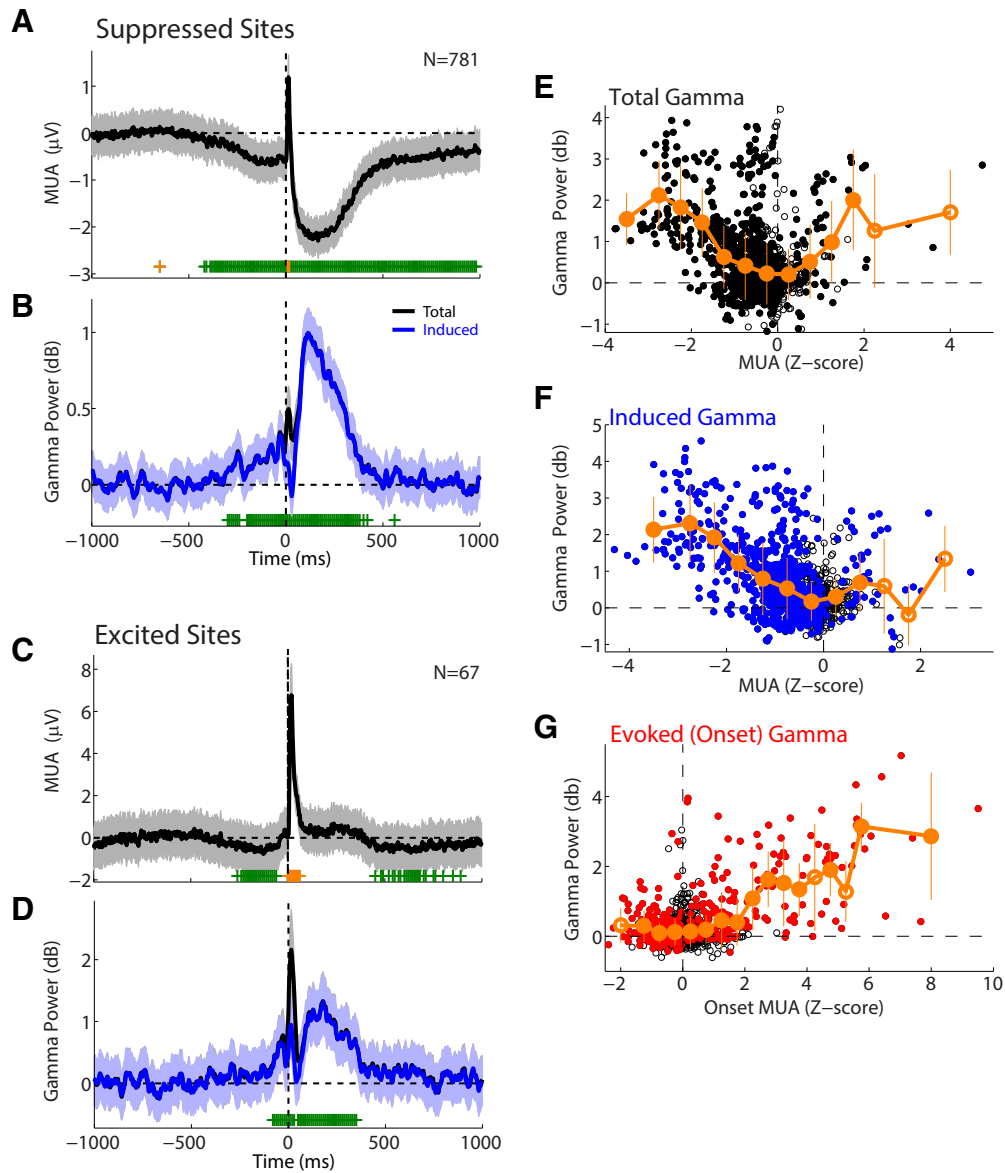


Figure 3. Population average responses in sites with suppressed and excited responses. **A**, Population average PSTH of MUA responses during vocalization for sites with significant vocalization-induced suppression ($p < 0.05$). SEM (shaded) and significant time bins (green represents reduced; orange represents increased) are shown. Number of sites is indicated. **B**, Population average γ responses for the suppressed sites shown in **A**. γ power is plotted for both total (black) and induced (blue) responses. **C**, **D**, PSTHs of MUA and γ activity for the smaller population of sites with vocalization-related excitation. **E**, Scatter plot comparing total γ power during vocalization with MUA responses across the population. MUA responses were z-score-normalized relative to baseline to compare responses across different sites. Mean γ responses, binned by MUA, are shown (orange). Error bars indicate bin SEM. Filled symbols represent significant bins ($p < 0.05$; signed-rank, FDR-corrected). Induced γ (**F**) and onset-evoked γ (**G**) are similarly plotted, showing correlations between induced γ and vocal suppression, and between evoked γ and increased MUA onset activity.

(typically ~ 10 ms). Onset period activity during playback (0–50 ms), like that of vocal production, was strongest for evoked rather than induced γ activity ($p < 0.001$, $z = 26.7$; Fig. 5G). We compared the strength of playback total γ activity with that of playback MUA responses (Fig. 5H) and found a weak, but significant, correlation ($r = 0.21$, $p < 0.001$). This correlation was stronger for onset MUA and evoked γ ($r = 0.77$, $p < 0.001$; Fig. 5I) than for induced γ ($r = -0.01$, $p = 0.57$; Fig. 5H). These results suggest that, unlike vocal production where nonsynchronized-induced γ oscillations are the primary observed response, passive auditory playback largely results in evoked γ oscillations. This comparison suggests that the onset-evoked γ response seen during vocal production may reflect the ascending auditory (sensory) inputs to the auditory cortex.

We next directly compared responses between vocal production and playback across the recorded population. Consistent with our previous results in single units (Eliades and Wang, 2017), there was only weak correlation between vocal and playback MUA ($r = 0.17$, $p < 0.001$), with stronger excitatory vocal responses at sites with strong playback responses, and more heterogeneous playback MUA responses (though still positive) in sites with vocal suppression (Fig. 6A). Interestingly, this correlation of MUA activity was stronger in the onset response ($r = 0.45$, $p < 0.001$). In contrast, total γ activity was highly correlated between vocal production and playback (Fig. 6B; $r = 0.63$, $p < 0.001$). Linear regression over the population showed a significant bias toward larger γ responses during vocalization (slope 0.37, CI [0.34, 0.39], $F = 815.5$). A direct comparison of

