

Visual modulation of firing and spectrotemporal receptive fields in mouse auditory cortex

James Bigelow^{a,c}, Ryan J. Morrill^{a,b,c}, Timothy Olsen^{a,c}, Andrea R. Hasenstaub^{a,b,c,*}

^a Coleman Memorial Laboratory, University of California, San Francisco, USA

^b Neuroscience Graduate Program, University of California, San Francisco, USA

^c Department of Otolaryngology–Head and Neck Surgery, University of California, San Francisco, 94143, USA

ABSTRACT

Recent studies have established significant anatomical and functional connections between visual areas and primary auditory cortex (A1), which may be important for cognitive processes such as communication and spatial perception. These studies have raised two important questions: First, which cell populations in A1 respond to visual input and/or are influenced by visual context? Second, which aspects of sound encoding are affected by visual context? To address these questions, we recorded single-unit activity across cortical layers in awake mice during exposure to auditory and visual stimuli. Neurons responsive to visual stimuli were most prevalent in the deep cortical layers and included both excitatory and inhibitory cells. The overwhelming majority of these neurons also responded to sound, indicating unimodal visual neurons are rare in A1. Other neurons for which sound-evoked responses were modulated by visual context were similarly excitatory or inhibitory but more evenly distributed across cortical layers. These modulatory influences almost exclusively affected sustained sound-evoked firing rate (FR) responses or spectrotemporal receptive fields (STRFs); transient FR changes at stimulus onset were rarely modified by visual context. Neuron populations with visually modulated STRFs and sustained FR responses were mostly non-overlapping, suggesting spectrotemporal feature selectivity and overall excitability may be differentially sensitive to visual context. The effects of visual modulation were heterogeneous, increasing and decreasing STRF gain in roughly equal proportions of neurons. Our results indicate visual influences are surprisingly common and diversely expressed throughout layers and cell types in A1, affecting nearly one in five neurons overall.

1. Introduction

Environmental events are often transduced by multiple sensory modalities, subserving multisensory perceptual processes such as spatial localization and communication. Although multisensory integration was once believed to be mediated primarily by higher-order cortices, mounting evidence indicates that earlier stations, including primary sensory cortices, extensively integrate information from other sensory and motor areas (Wallace et al., 2004; Ghazanfar and Schroeder, 2006; Budinger and Scheich, 2009; Smiley and Falchier, 2009; Kajikawa et al., 2012; Stehberg et al., 2014; Schneider and Mooney, 2018; King et al., 2019). In primary auditory cortex (A1), visual influences have been widely observed in many species and preparations, likely reflecting the pervasiveness of audiovisual interaction in natural behavior (Wallace et al., 2004; Banks et al., 2011; Budinger and Scheich, 2009; Stehberg et al., 2014; King et al., 2019). For instance, anatomical studies report direct projections from visual areas (Banks et al., 2011; Budinger and Scheich, 2009; Stehberg et al., 2014) and physiological studies have observed neurons with visual responses or visually modulated responses to sound (Bizley et al., 2007; Bizley and King, 2008; Kayser et al., 2009,

2010; Kobayasi et al., 2013). These insights have left open two important questions: first, what is the microcircuitry through which visual inputs shape processing in A1, and which neurons participate in this process? And second, which aspects of sound encoding are affected by visual influences, and in what ways?

Understanding audiovisual microcircuitry in A1 requires determining which anatomically and functionally defined neuron populations are driven or modulated by visual input. Cortical neurons fall into many inhibitory and excitatory types, which are organized into layers, each with different patterns of intracolumnar, intercolumnar, and subcortical connectivity (Tremblay et al., 2016). Experiments *in vitro* have confirmed that visual projections directly activate both excitatory and inhibitory cells in A1 (Banks et al., 2011). However, it is not yet known whether these cell types are differentially responsive to visual stimulation *in vivo*, and whether the effects of visual stimulation on different cell types are cell type- or layer-specific. Second, while a recent study in ferrets suggested visually modulated neurons may be most common in the supragranular and infragranular layers (Atilgan et al., 2018), and another from our lab found visually responsive neurons in the mouse were most common in the infragranular layers

* Corresponding author Sandler Neurosciences Center, 675 Nelson Rising Lane, Box 0444, San Francisco, CA, 94143, USA.

E-mail address: Andrea.Hasenstaub@ucsf.edu (A.R. Hasenstaub).

<https://doi.org/10.1016/j.crneur.2022.100040>

Received 8 February 2022; Received in revised form 26 April 2022; Accepted 6 May 2022

Available online 18 May 2022

2665-945X/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

(Morrill and Hasenstaub, 2018), we do not yet know whether visually responsive and visually modulated neurons are drawn from the same populations or reflect distinct subpopulations. Similarly, it remains unclear whether visually responsive infragranular neurons are part of the circuitry underlying sound processing in A1, or alternatively comprise a distinct subpopulation relaying visual information to auditory neurons. Therefore, determining how both visual responses and visual modulation are distributed among the layers and cell types, and determining which of these cell populations are also responsive to sound, will provide insight into the circuitry mediating audiovisual integration in A1.

Understanding which aspects of sound encoding are modulated by visual inputs is similarly essential to understanding the nature and extent of visual influences in A1. In A1, sound information can be encoded both in transient (at stimulus onset) and sustained changes in firing rate (FR). Onset responses are often stronger than sustained responses, and reflect stronger feedforward contributions from subcortical structures, while sustained responses are weaker and more closely tied to local network processing. This implies that onset and sustained responses may be differentially susceptible to contextual, including visual, influences. Beyond spike rate, A1 neurons often encode sound features and other events through changes in spike timing, with or without changes in average FR (deCharms and Merzenich, 1996; Malone et al., 2010; Insanally et al., 2019). Spectrotemporal receptive fields (STRFs) can be used to measure such responses because they reflect average stimulus values preceding each spike and are thus sensitive to both spike rate and timing (Atencio and Schreiner, 2013; Wu et al., 2006). Previous studies have shown that contextual variables such as attentional state may modulate sensitivity to spectrotemporal features (i.e., may cause STRF changes) without affecting time averaged FRs, or vice versa (Slee and David, 2015). Thus, assessing visual influences on these distinct sound response types is needed to understand both the prevalence of visual influences in A1 and their functional consequences.

Here, we examine how visual inputs modulate different aspects of spontaneous and sound-evoked activity in excitatory and inhibitory neurons in different layers of awake mouse A1. We used high-density translaminar probes to estimate the cortical depth of each neuron and used spike shapes to classify putative inhibitory and excitatory neurons. By presenting receptive field estimation stimuli in segments, we were able to simultaneously measure STRFs as well as transient/onset and sustained FR responses to sound. On half of the trials, we simultaneously presented visual stimuli to determine whether these responses were sensitive to visual context.

2. Materials and Methods

2.1. Subjects and surgical preparation

All procedures were carried out in AAALAC approved facilities, were approved by the Institutional Animal Care and Use Committee at the University of California, San Francisco, and complied with ARRIVE guidelines and recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. A total of 15 adult mice (6 female) served as subjects (median age 99 days, range 58–169 days). All mice had a C57BL/6 background and expressed optogenetic effectors targeting interneuron subpopulations, which were not manipulated in the current study. The mice were generated by crossing the Ai32 strain expressing Cre-dependent eYFP-tagged channelrhodopsin-2 (JAX stock #012569) with neuron-specific Cre driver lines (PV-Cre: JAX stock #012358; Sst-Cre: JAX stock #013044; VIP-Cre: JAX stock #010908; Ctgf-Cre: JAX stock # 028535; Penk-Cre: JAX stock #025112; NtsR1-Cre: MMRRC UCD stock #030648-UCD). Mice were housed in groups of two to five under a 12H–12H light-dark cycle. Surgical procedures were performed under isoflurane anesthesia with perioperative analgesics (lidocaine, meloxicam, and buprenorphine) and monitoring. A custom stainless steel headbar was affixed to the cranium above the right temporal lobe with dental cement, after

which subjects were allowed to recover for at least two days. Prior to electrophysiological recording, a small craniotomy (~1–2 mm diameter) centered above auditory cortex (~2.5–3.5 mm posterior to bregma and under the squamosal ridge) was made within a window opening in the headbar. The craniotomy was then sealed with silicone elastomer (Kwik-Cast, World Precision Instruments). The animal was observed until ambulatory (~5–10 min) and allowed to recover for a minimum of 2 h prior to electrophysiological recording. The craniotomy was again sealed with silicone elastomer after recording, and the animal was housed alone thereafter. Electrophysiological recordings were conducted up to five consecutive days following the initial craniotomy.

2.2. Auditory and visual stimuli

All stimuli were generated in MATLAB (Mathworks) and delivered using Psychophysics Toolbox Version 3 (Kleiner et al., 2007). Sounds were delivered through a free-field electrostatic speaker (ES1, Tucker-Davis Technologies) approximately 15–20 cm from the left (contralateral) ear using an external soundcard (Quad Capture, Roland) at a sample rate of 192 kHz. Sound levels were calibrated to 60 ± 5 dB at ear position (Model 2209 m, Model 4939 microphone, Brüel & Kjær). Visual stimuli were presented on a 19-inch LCD monitor with a 60 Hz refresh rate (ASUS VW199 or Dell P2016t) centered 25 cm in front of the mouse. Monitor luminance was calibrated to 25 cd/m^2 for 50% gray at eye position.

Search stimuli used for cortical depth estimation included click trains, noise bursts, and pure tone pips, plus the experimental stimuli described below. For a few recordings, only tone pips and experimental stimuli were presented due to time constraints. In some recordings, additional search stimuli were presented, such as frequency-modulated sweeps. Click trains comprised broadband 5 ms non-ramped white noise pulses presented at 4 Hz for 1 s at 60 dB with a ~1 s interstimulus interval (ISI), with 20–50 repetitions. Noise bursts consisted of 50 ms non-ramped band-passed noise with a uniform spectral distribution between 4 and 64 kHz presented at 60 dB in 500 unique trials with a ~350 ms ISI. Pure tones consisted of 100 ms sinusoids with 5-ms cosine-squared onset/offset ramps presented at a range of frequencies (4–64 kHz, 0.2 octave spacing) and attenuation levels (30–60 dB, 5 dB steps). Three repetitions of each frequency-attenuation combination were presented in pseudorandom order with an ISI of ~550 ms. Peristimulus-time histograms (PSTHs) quantifying time-binned multi-unit FRs were constructed for each stimulus. For tone pips, frequency-response area (FRA) functions were constructed from baseline-subtracted spike counts during the stimulus period averaged across trials at each frequency-attenuation combination. PSTHs and FRAs from an example recording are shown in Fig. 1C.

As depicted by Fig. 2, experiments included two trial types: (a) auditory trials, which presented sound only, and (b) audiovisual trials, which included both sound and visual stimuli. For both trial types, the auditory stimulus was a random double sweep (RDS), a continuous, spectrally sparse receptive field estimation stimulus capable of effectively driving activity across diversely tuned neurons in A1 (Gourévitch et al., 2015). The RDS comprised two uncorrelated random sweeps, each varying continuously over time between 4 and 64 kHz, with a maximum sweep modulation frequency of 20 Hz. Sample RDS frequency vectors are depicted in Fig. 2A and B. The RDS was delivered in 15-s non-repeating segments (40 trials, 10 min total stimulation; cf. Rutkowski et al., 2002). The inter-sound interval was ~9 s, with visual stimuli trailing and leading sounds by 2.5 s within this interval. Thus, intertrial intervals were ~9 s for consecutive auditory trials, ~4 s for consecutive audiovisual trials, and ~6.5 s for mixed trial sequences. The same 40 RDS segments were used for auditory and audiovisual trials to maintain identical stimulus statistics between conditions. Audiovisual trials were thus identical to auditory trials except for an additional visual contrast modulated noise (CMN) stimulus (Fig. 2B).