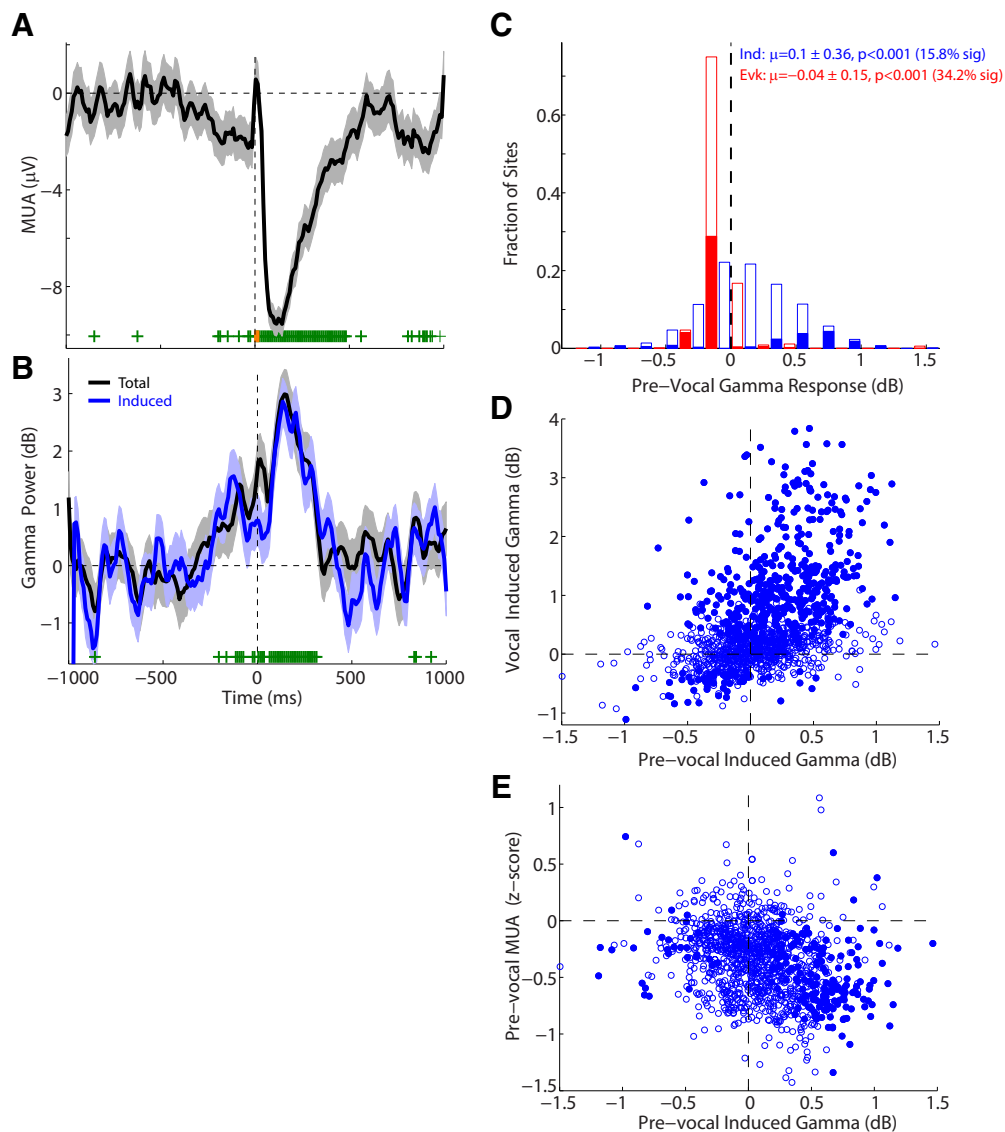


Figure 4. Pre-vocal suppression and γ activity. **A, B**, MUA and γ responses of an example auditory cortex site in which suppression began before the onset of vocalization. Induced γ was similarly increased in the prevocal period for this unit (up to 200 ms before vocal onset, **B**). **C**, Population histogram of induced (blue) and evoked (red) activity in the 250 ms prevocal period. Only a fraction of sites showed increases in prevocal induced γ power (15.8%), whereas a number showed decreased prevocal evoked power (34.2%). **D**, Raster plot showing a correlation between prevocal and vocal induced γ power ($r=0.44, p<0.001$; slope 0.95, 95% CI [0.84, 1.06]). **E**, Raster plot showing inverse correlation of prevocal induced γ power and prevocal MUA responses ($r=-0.28, p<0.001$), which is a positive correlation between γ responses and vocal suppression.

the change in total γ responses between vocalization and playback revealed the largest vocal increases in γ power for those sites that were more strongly suppressed during vocalization (i.e., negative vocal MUA; $r=-0.29, p<0.001$), and unchanged responses for excited sites (Fig. 6C). Separately examining evoked and induced γ changes between vocalization and playback revealed that the increase in total γ power was mostly due to the induced response (Fig. 6D, blue), which showed a significant negative correlation (vs vocal MUA; $r=-0.52, p<0.001$). In contrast, onset-evoked activity was decreased in sites with decreased MUA (Fig. 6D, red; $r=0.27, p<0.011$). Induced activity in the onset period and evoked activity in the sustained vocal period showed only modest changes. These results demonstrate a large increase in γ -band activity during vocalization compared with playback, particularly a large increase in induced γ power in suppressed units.

While these results demonstrate overall increased γ -band activity, they also suggest that evoked γ activity was decreased from

playback. However, we also observed that the background γ power seen in the baseline period was also increased between playback and vocal conditions. This difference may be attributable to the measurement of playback responses in a quiet sound booth with a passive animal, while vocalizations were measured in the noisy marmoset colony with a more attentive and engaged animal. Both contextual differences could have increased the overall background γ activity, and therefore artifactually decreased baseline-referenced vocal γ responses. We therefore compared raw γ power, not corrected for prevocal baseline, between vocal and playback conditions and found average increases in γ power (vocal-playback) were 3.23 ± 4.91 (total), 1.96 ± 5.15 (evoked), and 3.74 ± 4.92 (induced), all $p<0.001$ and all equal or greater to the change in baseline γ power ($1.97 \pm 4.82, p<0.001$). Alternatively, if we instead focus only on units with strong playback γ responses (those with playback greater than the vocal testing baseline, and $p<0.05$), baseline-referenced γ responses are still increased during vocalization over playback: total 1.54 ± 1.31 ($p<0.001$), evoked 0.22 ± 2.03

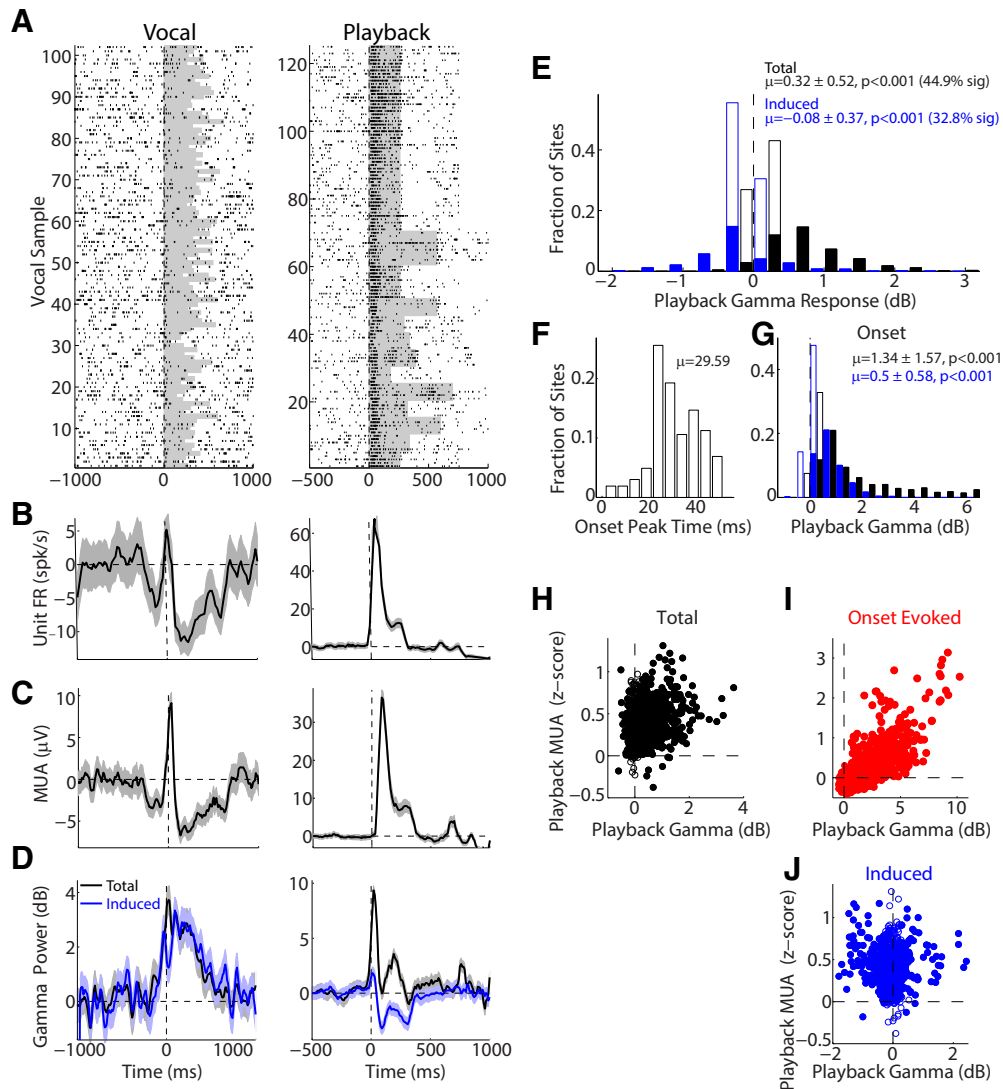


Figure 5. γ responses during playback of vocal stimuli. **A**, Raster is shown for a sample single-unit response to vocal production (left) and passive playback of vocal sound stimuli (right). **B**, **C**, Spike and MUA PSTHs for the vocal and playback responses shown in **A**. **D**, Total (black) and induced (blue) γ responses corresponding to the unit responses above. Unlike vocalization, where γ responses were largely induced, induced γ power during playback was decreased for this recording site. **E**, Population histogram of total (black) and induced (blue) γ during playback, with positive total γ responses, but weak to slightly reduced induced. Filled bins represent sites with significant playback γ responses ($p < 0.05$). Percent of significant sites is indicated. **F**, Timing distribution of playback total γ peak. **G**, Histogram for total (black) and induced (blue) γ during the onset period (0–50 ms). **H**, Scatter plot comparing playback total γ and MUA responses, showing a weak correlation ($r = 0.21$, $p < 0.001$). Strong correlation was seen for evoked γ responses in the onset period (**I**, $r = 0.77$, $p < 0.001$). However, there was no correlation seen for the weaker induced playback response (**J**, $r = -0.01$, $p = 0.57$).

($p = 0.33$), and induced 3.02 ± 1.85 ($p < 0.001$). These results suggest baseline changes in γ power cannot account for increases in total and induced γ during vocalization compared with playback. Additionally, these results further suggest that evoked γ responses may not change much between vocalization and playback, in contrast to the increase in induced activity.

Coherence of γ -band activity and vocal acoustics

Although γ oscillations have been seen in response to a variety of sensory stimuli, the origin of γ responses during vocalization is unclear. One possible confounding explanation could be the acoustics of the marmoset vocalizations themselves. Some marmoset vocalizations, notably the trill call, contain sinusoidal frequency modulations (Agamaite et al., 2015). This sinusoidal component tends to oscillate at ~ 30 Hz, in the same frequency range as γ oscillations. We therefore performed a phase coherence analysis between the LFPs and the frequency contours of

the vocalizations. Figure 7A shows a time-frequency plot of this phase coherence for a sample recording site, with 0 indicating no coherent oscillations and 1 indicating perfect synchronization. This analysis showed strong low-frequency onset coherence, but also a small delayed coherence in the 30 Hz range. Measurement of the peak coherence in the γ frequency band showed two peaks reaching as high as 0.5 (Fig. 7B, blue). It is not clear, however, whether this coherence was due to a specific alignment of γ oscillation phase to the vocalization, or whether it could have been due to a fluctuation in the overall response power with the vocalization. We performed a similar coherence analysis between the vocalization and both the MUA, which contains no low-frequency phase information, and the overall LFP power, where specific oscillation phase is discarded during calculation, and saw similar but slightly weaker coherence values (Fig. 7B). Across the population, there were generally similar average coherence values for the γ oscillation and MUA/LFP power (Fig. 6C). The