As described elsewhere (Niell and Stryker, 2008; Piscopo et al.,

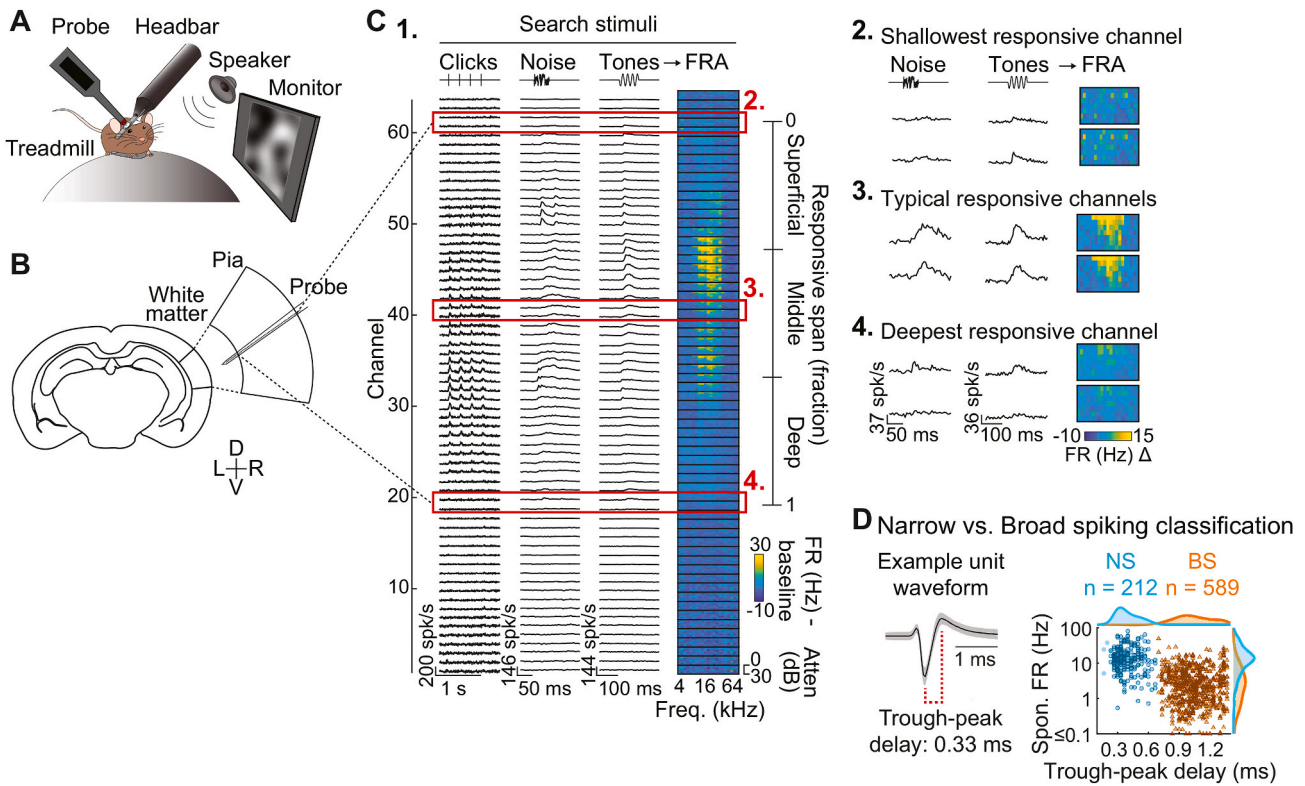


Fig. 1. Translaminar single unit recording in awake mouse auditory cortex. **(A)** Mice were head fixed atop a spherical treadmill. **(B)** Translaminar probes recorded neuronal activity across cortical layers. **(C)** Auditory cortical depth estimation. **(1.)** The putative span of cortex was estimated based on the span of channels with multiunit responses to search stimuli. Single units were then classified as Superficial, Middle, or Deep according to their relative depth within this span. **(2–4)** Magnified view of example responsive channels. **(D)** Classification of single units as narrow (putative inhibitory) or broad spiking (predominantly excitatory). Left: Example spike waveform (median ± MAD). Right: Narrow and broad spiking units clustered by trough-peak delay and FR.

2013), CMN is a broadband stimulus designed to maximally activate diversely tuned neurons in visual cortex. The stimulus is generated by first creating a random frequency spectrum in the Fourier domain.

The temporal frequency spectrum was flat with a low-pass cutoff at 10 Hz. The spatial frequency spectrum dropped off as $A(f) \sim 1/(f + f_c)$, with $f_c = 0.05$ cycles/°, where f is spatial frequency and A denotes the amplitude envelope. The spatial frequency content of the stimulus was thus biased toward low frequencies, approximating the spatial frequency preference distribution of visual cortical neurons (Niell and Stryker, 2008; Piscopo et al., 2013). A spatiotemporal movie was then created by inverting the three-dimensional spectrum. Finally, contrast modulation was imposed by multiplying the movie by a sinusoidally variable contrast function. The CMN stimulus was generated at 60×60 pixels, then interpolated to 900×900 pixels. The first and last frames of the CMN movie were uniform 50% gray, providing abrupt luminance changes at stimulus onset and offset from black during the intertrial interval. The CMN stimulus led and trailed the RDS sounds by 2.5 s to allow ample time for potential visual-evoked spiking responses to reach an adapted state prior to sound onset responses and persist throughout sound offset responses.

2.3. Electrophysiology

Recordings were conducted inside a sound attenuation chamber (Industrial Acoustics Company). Because anesthesia strongly influences auditory cortical activity (Wang et al., 2005; Hromádka et al., 2008; Kobayasi et al., 2013; Morrill and Hasenstaub, 2018), we conducted all recordings in awake, headfixed animals moving freely atop a spherical treadmill (Fig. 1A; Dombek et al., 2007; Niell and Stryker, 2010; Phillips and Hasenstaub, 2016; Phillips et al., 2017a, 2017b; Morrill and Hasenstaub, 2018; Bigelow et al., 2019). Recordings were made with

single shank, linear multichannel electrode arrays slowly lowered into cortex using a motorized microdrive (FHC). Arrays with 64 channels (Cambridge Neurotech H3; 20 μm site spacing, 1260 μm total span) were used for all recordings except one, which used a 32-channel array (Cambridge Neurotech H4; 25 μm site spacing, 775 μm total span). Prior to lowering the probe, the craniotomy was filled with 2% agarose to stabilize the brain surface. After reaching depths of approximately 800–1000 μm below the first observed action potentials, probes were allowed to settle for at least 20 min before recording. Continuous extracellular voltage traces were collected using an Intan RHD2164 amplifier chip with RHD USB Interface Board (Intan Technologies) at a sample rate of 30 kHz. Other events such as stimulus times were stored concurrently by the same system. All analyses were performed with MATLAB (Mathworks).

The span of recording channels (1260 μm) exceeded mouse cortical depth (~800 μm; Chang and Kawai, 2018; Paxinos and Franklin, 2019), typically resulting in a span of sound-responsive channels plus an additional subset of non-responsive channels outside of A1. As depicted in Fig. 1C, we used multi-unit responses evoked by search stimuli to designate the span of sound-responsive channels, which served as an estimate of cortical span. This was normalized by expressing each as a fraction of total depth (Morrill and Hasenstaub, 2018) and divided into three equal bins reflecting superficial, middle, and deep-layer neuron populations. In our prior study of visual responses in A1, we applied Di-I to recording probes for histologically referenced depth estimation (Morrill and Hasenstaub, 2018). We observed parallel depth distributions of visually responsive neurons in the current and prior studies, suggesting the current method achieved a rough approximation to the histological approach.

We targeted recordings in A1 using stereotaxic coordinates and anatomical landmarks such as characteristic vasculature patterns

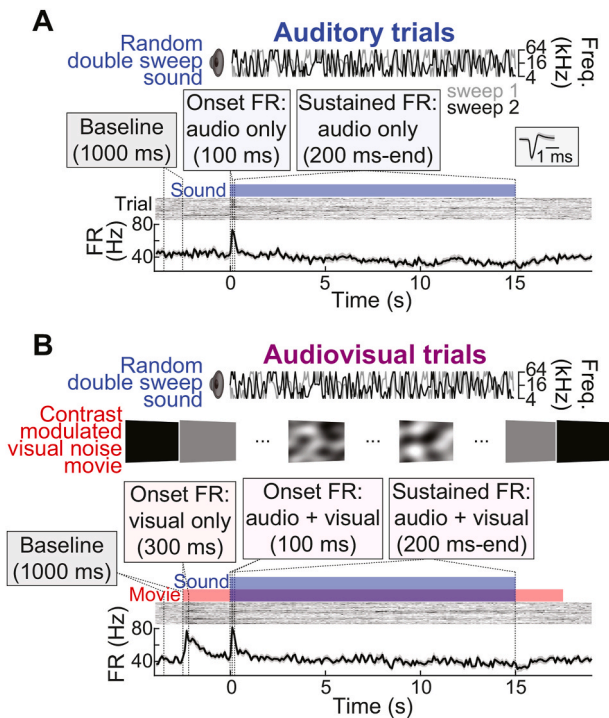


Fig. 2. Interleaved auditory and audiovisual trials. (A) Auditory trials comprised non-repeating 15-s RDS sounds separated by silent intertrial intervals. Example unit spiking activity is shown in PSTHs (lower) and binned-spike count matrices (upper). Baseline and evoked FR responses (onset, sustained) were measured within the indicated windows. Inset shows the unit spike waveform (median \pm MAD). (B) Audiovisual trials were identical to Auditory trials with the addition of a visual CMN stimulus.

(Joachimsthaler et al., 2014). Previous studies have reported significant differences in tone onset latencies between primary and non-primary fields (Joachimsthaler et al., 2014), with latencies between 5 and 18 ms for primary fields (median \sim 9 ms), and 8–32 ms for non-primary fields (median \sim 12–16 ms). We thus analyzed tone onset latencies to support designations of putative primary recording sites. Using multi-unit PSTHs with 2-ms bins and smoothed with a Savitzky-Golay filter (3rd order, 10-ms window), we defined onset latency as the first post-onset bin exceeding 2.5 standard deviations of the pre-stimulus FR (Morrill and Hasenstaub, 2018). We retained only putative primary sites for further analysis, defined as recordings for which the median latency across the responsive channel span was 14 ms or less (49 of 60 total recordings). The retained putative primary recordings universally exhibited robust multi-unit responses to click trains, noise bursts, and tone pips, and clear evidence of frequently-level tuning in the FRA plots (Fig. 1C) as well as spectrotemporal tuning in single-unit responses to RDS stimuli (Fig. 5).

To isolate single-unit activity from continuous multichannel traces, we used Kilosort 2.0 (Pachitariu et al., 2016; available: <https://github.com/MouseLand/Kilosort>) combined with auto- and cross-correlation analysis, refractory period analysis, and cluster isolation statistics. Although the majority of isolated units were held throughout the entire recording (\sim 28 min), isolation of individual neurons was occasionally disrupted or lost partway through the experiment. Thus, we estimated the active timespan for each unit by visual inspecting unit activity plots over time. For the retained subset of active trials, we matched RDS stimuli between conditions by analyzing only RDS segments common to both conditions. This ensured strict equivalence of auditory stimuli between conditions, isolating any observed differences to the presence of the visual CMN stimulus. We included only units with 10 or more active trials (5 per condition) in the final analysis. In total, 801 units were included in the analyses below. As in previous publications (Phillips

et al., 2017a; Bigelow et al., 2019), we were able to classify units as narrow-spiking (NS) or broad-spiking (BS) using a clear bimodal distribution of waveform trough-peak delays (Fig. 1D; NS, $<600 \mu\text{s}$, $n = 212$; BS, $\geq 600 \mu\text{s}$, $n = 589$). As reported elsewhere (Kawaguchi and Kubota, 1997; Niell and Stryker, 2008; Wu et al., 2008; Moore and Wehr, 2013; Li et al., 2015; Yu et al., 2019; Liu et al., 2020), NS units overwhelmingly reflect inhibitory neurons, the majority of which are parvalbumin positive (PV+) interneurons (a minority are somatostatin [SST+] inhibitory interneurons). The majority of BS units are excitatory neurons, although some are known to be SST+ and vasoactive intestinal peptide (VIP+) inhibitory interneurons. NS and BS units thus reflect neuron subpopulations dominated by inhibitory and excitatory neurons, respectively. Consistent with this classification, NS units had characteristically higher mean spontaneous FRs than BS units (BS: 3.94 Hz, NS: 16.84 Hz; Hromádka et al., 2008; Schiff and Reyes, 2012; spontaneous rates estimated from the baseline period shown in Fig. 2A; Wilcoxon Rank Sum test: $p < 10^{-61}$). For the spike-equant analysis of visual responses reported in Fig. S2, we calculated mean FRs using both spontaneous and driven spikes (including the full stimulus period plus spontaneous activity 1 s before and after the stimulus), yielding population mean rates of 18.11 and 4.45 Hz for NS and BS units, respectively.