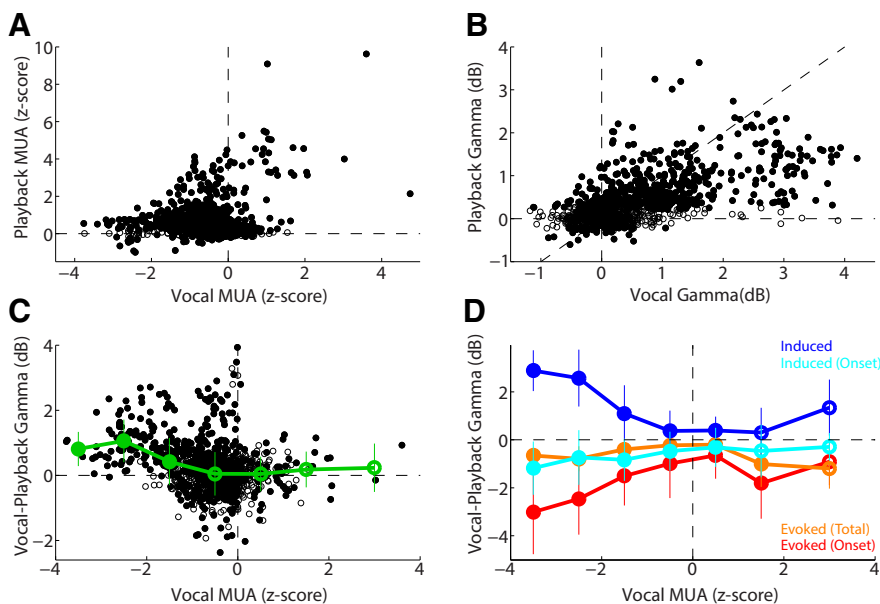


Figure 6. Comparison of vocal production and playback γ responses. **A**, Scatter plot comparing z-score-normalized MUA activity between vocal production and playback. Units with positive vocal MUAs (excited units) also had positive playback responses. Those with negative vocal MUA responses (suppressed) also had increased playback responses but were more diverse. **B**, Comparison of vocal and playback total γ -band power, showing a strong correlation ($r = 0.63$, $p < 0.001$), and bias toward stronger γ power during vocalization (slope 0.37, 95% CI [0.34, 0.39], or $\sim 2.5\times$ increase). **C**, Comparison of vocal-playback total γ power difference as a function of vocal MUA responses. Considerable variability was seen for weakly suppressed units (negative MUA), but there was an overall significant negative correlation between γ changes and vocal MUA ($r = -0.29$, $p < 0.001$). Mean and SEM γ changes, binned by vocal MUA response, are shown (green). Filled symbols represent significant bins. **D**, Comparison of vocal-playback γ changes for induced and evoked components, divided by onset (0–50 ms) and total/sustained timing (0–300 ms). Binned mean and SEM are shown. Induced γ exhibited large vocal increases for suppressed units (negative vocal MUA), whereas evoked γ showed apparent decreases.

slopes of this comparison were 0.79 [95% CI: 0.77, 0.82] for MUA and 0.78 [0.76, 0.80] for LFP. This suggests that overall envelope fluctuations could account for $\sim 80\%$ of the apparent phase coherence of the γ oscillation, rather than solely due to a specific entrainment of LFP oscillations to the vocal acoustic modulation.

Another possible explanation for a strong non-phase-synchronized induced γ response, rather than a synchronized evoked one, could be γ oscillations synchronizing to the variable phases of the trill vocalizations, rather than vocal onsets. This could yield a γ power that appeared not to be stimulus-synchronized if calculated only relative to vocal onset. We therefore recalculated total and induced γ power relative to both vocal onset and to the first three peak times in the trill frequency oscillation. One sample recording site showed only a small reduction in the induced γ power with the peak realignment (decreases of 13%–17%), suggesting alignment cannot account for most of the induced response (Fig. 7D). This was also true across the population, where there was little change in the total γ response (0.4%) with the realignment, and only a 2.6% reduction in the induced γ power (Fig. 7E).

As a further control, we also performed analysis of total and induced γ responses limited to only marmoset phee vocalizations, which do not contain a sinusoidal acoustic component. Figure 7F shows γ responses from a site exhibiting strong induced γ activity during phee vocalizations. Unfortunately, the marmosets in this study did not make a large number of phee vocalizations, and our population analysis is limited to a single animal and a small number of sites. Figure 7G shows γ responses during phee vocalizations across a number of sites, which exhibited

weak average total and induced γ activity (left). However, these sites also showed weak γ responses to sinusoidally modulated trill vocalizations as well. Indeed, these sites showed stronger γ activity to the nonsinusoidal phee than trills (Kruskal-Wallis: $p = 0.013$, $z = 2.48$). These same recording sites, however, did show strong γ responses to playback of both phee and trill vocal sounds (Fig. 7G, right), suggesting that a sinusoidal acoustic component is not necessary to cause a γ -band response.

γ -band activity and frequency receptive fields

One of the perplexing findings that has come out of previous studies of marmoset vocalization has been a lack of correlation between vocalization-induced suppression and the underlying sensory tuning of auditory cortex neurons (Eliades and Wang, 2003, 2017). Given the proposed role of suppression in vocal self-monitoring and control, one might expect suppression to be limited to those neurons around vocal frequencies, rather than more broadly distributed. This has not been found to be the case. We therefore calculated frequency receptive fields, using tone and bandpass noise stimuli, and measured the MUA CF response and sound-level sensitivity. All MUA responses exhibited monotonic sound-level responses,

with little variation in thresholds. However, comparisons of vocal responses for MUA and induced/evoked γ with CF showed considerable variability, with tendencies to cluster around the vocal fundamental and harmonic frequencies (Fig. 8A). In order to better examine these trends, we calculated a smoothed moving average of vocal responses ordered by CF (Fig. 8B). These moving averages showed the largest vocal responses in both induced and evoked γ power around vocal frequencies. These peaks exceed shuffle-calculated 95% CIs testing for random association, suggesting that both the induced and evoked vocal γ activity in auditory cortex is somewhat frequency-specific. In contrast, the MUA responses showed the largest vocal suppression in the mid-frequency range, but not specifically associated with vocal frequencies (Fig. 8B, black), as seen in previous results. MUA onset activity also failed to show clear frequency tuning. However, frequency specificity was seen for both evoked γ and MUA responses during vocal playback, consistent with a sensory response, although induced playback responses were weak.

Given the apparent transformation between frequency-tuned γ activity and nontuned MUA/spiking output, we sought to determine what combination of induced and evoked γ activity could result in a flattening of frequency tuning. We performed a simple analysis to linearly combine the γ -activity curves and compare with the MUA moving average. The simplest model, subtracting induced γ response from the onset evoked, yielded a result that reproduced the flattened MUA population frequency sensitivity in the vocal range (Fig. 8C, orange; comparison with MUA: $r = 0.54$, $p < 0.001$). In contrast, simply inverting the total

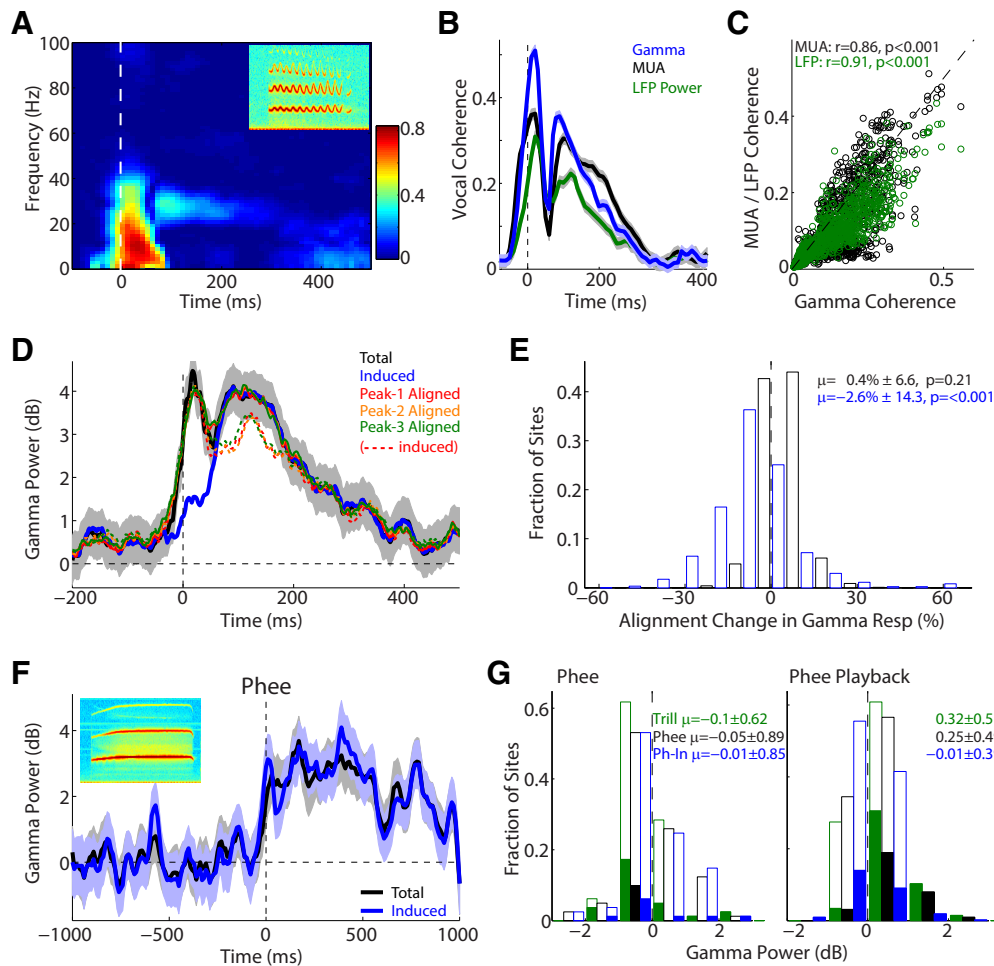


Figure 7. Effects of coherence and phase alignment on vocal γ . **A**, Time-frequency plot of phase coherence between LFP responses and vocal acoustic frequency contours for a sample recording. Coherence is bounded between [0, 1], where 1 indicates perfect phase alignment. Some delayed coherence is noted in the γ frequency range. Inset, Spectrogram of trill vocalization; y scale, 0–25 kHz; x scale, 600 ms. **B**, Plot of the peak coherence in the γ -frequency range over time (blue), for the example in **A**. Results are compared with similar coherence analysis for the MUA (black) and phase-removed LFP power (green), showing similar though weaker trends. **C**, Scatter plot comparing coherence in γ frequencies with that of the MUA and LFP power across the population, suggesting much of the apparent coherence in γ oscillations may have been due to the coherence in the overall response power, rather than the oscillations themselves. **D**, Sample PSTH of total and induced γ responses aligned by vocal onset (black/blue), and recalculated by aligning to the times of the first, second, and third frequency modulation peaks of trill vocalization acoustics. Solid lines indicate realigned evoked responses. Dashed lines indicate induced responses. There was a slight reduction in the average induced response (13%–17%). Onset peaks are seen for induced responses here due to phase alignment to trill oscillations rather than onset. **E**, Population histogram showing changes in total (black) and induced (blue) γ power during peak realignment versus onset alignment, showing only small changes in induced power (2.6% decrease). **F**, Sample unit γ response to phee vocalizations, which do not contain any frequency modulated acoustics, showing similar total and induced γ activity. Inset, Spectrogram of phee vocalization, y scale, 0–25 kHz; x scale, 2200 ms. **G**, Population histograms for γ responses to phee vocalization production (left) and playback (right). Total (black) and induced (blue) phee responses are shown; also shown are total γ responses to trill vocalizations for these same sites (green). Overall γ activity for these sites was weak, but stronger for phees than for trills (trills decreased $p < 0.001$, total and induced phee unchanged from prevocal: $p = 0.25$ and $p = 0.38$). During playback, phees evoked strong total γ power, though weaker than for trills (all $p < 0.001$).

γ activity, or the inverse of the summed onset evoked and induced, tended to preserve frequency tuning (gray: $r = 0.49$; green: 0.46, respectively, $p < 0.001$), although these models did a better job predicting the low-frequency responses than the evoked-induced model. These results suggest that the loss of frequency specificity during vocal suppression, compared with evoked and induced γ activities, might be accounted for by a simple model in which evoked γ activity is contributing an excitatory input while induced activity is contributing an inhibitory one.

Discussion

In this study, we examined γ frequency-band activity in the auditory cortex during self-initiated vocal production. Using simultaneous unit and field potential recordings, we demonstrated

several important findings with implications for both our understanding of the origin and function of cortical γ oscillations and for sensory-motor mechanisms in the auditory system. First, we found that γ -band power was increased during vocalization, in stark contrast to the dominant vocalization-induced suppression seen in both single-unit and multiunit responses. Second, we could divide these γ responses into stimulus onset-synchronized-evoked and non-synchronized-induced components. The evoked activity correlated with excitatory unit vocal responses and induced activity with vocal suppression. Third, evoked, but not induced, γ activity was seen during playback of vocal sound stimuli, and the largest differences between production and playback seen at sites with strong vocalization-induced suppression. Finally, vocal γ responses exhibited a specificity for sites tuned to vocal acoustic frequencies, in contrast to unit responses.