2.4. Spectrotemporal receptive field estimation

We calculated spectrotemporal fields (STRFs) using standard reverse-correlation techniques (Wu et al., 2006) as depicted in Fig. 4. RDS stimuli were discretized in 1/8 oct frequency bins and 5-ms time bins, sufficient resolution for modeling response properties of most A1 neurons (Thorson et al., 2015). We first calculated the spike-triggered average (STA) by adding the discretized stimulus segment preceding each spike to a cumulative total, and then dividing by the total spike count. For all data analyses, the peri-spike time analysis window spanned 0–100 ms prior to spike event times, sufficient for capturing latency-adjusted temporal response periods of most A1 neurons (Atencio and Schreiner, 2013; See et al., 2018). We use a broader window for presenting example data, spanning 200 ms before and 50 ms after spike event times. The 50 ms post-spike window indicates a causal values, i.e., those that would be expected by chance given the finite recording time, stimulus and spike timing statistics, and smoothing parameters (Gourévitch et al., 2015). The first 200 ms of the RDS response from each trial were dropped from all STA calculations analyses to minimize bias reflecting strong onset transients.

The STA thus reflects the average binned stimulus segment preceding spike events and can be viewed as a linear approximation to the optimal stimulus for driving neuronal firing (deCharms et al., 1998). As discussed by Rutkowski et al. (2002), the STA can be formalized as the probability (P) of a stimulus frequency f occurring at time $t_i - \tau$ given that a spike occurred, as expressed by the equation

$$P(S[f, \tau] | i) = \left(\sum_i^n S(f, t_i - \tau) \bullet \delta(t_i) \right) / n \quad \text{Eq. (1)}$$

where i indicates a spike, t_i is a spiketime, τ is the time analysis window, n is the spike count, and Σ indicates summing across spikes. $S(f, t)$ is the stimulus value at a given time-frequency bin, equaling one if an RDS frequency intersects the bin, two if both RDS frequencies coincide with the bin, and zero otherwise. $S(f, t_i - \tau)$ represents the windowed stimulus aligned to a spike time. $\delta(t_i)$ is equal to one if a spike occurs at time t_i and zero otherwise. With a minor modification, STRF time-frequency bins can be expressed in terms of deviation from mean driven FR (spikes/s - mean) using the terms defining the STA and Bayes' theorem,

$$P(i | S[f, \tau]) = P(S[f, \tau] | i) \bullet P(i) / P(S[f]) \quad \text{Eq. (2)}$$

where $P(i)$ is the probability of a spike occurring in a bin, equal to n_i/T , where T is the total stimulus time and $P(S[f])$ is the probability of a tone

frequency occurring in a bin. The mean driven FR is then subtracted from the STRF such that individual time-frequency bins reflect increases (positive, red) or decreases (negative, blue) from the mean driven rate (Fig. 4C). Finally, we used a uniform 3×3 bin window for smoothing STRFs to reduce overfitting to finite-sampled stimulus statistics. The STA and STRF expressed in units of spikes/s are multiples of each other since terms in the expression $P(i)/P(S|f)$ are constant and thus practically identical for the purposes of all data analyses, including comparisons between conditions. However, we opted to report STRFs represented in firing-rate change units to ease interpretation of stimulus driven changes in neuronal activity.

We calculated STRFs independently for each condition. As in previous studies (Fritz et al., 2003), we further calculated a difference STRF (Δ STRF) by subtracting the auditory STRF from the audiovisual STRF. In addition, we calculated ‘null’ STRFs using identical procedures to those described above except that the stimulus was reversed in time, while preserving the original spike event times. This modification breaks the temporal relationship between the stimulus and spike times, but preserves spike count and timing statistics (e.g., interspike interval distribution), as well as the statistical distributions of the stimulus (Bigelow and Malone, 2017). The resulting STRFs were used to estimate time-frequency bin values expected by chance, and thus assess the statistical significance of STRFs as described below (Statistical analysis). Null Δ STRFs were similarly obtained by subtracting null auditory STRF from the null audiovisual STRF.

2.5. Statistical analysis

As indicated by Fig. 2A, we measured spontaneous FRs used a window spanning 3.5 to 2.5 s prior to sound onset (immediately preceding visual stimulus onset for audiovisual trials). We measured sound onset FR responses using spikes occurring within the first 100 ms of the stimulus. We used an analysis window spanning 200 ms post stimulus onset to the end of the stimulus (15 s) for both sustained FRs and STRFs. Finally, we measured visual onset FR responses within 300 ms post visual stimulus onset, which accommodated the longer response latencies typical of visual responses in A1 (Morrill and Hasenstaub, 2018).

We further analyzed auditory and visual offset FR responses using 100- and 300-ms windows following stimulus offset, respectively. Unlike onset and sustained responses, offset responses required comparison against two baselines: first, the spontaneous window described above, and second, a window of equivalent duration (1 s) preceding the offset response (i.e., the final second of the stimulus). Differences from the second baseline ensured offset responses did not merely reflect the continuation of a sustained response. In addition, we only considered offset responses ‘significant’ if they were above or below both baselines (but not between them) to ensure offset responses did not simply reflect the return of a sustained FR change to spontaneous activity. We used the larger of the two p-values associated with these tests to assess significance of the offset response, and its associated baseline for calculating effect size described below.

For each unit, we assessed the significance of all FR responses by comparison to baseline using Wilcoxon signed-rank tests (paired) with $\alpha = 0.05$. For all tests of individual unit significance, we controlled false discovery rate (FDR) by implementing the Benjamini–Hochberg procedure with $q = 0.05$ across the unit population (Benjamini and Hochberg, 1995). For a standardized measure of response strength that accommodated both increases and decreases in FR from baseline, we defined effect sizes as the absolute difference between the evoked and spontaneous FR means, divided by the standard deviation of the spontaneous FR. This facilitated interpretation of response deviations from chance reflecting different analysis windows and unit types. We similarly assessed significant FR differences between conditions with Wilcoxon signed-rank tests ($\alpha = 0.05$, Benjamini–Hochberg FDR correction with $q = 0.05$), and estimated effect sizes as the absolute difference between conditions divided by the standard deviation of the auditory

condition.

To assess the significance of STRFs, we calculated correlation coefficients between two STRFs computed from random trial halves (Fig. S1; Escabí et al., 2014) and defined a reliability index as the mean across 1000 subsample iterations. We calculated p-values reflecting the proportion of the null distribution exceeding the mean of the data distribution (Fig. S1A), and adjusted p-values equal to zero (cases where none of the null correlations exceeded the data mean) to 0.000999 reflecting the resolution permitted by the number of subsample iterations. Finally, we multiplied these p-values by two for an estimate of two-tailed significance and adjusted using Benjamini–Hochberg FDR correction ($q = 0.05$). Because the null subsampled STRF distribution was skewed for some units (e.g., null reliability index >0), we further considered STRFs significant only if reliability was greater than 0.2. Similar to FR responses, we estimated STRF reliability effect sizes as the absolute difference between the data and null distributions, divided by the standard deviation of the null distribution. We assessed the significance of Δ STRFs using the same approach, except using subsampled Δ STRF correlation distributions (Fig. S1B).

We used two approaches for assessing differences in responsiveness and modulation at the neuron subpopulation level (NS and BS units, and units in different cortical depth bins). First, we used chi-square tests to determine whether the proportion of units with significant responses ($p < 0.05$, FDR corrected) differed among groups. Second, we used independent one-way analysis of variance (ANOVA) to determine whether effect sizes for units with significant responses differed by group. For uniformity in presenting the results, we use the same approach for testing between two groups (NS vs. BS units), wherein ANOVA and the Student’s *t*-test produce equivalent p-values with $F = t^2$. We considered each neuron subpopulation as independent (NS vs. BS units), and therefore did not report main effects reflecting pooling across subpopulations. The rationale for including both tests was that neuron subpopulations (e.g., NS, BS) could have different proportions of responsive units with similar response strength or vice versa.

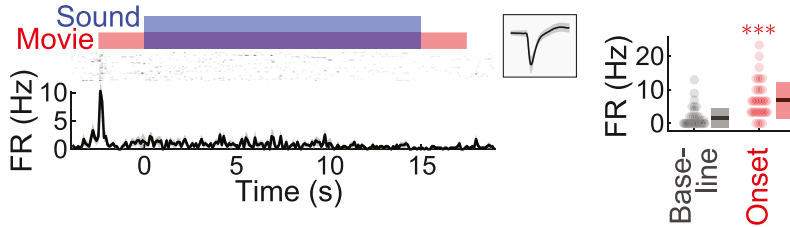
3. Results

Here, we presented continuous random sounds, continuous random movies, and sound-movie combinations (Fig. 1) to address two sets of questions regarding visually driven or modulated firing in awake mouse A1. First, we sought to resolve whether visually responsive neurons are part of the sound-responsive majority in A1, and if so, whether sound responses in these neurons are modulated by visual stimulation. Second, we sought to resolve whether visual context differently influences the diverse ways in which neurons respond to sound (transient onset firing, sustained firing, feature selectivity revealed by STRFs). In each analysis, we used cortical depth estimation and waveform classification techniques to identify differences in visual responses between putative inhibitory (NS) and excitatory (BS) neurons at different cortical depths (Fig. 1C and D).

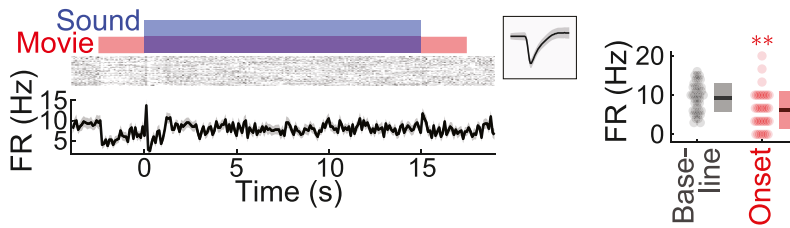
3.1. Visually responsive neurons in A1 are most prevalent in deep layers

We first measured the depth distribution of neurons responsive to visual stimulation alone (i.e., neurons for which FR significantly changes, either up or down, in the 300 ms following movie onset). In the middle layers, visual stimuli could either increase or suppress firing (e.g., Fig. 3C; excited: 17/29, 58.6%; suppressed: 12/29, 41.4%). In the deep layers, in contrast, visual responses were relatively common and almost exclusively excitatory (44/47, 93.6%). In total, roughly one in ten neurons were visually responsive (79/801, 9.9%), and consistent with our earlier study (Morrill and Hasenstaub, 2018), nearly two thirds of these units (47/79, 59.5%) were found in the depth bins corresponding to the deep layers of cortex (Fig. 3C). Chi-square tests confirmed the proportion of significant visual responses depended on depth for both unit types (NS: $\chi^2 = 18.05$, $p < 10^{-3}$, $\phi = 0.150$; BS: $\chi^2 =$

A Example unit: Visual onsets increase firing



B Example unit: Visual onsets suppress firing



C Visually responsive units are concentrated in deep layers

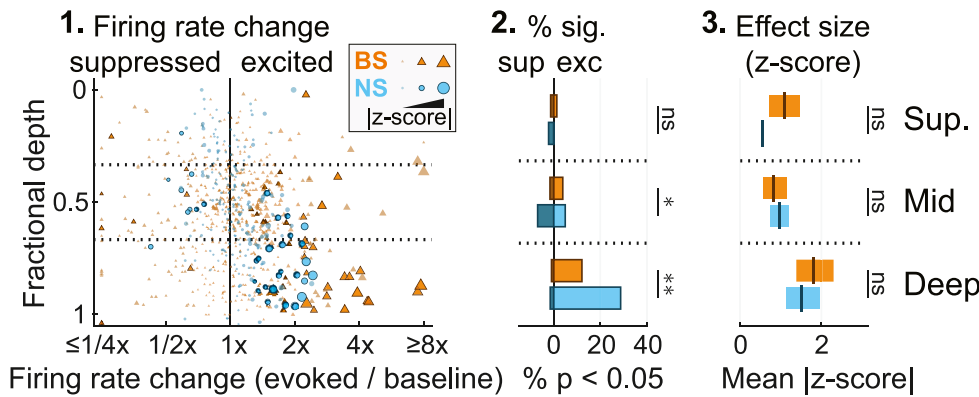


Fig. 3. Visually responsive neurons in A1. (A) Example unit with excited visual onset response. Left: Spiking response relative to the visual stimulus with unit spike waveform. Right: Baseline and visual onset FR responses (dots: single trials; boxes: mean \pm SD). (B) Example unit with suppressed visual onset response. (C) Population summary of visual onset FR responses. (1.) Visual onset FR changes from baseline for each unit by fractional depth. Outlined markers indicate significant responses ($p < 0.05$, FDR corrected). (2.) Percentages of units with significant excited and suppressed responses. (3.) Mean effect size (plus 99% confidence interval) for units with significant responses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns $p > 0.05$.