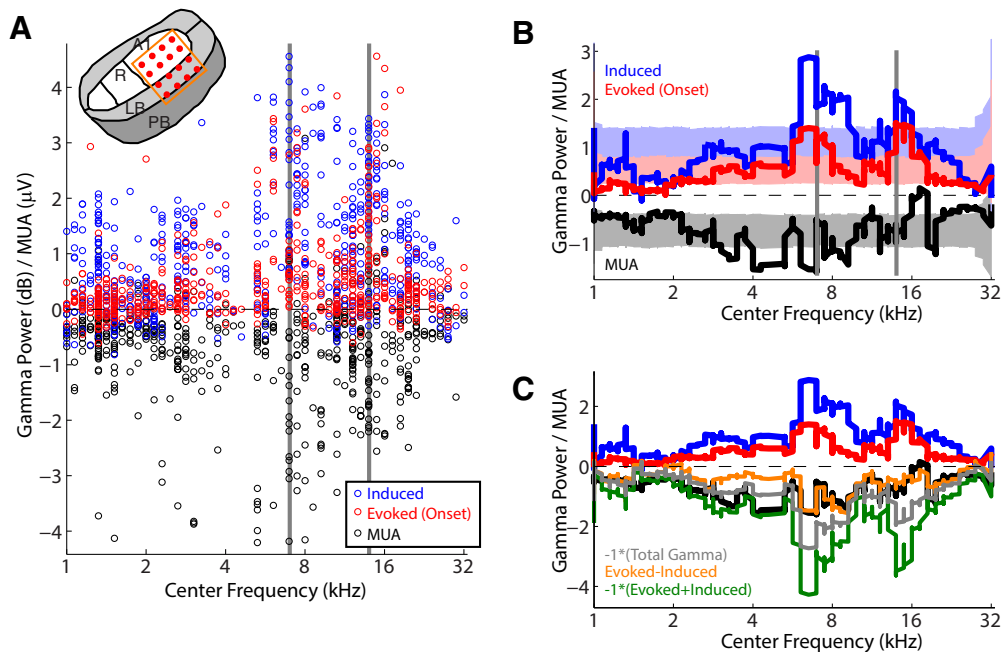


Figure 8. Comparison of vocal responses and frequency tuning. **A**, Comparison of vocal MUA (black) and induced/evoked (blue, red) γ responses by unit frequency tuning (CF). Vertical gray bars represent mean fundamental frequency of vocal acoustics and its harmonic. MUA response magnitudes have been divided by 2 to display with an equivalent magnitude to the γ power. Inset, Electrode positions within auditory cortex. **B**, Smoothed moving averages for the data plotted in **A** are shown (see Materials and Methods), along with 95% CIs from shuffled data for each of the moving averages (shaded). γ responses show peaks above the CIs around vocal frequencies, suggesting a significant association above chance. MUA responses did not show the same tuning pattern. **C**, Moving averages as above, shown with different combinations of γ power to model MUA tuning: inverse of total γ (gray; $r = 0.49$, $p < 0.001$), evoked-induced (orange; $r = 0.54$, $p < 0.001$), and inverse of onset-evoked + induced (green; $r = 0.46$, $p < 0.001$).

Origins of γ oscillation during vocal production

Recent evidence has emerged demonstrating that oscillations in the γ frequency range are a product of coupled activity between pyramidal cells and networks of local interneurons, specifically parvalbumin-expressing inhibitory interneurons (Cardin et al., 2009; Sohal et al., 2009; Chen et al., 2017). These oscillations appear to be generated within the cortex itself, rather than inherited from subcortical inputs, exhibit laminar specificity, and can be disrupted by local application of GABAergic blockade (Welle and Contreras, 2016). An important question, however, is the degree to which the frequency profile of oscillations is intrinsic to local network properties or a result of extrinsic sensory inputs with temporal characteristics within γ frequency ranges. In auditory cortex, there is a well-described frequency following response wherein cortical field potentials synchronize maximally to periodic stimulation at ~ 40 Hz (Galambos, 1992). However, animal studies have also shown γ -band responses to stimuli of different temporal characteristics, including tones and species-specific vocalizations (Brosch et al., 2002; Medvedev and Kanwal, 2008; Steinschneider et al., 2008).

While the current study observed strong γ -band activity during vocal production, an argument could be made that these were a result of cortical entrainment to the sinusoidal frequency modulation (typically ~ 30 Hz) seen in many marmoset vocalizations, including trill calls that made up a portion of the results. While many sites exhibited phase coherence between the LFP and vocal modulation, further analysis suggested that much of this was a result of coherence in the overall power envelope, rather than specific oscillatory phase synchronization. Additionally, γ responses were seen in many sites in response to nonsinusoidal phee vocal production and playback. We also examined whether the induced γ , which is not phase-synchronized to stimulus onset, could have been synchronized to acoustic modulations, but found only small decreases in induced γ power when realigning responses to these

acoustic features. Finally, the increase in induced γ during vocal production over playback, both with similarly modulated stimuli, and presence of prevocal-induced γ power, also argue against an origin from entrained responses. Together, these results suggest that γ -band activity during vocalization is likely largely intrinsic, generated from local cortical circuitry, rather than a synchronized response to periodic extrinsic inputs.

γ oscillations and cortical inhibition

Despite the clear role played by inhibitory interneurons in generating γ rhythms, evidence for a relationship between γ activity and inhibitory sensory processing is surprisingly lacking. Most studies examining γ responses in sensory cortex have focused on stimuli evoking strongly driving spiking responses and, as a result, have found a positive correlation between unit activity and γ power. One recent visual study attempted to dissociate the two, showing that γ power increased with increased grating size whereas spiking decreased (Ray and Maunsell, 2011). Similar studies in auditory cortex have found γ phase, but not power, changes in non-best-frequency inhibitory side-bands (O'Connell et al., 2011). The present results, demonstrating correlation between induced γ power and unit activity suppression, are strong evidence for the role of γ oscillations in inhibitory cortical processing. The absence of similar vocal increases in adjacent LFP frequency bands argues for a γ -specific phenomenon. Alternatively, such synchrony has been suggested as a mechanism to suppress competing inputs that are out of phase (Lakatos et al., 2007; O'Connell et al., 2011), and may be one potential mechanism by which auditory cortex neurons could increase their sensitivity to changes in vocal feedback (Eliades and Wang, 2008b), comparing top-down predictions of expected vocal acoustics with bottom-up afferent sensory input.

γ -band responses during human speech

One of the potential advantages in using cortical oscillations to study vocalization-related activity is the ability to compare responses between human and animal work. Previous human intracranial electrocorticography has focused on the high γ -band (>70 Hz), which is correlated with spiking activity (Ray and Maunsell, 2011), and has found vocal suppression qualitatively similar to that seen in marmosets (Crone et al., 2001b; Flinker et al., 2010). However, previous electrocorticography studies have not reported changes in lower γ (<70 Hz) power during speech production. Interestingly, high-frequency responses have been noted that matched the pitch of a speaker's voice \sim 100 Hz, (Behroozmand et al., 2016), which may show parallels to the phase synchronization seen in marmosets. It is still unclear, however, how closely the presents results match those seen in humans, a potential avenue for further interspecies comparisons.

Implications for auditory-vocal mechanisms

Theoretical models of speech motor control have suggested that the suppression of auditory cortex during vocal production results from sensory-motor comparisons between top-down predictions of expected vocal acoustics and bottom-up sensory feedback information (Houde and Nagarajan, 2011). Evidence that vocalization-induced suppression results from top-down efferent inputs comes from recordings showing suppression begins before the onset of vocalization (Eliades and Wang, 2003), where no auditory input is present, and by relative absence of suppression during playback of vocal sounds (Eliades and Wang, 2017). Despite this suppression, auditory cortex neurons remain sensitive to experimentally induced changes in vocal feedback, such as masking noise (Eliades and Wang, 2012) and pitch-shifts (Eliades and Wang, 2008b; Eliades and Tsunada, 2018), suggesting integration of bottom-up auditory feedback. However, because previous experiments have been limited to neuronal spiking outputs, it is unclear where and how these two signals are integrated.

Results of the present study reveal insights into possible mechanisms. γ -band oscillations, thought to be generated within cortex itself by local inhibitory interneurons, are increased during vocalization. Our findings therefore suggest that the site of vocalization-induced suppression is specifically located within the auditory cortex itself, not inherited from the ascending auditory pathway, and is a result of local cortical inhibition. The two separable components of the γ response are particularly interesting, as these may reflect simultaneous sensory and motor inputs to the auditory cortex, and cortex as a possible site of sensory-motor integration. The induced response strongly correlates with the degree of vocal suppression in unit activity, shows similar temporal characteristics, including peak time and prevocal onset, and is largely absent from playback responses. These findings suggest that induced γ during vocalization may be a specific marker of top-down efferent activity. Such nonsynchronized γ responses have been previously implicated in similar top-down phenomena during different tasks, including associative learning (Miltner et al., 1999; Jeschke et al., 2008) and sensory-motor integration (Murthy and Fetz, 1992; Sanes and Donoghue, 1993). Induced γ activity may therefore represent the result of local inhibitory networks under the influence of top-down modulation from motor, premotor, or prefrontal areas (Hage and Nieder, 2013, 2016; Roy et al., 2016) involved in planning and generating vocal production, many of which also show evidence of both auditory and vocal responses (Hage and Nieder, 2015; Hage, 2018).

In contrast to the induced component, evoked γ activity peaks much earlier in the response, correlates with excitatory onset spiking, and is present in both vocal production and playback. These findings suggest evoked γ as a marker for bottom-up sensory inputs from the ascending auditory pathway. Similar evoked γ has been seen in auditory cortex in response to a variety of sound stimuli. Further work examining evoked responses during altered feedback may yield additional insights. Interestingly, these evoked responses showed greater similarity between vocalization and playback than seen for induced activity. This further suggests that the ascending pathway does not exhibit the suppression seen in cortex, a subject of controversy (Eliades and Wang, 2003, 2019), and that vocal suppression is largely a cortical phenomenon (Rummell et al., 2016). Finally, we also note that both induced and evoked γ responses exhibited specificity for cortical sites tuned around vocal frequencies. This is consistent with evoked γ as auditory input but also suggests that inputs responsible for induced γ are similarly tuned, perhaps reflecting frequency specificity of top-down sensory predictions. A simple excitatory/inhibitory linear combination of these inputs may explain the lack of frequency specificity in spiking suppression but would benefit from further studies and modeling. Together with previous findings of feedback sensitivity (Eliades and Tsunada, 2018), our results implicate the auditory cortex as an important nexus for the integration of sensory prediction and auditory feedback to support vocal self-monitoring and control.

References

- Agamaite JA, Chang CJ, Osmanski MS, Wang X (2015) A quantitative acoustic analysis of the vocal repertoire of the common marmoset (*Callithrix jacchus*). *J Acoust Soc Am* 138:2906–2928.
- Barth DS, MacDonald KD (1996) Thalamic modulation of high-frequency oscillating potentials in auditory cortex. *Nature* 383:78–81.
- Bartos M, Vida I, Jonas P (2007) Synaptic mechanisms of synchronized gamma oscillations in inhibitory interneuron networks. *Nat Rev Neurosci* 8:45–56.
- Behroozmand R, Oya H, Nourski KV, Kawasaki H, Larson CR, Brugge JF, Howard MA, Greenlee JDW (2016) Neural correlates of vocal production and motor control in human Heschl's gyrus. *J Neurosci* 36:2302–2315.
- Brosch M, Budinger E, Scheich H (2002) Stimulus-related gamma oscillations in primate auditory cortex. *J Neurophysiol* 87:2715–2725.
- Buzsaki G, Wang XJ (2012) Mechanisms of gamma oscillations. *Annu Rev Neurosci* 35:203–225.
- Cardin JA, Carlen M, Meletis K, Knoblich U, Zhang F, Deisseroth K, Tsai LH, Moore CI (2009) Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* 459:663–667.
- Chang EF, Niziolek CA, Knight RT, Nagarajan SS, Houde JF (2013) Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proc Natl Acad Sci USA* 110:2653–2658.
- Chen G, Zhang Y, Li X, Zhao X, Ye Q, Lin Y, Tao HW, Rasch MJ, Zhang X (2017) Distinct inhibitory circuits orchestrate cortical beta and gamma band oscillations. *Neuron* 96:1403–1418.e1406.
- Crone NE, Boatman D, Gordon B, Hao L (2001a) Induced electrocorticographic gamma activity during auditory perception: Brazier Award-winning article, 2001. *Clin Neurophysiol* 112:565–582. 2001.
- Crone NE, Hao L, Hart J Jr, Boatman D, Lesser RP, Irizarry R, Gordon B (2001b) Electrocorticographic gamma activity during word production in spoken and sign language. *Neurology* 57:2045–2053.
- Eckhorn R, Bauer R, Jordan W, Brosch M, Kruse W, Munk M, Reitboeck HJ (1988) Coherent oscillations: a mechanism of feature linking in the visual cortex? Multiple electrode and correlation analyses in the cat. *Biol Cybern* 60:121–130.
- Eliades SJ, Wang X (2003) Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *J Neurophysiol* 89:2194–2207.
- Eliades SJ, Wang X (2005) Dynamics of auditory-vocal interaction in monkey auditory cortex. *Cereb Cortex* 15:1510–1523.
- Eliades SJ, Wang X (2008a) Chronic multi-electrode neural recording in free-roaming monkeys. *J Neurosci Methods* 172:201–214.