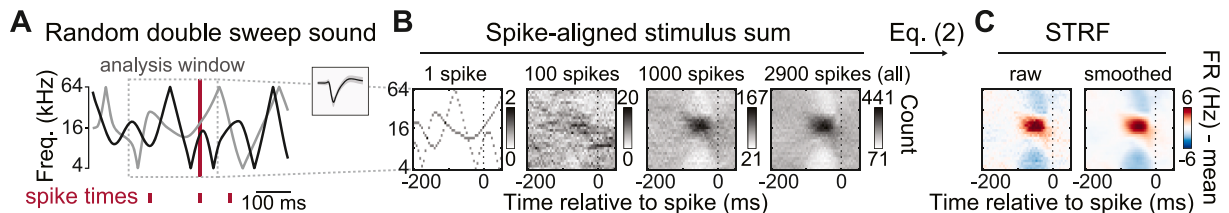
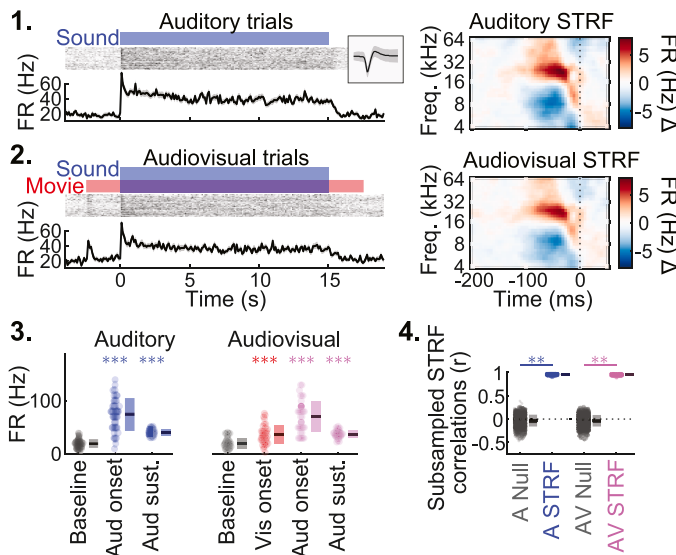


Fig. 4. Spectrotremoral receptive field (STRF) calculation. (A) Example RDS segment. (B) RDS segments aligned to each spike were added to a cumulative sum. (C) Transforming the spike-aligned stimulus sum according to Eq. (2) yielded an STRF estimate reflecting increases (red) and decreases (blue) in FR relative to the mean. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

13.03, $p = 0.0015$, $\phi = 0.128$). To determine whether visual responses were stronger in the deeper layers, we next measured the effect size, defined as the absolute difference between the visual-evoked and spontaneous FR means divided by the standard deviation of the spontaneous FR. This provided a standardized measure of visual response strength that does not over-weight units with high FRs, and that accommodates both increases and decreases in FR. One-way ANOVA confirmed that visual response effect sizes for units with significant responses depended on cortical depth for both unit types, with the strongest responses in the deepest bin (NS: $F = 3.82$, $p = 0.034$, $\eta^2 = 0.203$; BS: $F = 8.30$, $p < 10^{-3}$, $\eta^2 = 0.279$). Thus, both the prevalence of visually responsive neurons and the strength of their responses were greatest in the deep cortical layers.

We further tested whether visual responsiveness differed between NS and BS units. We observed visual responses in a significantly larger percentage of NS units (33/212, 15.6%) than BS units (46/589, 7.8%), but among the visually responsive units, the strength of the visual responses was not significantly different between NS and BS units (all F -ratios < 1.4 , all p -values > 0.24). As noted in the Materials and Methods (Fig. 1D), NS cells typically fired at much higher rates than BS cells, raising the possibility that the larger percentage of visually responsive NS units could be explained by the concomitant increase in statistical power. To test this, we randomly subsampled spikes from NS units with FRs above the BS mean (see Materials and Methods), creating a pseudo-population of NS units with mean FR equivalent to BS units (4.45 Hz; Fig. S2A). We repeated the visual onset response analyses for this

A Example unit: Visually responsive unit with auditory FR responses (onset, sustained) and STRF



B Most visually responsive neurons have significant responses to sound

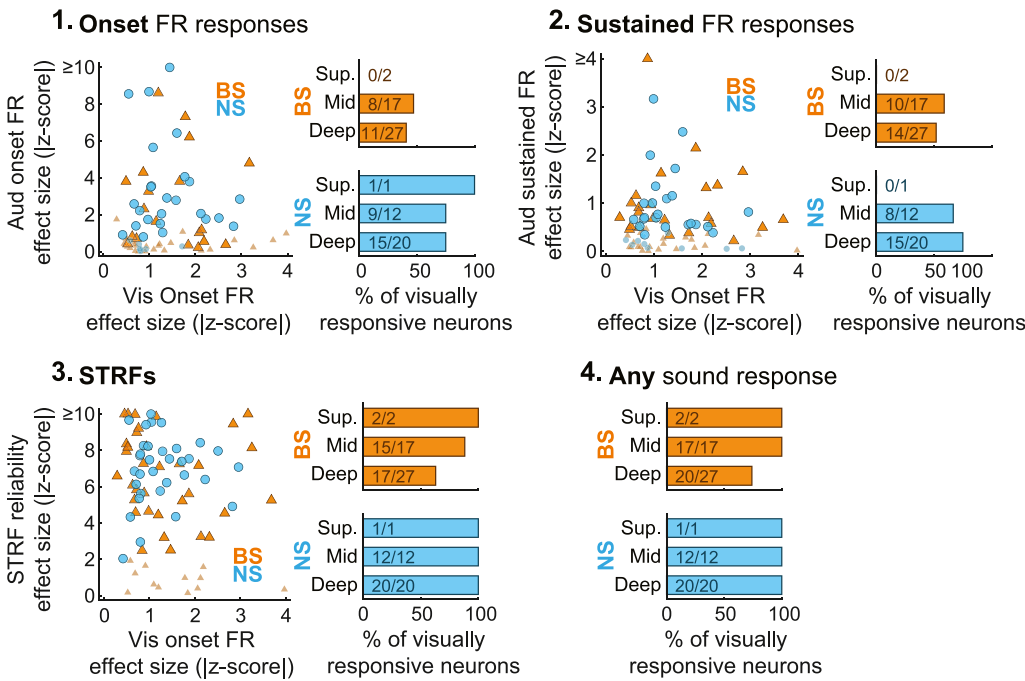


Fig. 5. Most visually responsive neurons in A1 also respond to sound. (A) Example unit with visual and auditory responses. (1–2.) Spiking activity and STRFs for auditory and audiovisual trials. (3.) Baseline and stimulus-evoked FR responses for each trial type (dots: single trials; boxes: mean \pm SD). (4.) Subsampled STRF correlations for each trial type (dots: single subsample iterations; boxes: mean \pm SD). (B) Auditory responses in visually responsive neurons. (1.) Auditory onset FR responses in visually responsive neurons. Left: Effect size for visual and auditory onset FR changes from baseline. Outlined markers indicate units with significant sound responses ($p < 0.05$, FDR corrected). Right: Percentages of visually responsive neurons with significant auditory onset firing responses. (2.) Auditory sustained FR responses in visually responsive neurons. (3.) Auditory STRFs in visually responsive neurons. (4.) Percentages of visually responsive neurons with any significant auditory response (onset FR, sustained FR, or STRF). ** $p < 0.01$, *** $p < 0.001$.

pseudo-population, and readjusted the p-values for the BS and pseudo-NS cells together using Benjamini–Hochberg FDR control. Following this adjustment, there were no longer any significant differences in the proportions of visually responsive NS and BS units (Fig. S2B; all χ^2 statistics < 1.24 , all p-values > 0.26), or in response strength (all F-ratios < 2.1 , all p-values > 0.17). Thus, we do not conclude that NS units are more likely to respond to visual stimulation. Rather, we conclude that visually evoked firing differences are simply easier to detect in NS units because of their generally higher spike rates.

Because stimulus offset responses have been observed in primary visual cortex (e.g., Liang et al., 2008), we also looked for significant FR changes following termination of the visual CMN stimulus. However, we found that no units had significant visual offset responses after FDR correction. We therefore defined visually responsive units throughout

the remainder of these analyses based on their visual onset responses only.

3.2. Most visually responsive neurons in A1 also respond to sound

We next determined whether the visually responsive neurons in our sample were primarily unimodal (visual only) or bimodal (visual plus auditory). To do this, we measured the proportions of units with significant visual onset responses that also responded to sound in each of three different ways: sound-onset FR responses, sustained FR responses, or STRFs (with or without concomitant changes in FR). We did not separately consider offset responses in this analysis because they were uncommon (NS: 15/212, 7.1%; BS: 27/589, 4.6%) and almost invariably occurred with at least one other sound response type (NS: 15/15, 100%;

BS: 23/27, 85.2%). As shown in Fig. 5B1–2, roughly half of all visually responsive BS units and three fourths of visually responsive NS units also had significant sound-evoked onset or sustained FR responses (BS onset: 19/46, 41.3%; BS sustained: 24/46, 52.2%; NS onset: 25/33, 75.8%; NS sustained: 23/33, 69.7%). The great majority of visually responsive neurons also had significant STRFs (Fig. 5B3; BS: 34/46, 73.9%; NS: 33/33, 100%). Thus, the overwhelming majority of visually responsive neurons also responded to sound in at least one of these three ways (Fig. 3B4; BS: 39/46, 84.8%; NS: 33/33, 100%), suggesting most visually responsive neurons in A1 also process sound.

We noted that many visually responsive neurons had significant STRFs but did not show a significant change in FR during the sound presentation. This implied that sound.

Responses in the general A1 neuron population may not always be evident from averaged FR responses. We confirmed this possibility by comparing the proportion of all neurons in the dataset with significant FR responses and/or STRFs (Fig. S3). We observed significant sound onset FR responses in just over one third of BS units (231/589, 39.2%), and two thirds of NS units (148/212, 69.8%). Just under half of both unit types had significant sustained FR responses (BS: 266/589, 45.2%; NS: 100/212, 47.2%). In contrast, the vast majority of units showed significant STRFs, even those lacking significant average FR changes (BS: 483/589, 82.0%; NS: 199/212, 93.9%). Examples of significant STRFs with and without significant FR responses are shown in Figs. S3A and S3B, respectively. These findings clarify that A1 neurons in our dataset were generally responsive to sound with very few exceptions (BS: 527/589, 89.5%; NS: 207/212, 97.6%), and underscore that spike timing-based analyses can reveal responses which do not produce changes in overall spike rate.

3.3. Visual signals differentially modulate sound onset and sustained FR responses

We next investigated whether distinct sound response types were differentially modulated by visual context by comparing changes in onset and sustained FR responses between conditions (Fig. 6). Group data in Fig. 6C show a striking imbalance in the proportions of units with visually modulated sound onset and sustained FR responses. Indeed, only one unit had a visually modulated onset FR response, whereas 15/212 (7.1%) of NS units and 30/589 (5.1%) of BS units had visually modulated sustained FR responses. Chi-square tests confirmed that in the middle and deep bins, sustained FR responses in both BS and NS cells were more likely to be visually modulated than onset FR responses (BS, middle: $\chi^2 = 14.33$, $p < 10^{-3}$, $\phi = 0.110$; BS, deep: $\chi^2 = 13.42$, $p < 10^{-3}$, $\phi = 0.107$; NS, middle: $\chi^2 = 6.73$, $p = 0.009$, $\phi = 0.126$; NS, deep: $\chi^2 = 6.29$, $p = 0.012$, $\phi = 0.122$).

We wondered whether the differences in the visual modulation of onset versus sustained FR responses might be explained by characteristic differences in the strength of onset and sustained FR responses to sound stimuli, as might be expected from previous studies reporting that multisensory integration depends on unimodal response strength (Stein and Stanford, 2008). As in previous studies (Lu et al., 2001; Wang et al., 2005; Malone et al., 2015), onset responses tended to have larger effect sizes and were generally more common than sustained responses (Fig. 7A and B). For BS units, effect sizes were significantly greater for onset than sustained FR responses across depth bins (Fig. 7B, left; superficial: $F = 7.28$, $p < 10^{-3}$, $\eta^2 = 0.123$; middle: $F = 21.77$, $p < 10^{-5}$, $\eta^2 = 0.079$; deep: $F = 5.57$, $p = 0.019$, $\eta^2 = 0.029$). Proportions of BS units with significant onset and sustained FR responses was comparable across depths, with the exception that onset responses were slightly less common in the superficial bin (Fig. 7B, left; $\chi^2 = 4.17$, $p = 0.041$, $\phi =$

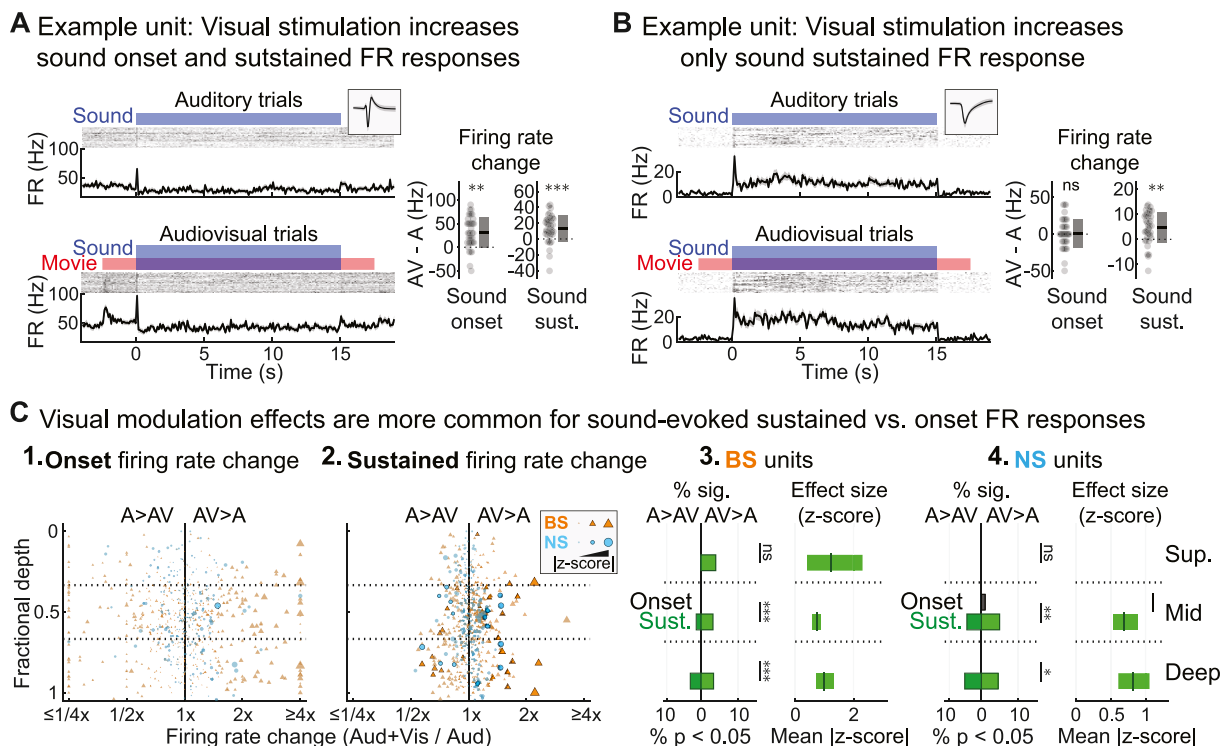


Fig. 6. Visual signals differentially modulate sound-evoked onset and sustained FR responses. (A) Example unit with visually modulated onset and sustained FR responses. Left: Spiking activity for auditory and audiovisual trials are shown in the top and bottom panels, respectively. Right: Differences in sound evoked FR responses between conditions (dots: single trials; boxes: mean \pm SD). (B) Example unit for which only the sustained FR response was modulated by visual context. (C) Population summary of visually modulated sound onset and sustained FR responses. (1–2.) FR changes between condition for each unit by fractional depth. Outlined markers indicate significant visually modulated units ($p < 0.05$, FDR corrected). (3–4.) Summary of significant visually modulated sound onset and sustained responses in BS and NS units. Left: Percentages of visually modulated units. Right: Mean effect size (plus 99% confidence interval) for significant visually modulated units. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns $p > 0.05$.

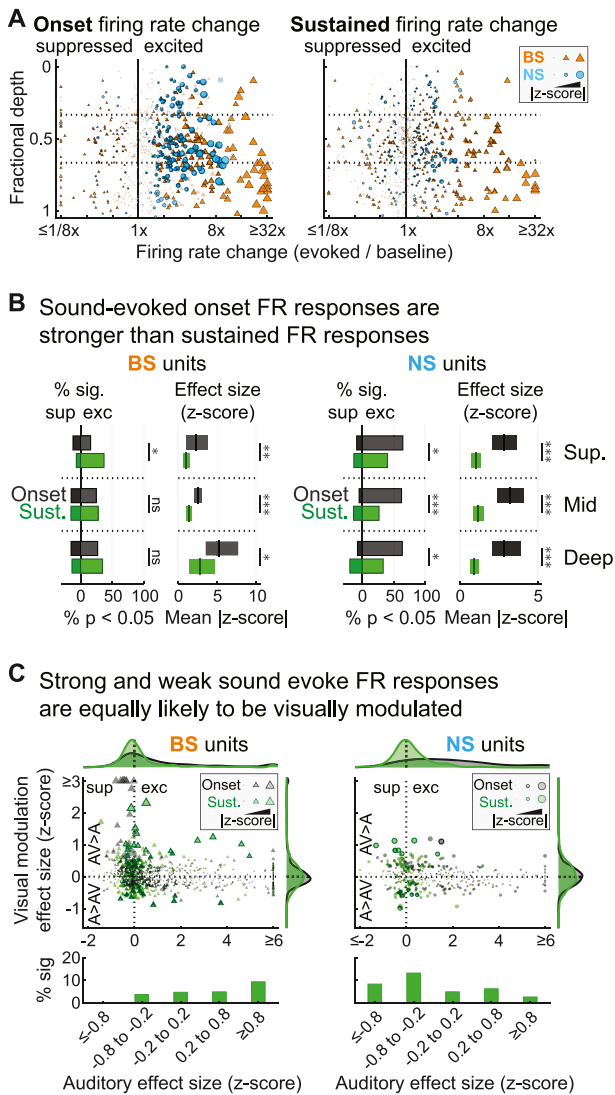


Fig. 7. Weak and strong sound evoked FR responses are equally likely to be modulated by visual context. (A) Summary of sound onset and sustained FR response effect sizes in the auditory condition (changes from spontaneous). FR changes from baseline for each unit by fractional depth. Outlined markers indicate significantly responsive units ($p < 0.05$, FDR corrected). (B) Summary of significant sound onset and sustained responses in BS and NS units. For each unit type, percentages of significantly responsive units are shown at left and mean effect sizes (plus 99% confidence interval) for significant responsive units are shown at right. (C) Summary of sound onset and sustained FR response effect sizes in the auditory condition (changes from spontaneous) and how these relate to visual modulation effects (changes between conditions). For each unit type, the scatter plot indicates visual modulation and auditory response effect size, and the bar plot below indicates percentages of visually modulated units in bins reflecting weak ($|0$ to 0.2), moderate (0.2 to 0.8), and strong auditory response effect size ($> |0.8|$). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns $p > 0.05$.

0.059). For NS units at all depth bins (Fig. 7B, right), onset FR responses were both significantly more common (superficial: $\chi^2 = 3.88$, $p = 0.049$, $\phi = 0.096$; middle: $\chi^2 = 14.53$, $p < 10^{-3}$, $\phi = 0.185$; deep: $\chi^2 = 4.64$, $p = 0.031$, $\phi = 0.105$) and had larger effect sizes (superficial: $F = 22.06$, $p < 10^{-4}$, $\eta^2 = 0.286$; middle: $F = 21.40$, $p < 10^{-4}$, $\eta^2 = 0.167$; deep: $F = 18.71$, $p < 10^{-4}$, $\eta^2 = 0.190$). Thus, sustained FR responses tended to be weaker than onset FR responses but were more likely to be visually modulated. These findings appeared consistent with the principle of inverse effectiveness, which predicts stronger multisensory influences in units with weak unimodal responses (Stein and Stanford, 2008).

To determine whether our data fit the predicted inverse relationship

between multisensory influences and unisensory response strength, we directly examined the relationship between auditory and visual modulation effect sizes (Fig. 7C). Surprisingly, significant visual modulation effects were equally common in units with strong and weak responses to sound. For statistical confirmation of this finding, we first separated units into five groups reflecting a range of sound-evoked sustained FR response strengths based on changes from spontaneous firing (z-score): ‘weak’ (-0.2 to 0.2), ‘moderate’ ($0.2-0.8$ or -0.2 to -0.8), and ‘strong’ (>0.8 or <-0.8). Chi-square tests confirmed that the proportion of units with significant visual modulation effects did not differ among these groups (Fig. 7C, lower panels; NS: $\chi^2 = 4.68$, $p = 0.321$; BS: $\chi^2 = 6.19$, $p = 0.185$). Extending these outcomes, auditory effect size (absolute z-score) was not significantly predictive of units with significant visual modulation effects using logistic regression (all p-values > 0.370). Thus, inverse effectiveness did not specifically explain why visual modulation effects were almost exclusively observed for sustained FR responses. We conclude that visual influences on auditory encoding in A1 strongly depend on sound response dynamics (onset vs. sustained FR), independent of auditory response strength.

We finally considered whether units in which visual inputs modulated sustained FR were concentrated in the deep layers, similar to the visually responsive neurons described in Fig. 3. Thus, we directly compared the depth distributions and effect sizes of visually responsive and modulated units (Fig. 8). We saw that both for BS and NS cells, the depth distribution of visually modulated units differed significantly from visually responsive units. In the deepest bin, there were fewer visually modulated cells (NS: $\chi^2 = 9.39$, $p = 0.002$, $\phi = 0.149$; BS: $\chi^2 = 5.43$, $p = 0.020$, $\phi = 0.068$), and they had smaller effect sizes (NS: $F = 5.33$, $p = 0.030$, $\eta^2 = 0.182$; BS: $F = 8.28$, $p = 0.007$, $\eta^2 = 0.179$). Neither the prevalence of visually modulated units nor the effect size was systematically stronger in deeper layers (all χ^2 statistics < 4.36 , all p-values > 0.113 ; all F-ratios < 2.17 , all p-values > 0.135). These findings suggest that visually responsive and visually modulated neuron populations have distinct laminar organizations.

3.4. Visual signals may modulate spectrotemporal receptive fields even without FR changes

Considering that many units without sound-evoked FR changes had significant STRFs (Fig. S3), we next addressed whether visual context could similarly modulate units’ STRFs without affecting their FRs, by measuring the overlap between units with visually modulated STRFs and visually modulated sustained FR responses. To determine which units had significant visually modulated STRFs, we calculated the statistical reliability of each unit’s difference STRF (Δ STRF, the auditory STRF subtracted from the audiovisual STRF based on randomly subsampled trials; Fig. S1B; Fig. 9A). In total, 16/212 NS units (7.6%) and 39/589 of BS units (6.6%) had STRFs that were significantly modulated by visual stimulation (Fig. 9B). We reported above that the prevalence and strength of visual responses (Fig. 3), but not visually modulated sustained FR responses (Fig. 6), depended significantly on cortical depth. Extending this comparison, we similarly found that the prevalence and strength of visually modulated STRFs did not significantly depend on depth for either unit type (all χ^2 statistics < 2.64 , all p-values > 0.268 ; all F-ratios > 1.19 , all p-values > 0.335).