- Eliades SJ, Wang X (2008b) Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* 453:1102–1106.
- Eliades SJ, Wang X (2012) Neural correlates of the Lombard effect in primate auditory cortex. *J Neurosci* 32:10737–10748.
- Eliades SJ, Wang X (2013) Comparison of auditory-vocal interactions across multiple types of vocalizations in marmoset auditory cortex. *J Neurophysiol* 109:1638–1657.
- Eliades SJ, Wang X (2017) Contributions of sensory tuning to auditory-vocal interactions in marmoset auditory cortex. *Hear Res* 348:98–111.
- Eliades SJ, Tsunada J (2018) Auditory cortical activity drives feedback-dependent vocal control in marmosets. *Nat Commun* 9:2540.
- Eliades SJ, Wang X (2019) Corollary discharge mechanisms during vocal production in marmoset monkeys. *Biol Psychiatry Cogn Neurosci Neuroimaging* 4:805–812.
- Flinker A, Chang EF, Kirsch HE, Barbaro NM, Crone NE, Knight RT (2010) Single-trial speech suppression of auditory cortex activity in humans. *J Neurosci* 30:16643–16650.
- Ford JM, Mathalon DH, Heinks T, Kalba S, Faustman WO, Roth WT (2001) Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. *Am J Psychiatry* 158:2069–2071.
- Fries P, Reynolds JH, Rorie AE, Desimone R (2001) Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* 291:1560–1563.
- Galambos R (1992) A comparison of certain gamma band (40 Hz) rhythms in cat and man. In: *Induced rhythms in the brain: brain dynamics*. (Basar E, Bullock TH, eds), pp 201–216. Boston: Birkhauser.
- Galambos R, Makeig S, Talmachoff PJ (1981) A 40-Hz auditory potential recorded from the human scalp. *Proc Natl Acad Sci USA* 78:2643–2647.
- Greenlee JD, Jackson AW, Chen F, Larson CR, Oya H, Kawasaki H, Chen H, Howard MA (2011) Human auditory cortical activation during self-vocalization. *PLoS One* 6:e14744.
- Greenlee JD, Behroozmand R, Larson CR, Jackson AW, Chen F, Hansen DR, Oya H, Kawasaki H, Howard MA (2013) Sensory-motor interactions for vocal pitch monitoring in non-primary human auditory cortex. *PLoS One* 8:e60783.
- Hage SR (2018) Auditory and audio-vocal responses of single neurons in the monkey ventral premotor cortex. *Hear Res* 366:82–89.
- Hage SR, Nieder A (2013) Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nat Commun* 4:2409.
- Hage SR, Nieder A (2015) Audio-vocal interaction in single neurons of the monkey ventrolateral prefrontal cortex. *J Neurosci* 35:7030–7040.
- Hage SR, Nieder A (2016) Dual neural network model for the evolution of speech and language. *Trends Neurosci* 39:813–829.
- Houde JF, Nagarajan SS (2011) Speech production as state feedback control. *Front Hum Neurosci* 5:82.
- Houde JF, Chang EF (2015) The cortical computations underlying feedback control in vocal production. *Curr Opin Neurobiol* 33:174–181.
- Jeschke M, Lenz D, Budinger E, Herrmann CS, Ohl FW (2008) Gamma oscillations in gerbil auditory cortex during a target-discrimination task reflect matches with short-term memory. *Brain Res* 1220:70–80.
- Joliot M, Ribary U, Llinas R (1994) Human oscillatory brain activity near 40 Hz coexists with cognitive temporal binding. *Proc Natl Acad Sci USA* 91:11748–11751.
- Lakatos P, Szilagy N, Pincze Z, Rajkai C, Ulbert I, Karmos G (2004) Attention and arousal related modulation of spontaneous gamma-activity in the auditory cortex of the cat. *Brain Res Cogn Brain Res* 19:1–9.
- Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53:279–292.
- Marshall L, Molle M, Bartsch P (1996) Event-related gamma band activity during passive and active oddball tasks. *Neuroreport* 7:1517–1520.
- Medvedev AV, Kanwal JS (2008) Communication call-evoked gamma-band activity in the auditory cortex of awake bats is modified by complex acoustic features. *Brain Res* 1188:76–86.
- Miltner WH, Braun C, Arnold M, Witte H, Taub E (1999) Coherence of gamma-band EEG activity as a basis for associative learning. *Nature* 397:434–436.
- Muller-Preuss P, Ploog D (1981) Inhibition of auditory cortical neurons during phonation. *Brain Res* 215:61–76.
- Murthy VN, Fetz EE (1992) Coherent 25- to 35-Hz oscillations in the sensorimotor cortex of awake behaving monkeys. *Proc Natl Acad Sci USA* 89:5670–5674.
- Niziolek CA, Nagarajan SS, Houde JF (2013) What does motor efference copy represent? Evidence from speech production. *J Neurosci* 33:16110–16116.
- Numminen J, Salmelin R, Hari R (1999) Subject's own speech reduces reactivity of the human auditory cortex. *Neurosci Lett* 265:119–122.
- O'Connell MN, Falchier A, McGinnis T, Schroeder CE, Lakatos P (2011) Dual mechanism of neuronal ensemble inhibition in primary auditory cortex. *Neuron* 69:805–817.
- Paus T, Perry DW, Zatorre RJ, Worsley KJ, Evans AC (1996) Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *Eur J Neurosci* 8:2236–2246.
- Rauschecker JP, Tian B (2004) Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. *J Neurophysiol* 91:2578–2589.
- Ray S, Maunsell JH (2011) Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol* 9:e1000610.
- Roy S, Zhao L, Wang X (2016) Distinct neural activities in premotor cortex during natural vocal behaviors in a new world primate, the common marmoset (*Callithrix jacchus*). *J Neurosci* 36:12168–12179.
- Rummell BP, Klee JL, Sigurdsson T (2016) Attenuation of responses to self-generated sounds in auditory cortical neurons. *J Neurosci* 36:12010–12026.
- Sanes JN, Donoghue JP (1993) Oscillations in local field potentials of the primate motor cortex during voluntary movement. *Proc Natl Acad Sci USA* 90:4470–4474.
- Schoffelen JM, Poort J, Oostenveld R, Fries P (2011) Selective movement preparation is subserved by selective increases in corticomuscular gamma-band coherence. *J Neurosci* 31:6750–6758.
- Singer W (1999) Neuronal synchrony: a versatile code for the definition of relations? *Neuron* 24:49–65. 111–125.
- Sohal VS, Zhang F, Yizhar O, Deisseroth K (2009) Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. *Nature* 459:698–702.
- Steinschneider M, Fishman YI, Arezzo JC (2008) Spectrotemporal analysis of evoked and induced electroencephalographic responses in primary auditory cortex (A1) of the awake monkey. *Cereb Cortex* 18:610–625.
- Sukov W, Barth DS (2001) Cellular mechanisms of thalamically evoked gamma oscillations in auditory cortex. *J Neurophysiol* 85:1235–1245.
- Super H, Roelfsema PR (2005) Chronic multiunit recordings in behaving animals: advantages and limitations. *Prog Brain Res* 147:263–282.
- Tallon-Baudry C, Bertrand O (1999) Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn Sci (Regul Ed)* 3:151–162.
- Tiitinen H, Sinkkonen J, Reinikainen K, Alho K, Lavikainen J, Naatanen R (1993) Selective attention enhances the auditory 40-Hz transient response in humans. *Nature* 364:59–60.
- Vianney-Rodrigues P, Iancu OD, Welsh JP (2011) Gamma oscillations in the auditory cortex of awake rats. *Eur J Neurosci* 33:119–129.
- Welle CG, Contreras D (2016) Sensory-driven and spontaneous gamma oscillations engage distinct cortical circuitry. *J Neurophysiol* 115:1821–1835.
- Whitford TJ (2019) Speaking-induced suppression of the auditory cortex in humans and its relevance to schizophrenia. *Biol Psychiatry Cogn Neurosci Neuroimaging* 4:791–804.
- Whittington MA, Traub RD (2003) Interneuron diversity series: inhibitory interneurons and network oscillations in vitro. *Trends Neurosci* 26:676–682.
- Womelsdorf T, Fries P (2007) The role of neuronal synchronization in selective attention. *Curr Opin Neurobiol* 17:154–160.