As seen in Fig. 10, there was surprisingly little overlap between units with visually modulated STRFs and visually modulated sustained FRs (NS: 8/24, 33.3%; BS 11/58, 19.0%). We conducted permutation tests to determine whether the observed degree of overlap (units for which visual stimulation modulated both FRs and STRFs) fell within the limits expected by chance. By shuffling the labels of neurons with modulated FRs and modulated STRFs, we created null distributions reflecting the proportions of units for which modulation effects coincided by chance (10^4 iterations). For both unit types, the observed fractions of neurons with overlapping modulation effects were significantly smaller than expected by chance (NS: observed = 0.292, null median = 0.524, $p =$

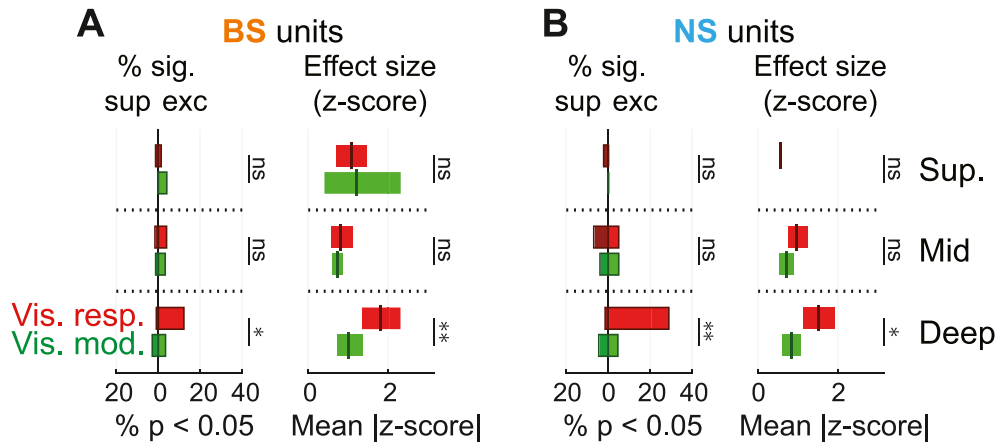


Fig. 8. Units with visually modulated responses to sound are rarer and less concentrated in the deep cortical layers than visually responsive units. (A–B) Comparison of significant visually responsive and modulated units for BS and NS units. Left: Percentages of units with significant effects. Right: Mean effect size (plus 99% confidence interval) for units with significant effects. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns $p > 0.05$.

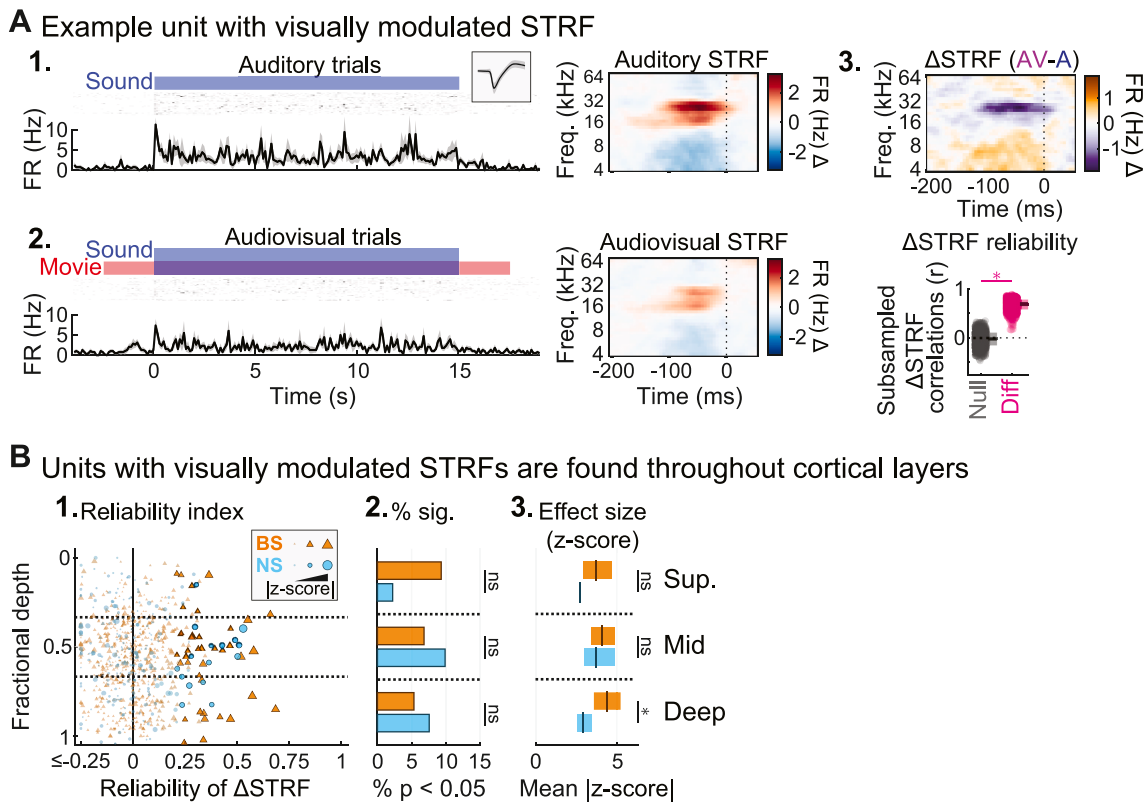


Fig. 9. Units with visually modulated STRFs found throughout cortical layers. (A) Example unit with a visually modulated STRF. (1–2.) Spiking activity and STRFs for auditory and audiovisual trials. (3.) Upper: Δ STRF reflecting the difference between audiovisual and auditory STRFs. Lower: Subsampled Δ STRF correlations (dots: single subsample iterations; boxes: mean \pm SD). (B) Population summary of visually modulated STRFs. (1.) Δ STRF reliability for each unit by fractional depth. Outlined markers indicate units with significant modulation ($p < 0.05$, FDR corrected). (2.) Percentages of units with significant Δ STRF reliability. (3.) Mean effect size (plus 99% confidence interval) for units with significant Δ STRF reliability. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns $p > 0.05$.

0.012; BS: observed = 0.190, null median = 0.592, $p < 10^{-4}$). This outcome underscores the notion that units with modulated FRs and modulated STRFs occur in largely distinct neuronal populations, perhaps reflecting distinct mechanisms. In total, at least one aspect of sound encoding was modulated by visual context for roughly one in ten neurons (NS 24/212, 11.3%; BS 58/589, 9.9%).

The differences in cortical depth distribution between units with significant visual responses and units with visually modulated responses to sound suggest these neurons may correspond to distinct unit

populations. In direct support of this possibility, we found that most units with visual onset responses did not show any type of significant visually modulated sound response (neither FR nor STRF), and vice versa (Fig. 11). Overall, after FDR correction, we saw that visual stimulation affects firing in at least one way – by driving a direct visual response, by modulating sound-evoked FRs, or by modulating STRFs – in nearly one in five A1 neurons (NS 52/212, 24.5%; BS 95/589, 16.1%; combined 147/801, 18.4%).

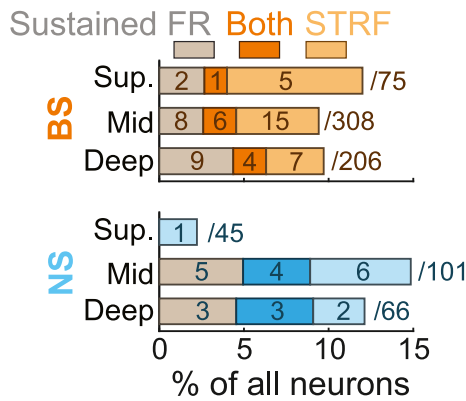


Fig. 10. Visual signals may modulate spectrotemporal receptive fields even without FR changes. Bar plots represent overlap between units with visually modulated sustained FR responses and STRFs for each unit type.

3.5. Visual context modulates STRF gain but preserves spectrotemporal tuning

Because many units had reliable Δ STRFs (i.e., had STRFs that were significantly changed by simultaneous visual stimulation), we next asked whether these differences corresponded to changes in what features the neurons were selectively responsive to (i.e., changes in STRF tuning), or to changes in how strongly the neurons responded to the

same stimulus features (i.e., changes in gain). To test for spectrotemporal tuning changes, we calculated the correlation between the sound-alone and audiovisual STRFs. For this analysis, we included only units for which spectrotemporal tuning in the Auditory condition could be reliably measured ($r > 0.5$) to ensure sound-unresponsive units did not skew the results. As shown in Fig. 12, sound-alone and audiovisual STRFs were highly correlated (NS median $r = 0.95$; BS median $r = 0.89$), implying that spectrotemporal tuning was largely preserved. Absolute correlations between auditory STRFs and Δ STRFs were similarly high (NS median: $r = 0.72$; BS median: $r = 0.62$), implying that the greatest changes in audiovisual STRFs occurred at the time-frequency bins eliciting the strongest changes in responses – a feature characteristic of gain changes but not tuning changes. Together, these results implied that visual stimulation did not substantially alter neurons’ tuning but instead may have altered neurons’ gain.

To address whether visual stimulation affected STRF gain, we measured the slope of the relationship between the sound-only and audiovisual STRFs using standardized major axis regression (a variant of linear regression that treats measurement error in both variables symmetrically; Warton et al., 2006). For this analysis, we included only units for which sound-only and audiovisual STRFs were highly correlated to ensure accurate slope measurements ($r^2 > 0.5$; range highlighted in Fig. 13). Group data in Figs. 13C and F shows that visual stimulation indeed resulted in STRF gain changes, which included both increases and decreases relative to the auditory condition. In extreme cases, gain nearly doubled or halved. Sound-only and audiovisual STRFs were highly correlated even for units with the largest gain changes. However,

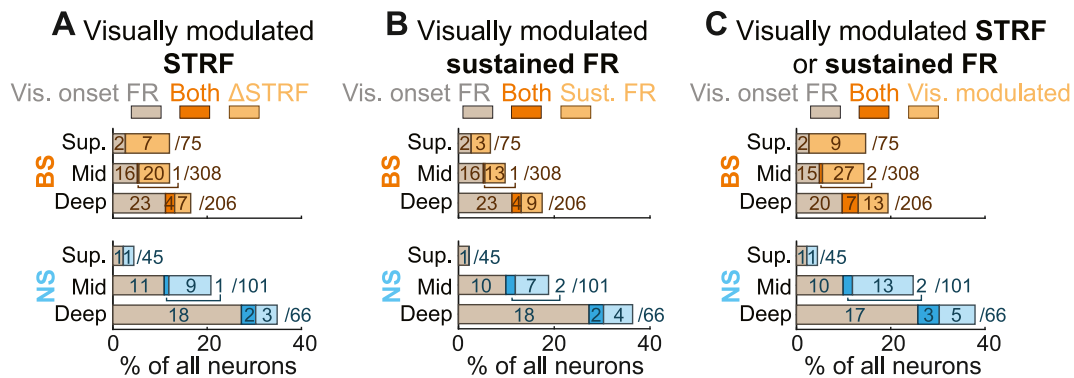


Fig. 11. Visual stimuli rarely drive firing and modulate sound responses in the same units. Bar plots show overlap between units with (A) visual responses and visually modulated STRFs, (B) visual responses and visually modulated sustained FR response, and (C) visual responses and visually modulated STRFs or sustained FR responses.

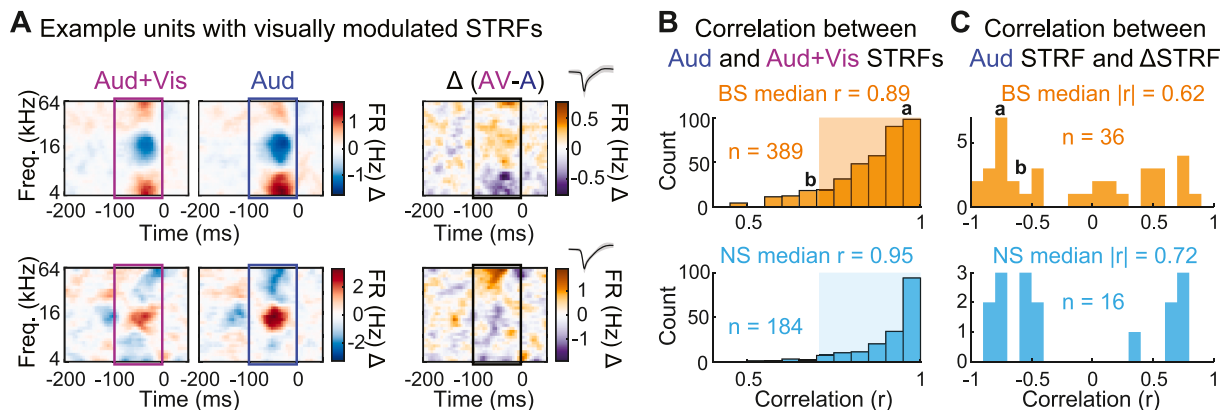


Fig. 12. Auditory spectrotemporal tuning is robust to visual context. (A) Top: Example unit with highly correlated STRFs between conditions. Bottom: Example unit with moderately correlated STRFs between conditions. (B) Distributions of STRF correlations between conditions for BS (top) and NS units (bottom), indicating similar spectrotemporal tuning. Shaded region indicates units analyzed in (B). (C) Distributions of correlations between Auditory and Δ STRFs for BS (top) and NS units (bottom), similarly suggesting preserved STRF structure.

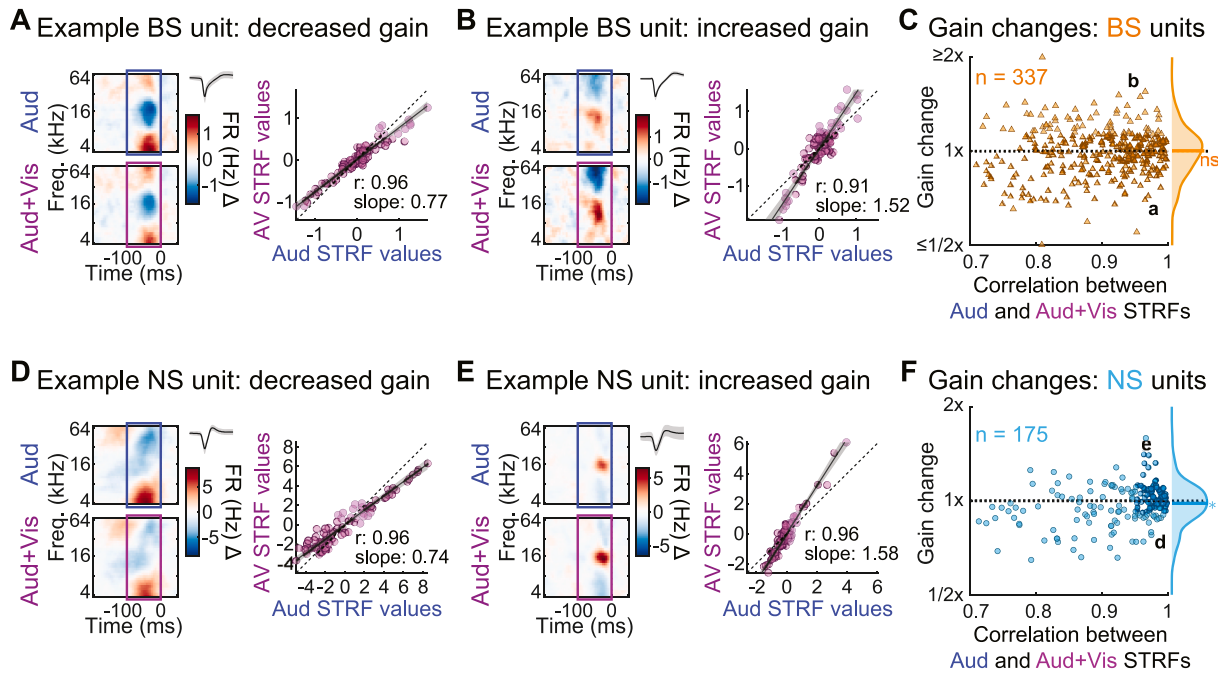


Fig. 13. Visual inputs may modify STRF response gain despite preserved spectrotemporal tuning. (A) Example BS unit with decreased gain in the Audiovisual condition. Left: STRFs for each condition with unit waveform. Right: best fit lines to STRF time-frequency bins from each condition (shading: 95% confidence intervals). (B) Example BS unit with gain increase in the Audiovisual condition. (C) STRF gain changes and correlations between conditions for BS units. (D–E) Example NS units with decreased and increased gain and in the Audiovisual condition. (F) STRF gain changes and correlations between conditions for NS units. * $p < 0.05$, ns $p > 0.05$.

there was little systematic trend for visual stimulation to increase or decrease gain. For BS units, the median gain change was not significantly different from one (median 1.003; Wilcoxon signed-rank test: $p = 0.888$), implying increases and decreases were approximately balanced (Fig. 6Bc). There was a slight trend for NS units' gain to be reduced by visual stimulation, but this trend was both small and only barely significant (median 0.982; Wilcoxon signed-rank test: $p = 0.036$). Thus, we conclude that the effects of visual stimulation on neural responses can be quite substantial in individual neurons and are heterogeneous across neural populations.

4. Discussion

Perception is inherently multisensory under natural conditions (Sugihara et al., 2006; Allman and Meredith, 2007; Stein and Stanford, 2008; Bigelow and Poremba, 2016). Concomitantly, recent evidence indicates multisensory interaction is pervasive throughout the brain, including within primary sensory areas (Ghazanfar and Schroeder, 2006; Bizley et al., 2007, 2016; Bizley and King, 2008; Banks et al., 2011; Iurilli et al., 2012; McClure and Polack, 2019). The extent to which crossmodal inputs may influence or even dominate spiking activity of local neurons remains an open question. In the current study, roughly one in ten A1 neurons responded to unimodal visual stimuli, consistent with previous estimates which typically range from ~5 to 15% (Wallace et al., 2004; Bizley and King, 2008; Kobayasi et al., 2013; Morrill and Hasenstaub, 2018). Additional analyses revealed that most of these neurons (>90%) were also responsive to sound. STRF analysis proved essential to this conclusion, as sound responses were only detected in slightly over half of neurons using average FRs (Fig. 5). These observations extend earlier findings which show that a minority of A1 neurons respond to visual inputs, and clarify that such responses occur in addition to, rather than instead of, responses to sound.

In an additional subpopulation of neurons, we observed visually modulated responses to sound (Figs. 6–13), which were almost completely limited to sustained FR responses and STRFs. We examined

inverse effectiveness as a potential explanation for the lack of visually modulated onset responses, which predicts that these strong baseline responses may be only weakly modified by crossmodal input. However, additional analysis indicated that visual modulation effects for sustained FR responses were not significantly related to unimodal auditory response strength. These findings add to a small but growing body of studies suggesting inverse effectiveness may not universally govern multisensory integrative responses (Holmes, 2009). Additional work is needed to determine whether differences between onset and sustained FR responses such as stimulus selectivity (Wang et al., 2005) or response variability (Churchland et al., 2010) may give rise to their distinct sensitivity to visual context.

Visually responsive and modulated subpopulations were mostly nonoverlapping (Fig. 11), consistent with prior work suggesting that visual responses are neither necessary nor sufficient for visually modulated responses to sound (Bizley et al., 2007). A major question motivating the present study was whether such visual modulatory influences might differently affect overall excitability (average FR changes) and sensitivity to spectrotemporal features (STRFs). Indeed, we observed only limited overlap between units with visually modulated sustained FR responses and STRFs, with over one third of all visually modulated units showing STRF changes alone. Because STRF and sustained FR analyses reflected identical stimulus and spike distributions, differences in visual context modulation primarily reflect the sensitivity of STRFs to spike timing above and beyond spike rate. Our observations thus reinforce previous studies concluding that spike rate and timing changes in A1 carry non-redundant stimulus information (Brugge and Merzenich, 1973; deCharms and Merzenich, 1996; Recanzone, 2000; Lu et al., 2001; Wang et al., 2005; Malone et al., 2010, 2015; Slee and David, 2015; Insanally et al., 2019; Liu et al., 2019).

Visually responsive neurons were most prevalent in the deep cortical layers, consistent with our earlier study (Morrill and Hasenstaub, 2018). By contrast, we found no evidence that the proportion of visually modulated neurons or the strength of their responses significantly depended on cortical depth. This was surprising, considering both the

depth dependence of visual responses in our own unit sample and previous studies reporting stronger audiovisual responses in ferrets (Atilgan et al., 2018) and humans (Gau et al., 2020). Visual modulation was also observed in the supragranular layers of the ferret study (Atilgan et al., 2018). Although supragranular visual projections are present in mouse (Banks et al., 2011), supragranular visual responses and modulation effects were uncommon in our physiological recordings (Morrill and Hasenstaub, 2018). Additional studies are needed to determine whether these discrepancies reflect differences in species (ferret vs. mouse), stimulus paradigm (correlated vs. uncorrelated), or other variables.

An additional extension of our earlier finding that some A1 neurons respond to visual stimuli is that, for a minority of these neurons, the responses were suppressed relative to spontaneous firing. Interestingly, most neurons with suppressed responses were found in the middle cortical depth bin, superficial to the deep-layer majority. Previous studies have reported that L6 corticothalamic pyramidal cells (L6CTs) selectively activate inhibitory interneurons with cell bodies in L6 and axonal arborizations extending throughout layers (Bortone et al., 2014). In response to activation of L6CTs, these transaminar axonal projections mediate suppressed responses above L6. Thus, one potential explanation for suppressed responses in our sample is activation of L6CTs by visual stimulation, which in turn suppress responses in more superficial neurons. Functional connectivity studies using L6-specific mouse lines are needed to investigate this possible circuit.

Numerous studies have found that sensory encoding may be either enhanced or diminished by input from another modality (Perrault et al., 2003; Romanski, 2007; Stein and Stanford, 2008; Sugihara et al., 2006; Kobayashi et al., 2013). Similar outcomes have been obtained with both correlated and uncorrelated bimodal signals (Dahl et al., 2010), with several studies suggesting enhanced encoding is more common with correlated stimuli (Diehl and Romanski, 2014; Meijer et al., 2017). In correlated paradigms, such as natural audiovisual speech, encoding benefits of bimodal stimulation likely reflect synergistic interaction of temporally and/or spatially coincident feature selectivity within each respective modality (Stein and Stanford, 2008). In uncorrelated bimodal stimulus paradigms, including the present study, improved encoding with the addition of an uncorrelated signal from another modality may reflect stochastic resonance, in which sensitivity to a weakly detectable stimulus may be increased by random noise within or between modalities (Wiesenfeld and Moss, 1995; Shu et al., 2003; Hasenstaub et al., 2005; Crosse et al., 2016; Malone et al., 2017). Consistent with these observations, STRF gain was increased by uncorrelated visual stimulation for approximately half of the neurons in our sample (Fig. 13), suggesting weak or inconsistent responses to sound features in these neurons may have been strengthened or regularized by crossmodal input. These and similar findings by previous studies establish neural phenomena parallel to psychophysical experiments reporting significantly improved detection of weakly perceptible events following the addition of both correlated and uncorrelated stimulation within a second modality (Ward et al., 2010; Huang et al., 2017; Gleiss and Kayser, 2014; Krauss et al., 2018).

5. Data availability statement

Data for this study are available upon request to the corresponding author.

CRedit authorship contribution statement

James Bigelow: Conceptualization, Methodology, Investigation, Funding acquisition, Formal analysis, Visualization, Writing – original draft, Data curation. **Ryan J. Morrill:** Software, Investigation, Writing – review & editing. **Timothy Olsen:** Investigation, Writing – review & editing. **Andrea R. Hasenstaub:** Conceptualization, Project administration, Supervision, Funding acquisition, Formal analysis, Visualization, Resources, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Stephanie Bazarini for assistance with data curation. This work was supported by National Institutes of Health (F32DC016846 to J.B. and R01DC014101, R01NS116598, R01MH122478, and R01EY025174 to A.R.H.), the National Science Foundation (R.J.M.), the Klingenstein Foundation (A.R.H.), Hearing Research Inc. (A.R.H.), PBBR Breakthrough Fund (A.R.H.), and the Coleman Memorial Fund (A.R.H.).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.crneur.2022.100040>.

References

- Allman, B.L., Meredith, M.A., 2007. Multisensory processing in “unimodal” neurons: cross-modal subthreshold auditory effects in cat extrastriate visual cortex. *J. Neurophysiol.* (Bethesda) 98, 545–549.
- Atencio, C.A., Schreiner, C.E., 2013. Stimulus choices for spike-triggered receptive field analysis. In: Depireux, D.A., Elhilali, M. (Eds.), *Handbook of Modern Techniques in Auditory Cortex*. Nova Biomedical, New York, pp. 61–100.
- Atilgan, H., Town, S.M., Wood, K.C., Jones, G.P., Maddox, R.K., Lee, A.K., Bizley, J.K., 2018. Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding. *Neuron* 97, 640–655.
- Banks, M.I., Uhlrich, D.J., Smith, P.H., Krause, B.M., Manning, K.A., 2011. Descending projections from extrastriate visual cortex modulate responses of cells in primary auditory cortex. *Cerebr. Cortex* 21, 2620–2638.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B* 57, 289–300.
- Bigelow, J., Malone, B.J., 2017. Cluster-based analysis improves predictive validity of spike-triggered receptive field estimates. *PLoS One* 12, e0183914.
- Bigelow, J., Morrill, R.J., Dekloe, J., Hasenstaub, A.R., 2019. Movement and VIP interneuron activation differentially modulate encoding in mouse auditory cortex. *eNeuro* 6 (5).
- Bigelow, J., Poremba, A., 2016. Audiovisual integration facilitates monkeys' short-term memory. *Anim. Cognit.* 19, 799–811.
- Bizley, J.K., Jones, G.P., Town, S.M., 2016. Where are multisensory signals combined for perceptual decision-making? *Curr. Opin. Neurobiol.* 40, 31–37.
- Bizley, J.K., King, A.J., 2008. Visual–auditory spatial processing in auditory cortical neurons. *Brain Res.* 1242, 24–36.
- Bizley, J.K., Nodal, F.R., Bajo, V.M., Nelken, I., King, A.J., 2007. Physiological and anatomical evidence for multisensory interactions in auditory cortex. *Cerebr. Cortex* 17, 2172–2189.
- Bortone, D.S., Olsen, S.R., Scanziani, M., 2014. Translaminar inhibitory cells recruited by layer 6 corticothalamic neurons suppress visual cortex. *Neuron* 82, 474–485.
- Brugge, J.F., Merzenich, M.M., 1973. Responses of neurons in auditory cortex of the macaque monkey to monaural and binaural stimulation. *J. Neurophysiol.* (Bethesda) 36, 1138–1158.
- Budinger, E., Scheich, H., 2009. Anatomical connections suitable for the direct processing of neuronal information of different modalities via the rodent primary auditory cortex. *Hear. Res.* 258, 16–27.
- Chang, M., Kawai, H.D., 2018. A characterization of laminar architecture in mouse primary auditory cortex. *Brain Struct. Funct.* 223, 4187–4209.
- Churchland, M.M., Byron, M.Y., Cunningham, J.P., Sugrue, L.P., Cohen, M.R., Corrado, G.S., Newsome, W.T., Clark, A.M., Hosseini, P., Scott, B.B., Bradley, D.C., 2010. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.* 13, 369–378.
- Crosse, M.J., Di Liberto, G.M., Lalor, E.C., 2016. Eye can hear clearly now: inverse effectiveness in natural audiovisual speech processing relies on long-term crossmodal temporal integration. *J. Neurosci.* 36, 9888–9895.
- Dahl, C., Logothetis, N.K., Kayser, C., 2010. Modulation of visual responses in the superior temporal sulcus by audio-visual congruency. *Front. Integr. Neurosci.* 4, 10.
- deCharms, R.C., Blake, D.T., Merzenich, M.M., 1998. Optimizing sound features for cortical neurons. *Science* 280, 1439–1444.
- deCharms, R.C., Merzenich, M.M., 1996. Primary cortical representation of sounds by the coordination of action-potential timing. *Nature* 381, 610–613.
- Diehl, M.M., Romanski, L.M., 2014. Responses of prefrontal multisensory neurons to mismatching faces and vocalizations. *J. Neurosci.* 34, 11233–11243.
- Dombeck, D.A., Khabbaz, A.N., Collman, F., Adelman, T.L., Tank, D.W., 2007. Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron* 56, 43–57.

- Escabi, M.A., Read, H.L., Viventi, J., Kim, D.H., Higgins, N.C., Storace, D.A., Liu, A.S., Gifford, A.M., Burke, J.F., Campisi, M., Kim, Y.S., 2014. A high-density, high-channel count, multiplexed μ ECoG array for auditory-cortex recordings. *J. Neurophysiol.* (Bethesda) 112, 1566–1583.
- Fritz, J., Shamma, S., Elhilali, M., Klein, D., 2003. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* 6, 1216–1223.
- Gau, R., Bazin, P.L., Trampel, R., Turner, R., Noppeney, U., 2020. Resolving multisensory and attentional influences across cortical depth in sensory cortices. *Elife* 9, e46856.
- Ghazanfar, A.A., Schroeder, C.E., 2006. Is neocortex essentially multisensory? *Trends Cognit. Sci.* 10, 278–285.
- Gleiss, S., Kayser, C., 2014. Acoustic noise improves visual perception and modulates occipital oscillatory states. *J. Cognit. Neurosci.* 26, 699–711.
- Gourévitch, B., Occelli, F., Gaucher, Q., Aushana, Y., Edeline, J.M., 2015. A new and fast characterization of multiple encoding properties of auditory neurons. *Brain Topogr.* 28, 379–400.
- Hasenstaub, A., Shu, Y., Haider, B., Kraushaar, U., Duque, A., McCormick, D.A., 2005. Inhibitory postsynaptic potentials carry synchronized frequency information in active cortical networks. *Neuron* 47, 423–435.
- Holmes, N.P., 2009. The principle of inverse effectiveness in multisensory integration: some statistical considerations. *Brain Topogr.* 21, 168–176.
- Hromádka, T., DeWeese, M.R., Zador, A.M., 2008. Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol.* 6, e16.
- Huang, J., Sheffield, B., Lin, P., Zeng, F.G., 2017. Electro-tactile stimulation enhances cochlear implant speech recognition in noise. *Sci. Rep.* 7, 1–5.
- Insanally, M.N., Carcea, I., Field, R.E., Rodgers, C.C., DePasquale, B., Rajan, K., DeWeese, M.R., Albanna, B.F., Froemke, R.C., 2019. Spike-timing-dependent ensemble encoding by non-classically responsive cortical neurons. *Elife* 8, e42409.
- Iurilli, G., Ghezzi, D., Olcese, U., Lassi, G., Nazzaro, C., Tonini, R., Tucci, V., Benfenati, F., Medini, P., 2012. Sound-driven synaptic inhibition in primary visual cortex. *Neuron* 73, 814–828.
- Joachimsthaler, B., Uhlmann, M., Miller, F., Ehret, G., Kurt, S., 2014. Quantitative analysis of neuronal response properties in primary and higher-order auditory cortical fields of awake house mice (*Mus musculus*). *Eur. J. Neurosci.* 39, 904–918.
- Kajikawa, Y., Falchier, A., Musacchia, G., Lakatos, P., Schroeder, C.E., 2012. Audiovisual integration in nonhuman primates: a window into the anatomy and physiology of cognition. In: Murray, M.M., Wallace, M.T. (Eds.), *The Neural Bases of Multisensory Processes*. CRC Press/Taylor & Francis, Boca Raton, FL.
- Kawaguchi, Y., Kubota, Y., 1997. GABAergic cell subtypes and their synaptic connections in rat frontal cortex. *Cerebr. Cortex* 7, 476–486.
- Kayser, C., Logothetis, N.K., Panzeri, S., 2010. Visual enhancement of the information representation in auditory cortex. *Curr. Biol.* 20, 19–24.
- Kayser, C., Petkov, C.I., Logothetis, N.K., 2009. Multisensory interactions in primate auditory cortex: fMRI and electrophysiology. *Hear. Res.* 258, 80–88.
- King, A.J., Hammond-Kenny, A., Nodal, F.R., 2019. Multisensory processing in the auditory cortex. In: Lee, A., Wallace, M., Coffin, A., Popper, A., Fay, R. (Eds.), *Multisensory Processes*. Springer Handbook of Auditory Research. Springer, Cham, pp. 105–133.
- Kleiner, M., Brainard, D., Pelli, D., 2007. What's new in Psychtoolbox-3? *ECVP Abstr Suppl.* 14.
- Kobayasi, K.I., Suwa, Y., Riquimaroux, H., 2013. Audiovisual integration in the primary auditory cortex of an awake rodent. *Neurosci. Lett.* 534, 24–29.
- Krauss, P., Tziridis, K., Schilling, A., Schulze, H., 2018. Cross-modal stochastic resonance as a universal principle to enhance sensory processing. *Front. Neurosci.* 12, 578.
- Liang, Z., Shen, W., Sun, C., Shou, T., 2008. Comparative study on the offset responses of simple cells and complex cells in the primary visual cortex of the cat. *Neuroscience* 156, 365–373.
- Li, L.Y., Xiong, X.R., Ibrahim, L.A., Yuan, W., Tao, H.W., Zhang, L.I., 2015. Differential receptive field properties of parvalbumin and somatostatin inhibitory neurons in mouse auditory cortex. *Cerebr. Cortex* 25, 1782–1791.
- Liu, J., Whiteway, M.R., Sheikhattar, A., Butts, D.A., Babadi, B., Kanold, P.O., 2019. Parallel processing of sound dynamics across mouse auditory cortex via spatially patterned thalamic inputs and distinct areal intracortical circuits. *Cell Rep.* 27, 872–885.
- Liu, L., Xu, H., Wang, J., Li, J., Tian, Y., Zheng, J., He, M., Xu, T.L., Wu, Z.Y., Li, X.M., Duan, S.M., 2020. Cell type-differential modulation of prefrontal cortical GABAergic interneurons on low gamma rhythm and social interaction. *Sci. Adv.* 6, eaay4073.
- Lu, T., Liang, L., Wang, X., 2001. Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat. Neurosci.* 4, 1131–1138.
- Malone, B.J., Heiser, M.A., Beitel, R.E., Schreiner, C.E., 2017. Background noise exerts diverse effects on the cortical encoding of foreground sounds. *J. Neurophysiol.* (Bethesda) 118, 1034–1054.
- Malone, B.J., Scott, B.H., Semple, M.N., 2010. Temporal codes for amplitude contrast in auditory cortex. *J. Neurosci.* 30, 767–784.
- Malone, B.J., Scott, B.H., Semple, M.N., 2015. Diverse cortical codes for scene segmentation in primate auditory cortex. *J. Neurophysiol.* (Bethesda) 113, 2934–2952.
- McClure Jr., J.P., Polack, P.O., 2019. Pure tones modulate the representation of orientation and direction in the primary visual cortex. *J. Neurophysiol.* (Bethesda) 121, 2202–2214.
- Meijer, G.T., Montijn, J.S., Pennartz, C.M., Lansink, C.S., 2017. Audiovisual modulation in mouse primary visual cortex depends on cross-modal stimulus configuration and congruency. *J. Neurosci.* 37, 8783–8796.
- Moore, A.K., Wehr, M., 2013. Parvalbumin-expressing inhibitory interneurons in auditory cortex are well-tuned for frequency. *J. Neurosci.* 33, 13713–13723.
- Morrill, R.J., Hasenstaub, A.R., 2018. Visual information present in infragranular layers of mouse auditory cortex. *J. Neurosci.* 38, 2854–2862.
- Niell, C.M., Stryker, M.P., 2008. Highly selective receptive fields in mouse visual cortex. *J. Neurosci.* 28, 7520–7536.
- Niell, C.M., Stryker, M.P., 2010. Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* 65, 472–479.
- Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M., Kennedy, D.H., 2016. Kilosort: Realtime Spike-Sorting for Extracellular Electrophysiology with Hundreds of Channels. *bioRxiv*, 061481.
- Paxinos, G., Franklin, K.B.J., 2019. *Paxinos and Franklin's the Mouse Brain in Stereotaxic Coordinates*. Academic press, Cambridge, MA.
- Perrault Jr., T.J., Vaughan, J.W., Stein, B.E., Wallace, M.T., 2003. Neuron-specific response characteristics predict the magnitude of multisensory integration. *J. Neurophysiol.* (Bethesda) 90, 4022–4026.
- Phillips, E.A., Hasenstaub, A.R., 2016. Asymmetric effects of activating and inactivating cortical interneurons. *Elife* 5, e18383.
- Phillips, E.A., Schreiner, C.E., Hasenstaub, A.R., 2017a. Cortical interneurons differentially regulate the effects of acoustic context. *Cell Rep.* 20 (4), 771–778.
- Phillips, E.A., Schreiner, C.E., Hasenstaub, A.R., 2017b. Diverse effects of stimulus history in waking mouse auditory cortex. *J. Neurophysiol.* (Bethesda) 118, 1376–1393.
- Piscopo, D.M., El-Danaf, R.N., Huberman, A.D., Niell, C.M., 2013. Diverse visual features encoded in mouse lateral geniculate nucleus. *J. Neurosci.* 33, 4642–4656.
- Recanzone, G.H., 2000. Response profiles of auditory cortical neurons to tones and noise in behaving macaque monkeys. *Hear. Res.* 150, 104–118.
- Romanski, L.M., 2007. Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. *Cerebr. Cortex* 17, i61–69.
- Rutkowski, R.G., Shackleton, T.M., Schnupp, J.W., Wallace, M.N., Palmer, A.R., 2002. Spectrotemporal receptive field properties of single units in the primary, dorsocaudal and ventrorostral auditory cortex of the Guinea pig. *Audiol. Neuro. Otol.* 7, 214–227.
- Schiff, M.L., Reyes, A.D., 2012. Characterization of thalamocortical responses of regular-spiking and fast-spiking neurons of the mouse auditory cortex in vitro and in silico. *J. Neurophysiol.* (Bethesda) 107, 1476–1488.
- Schneider, D.M., Mooney, R., 2018. How movement modulates hearing. *Annu. Rev. Neurosci.* 41, 553–572.
- See, J.Z., Atencio, C.A., Sohal, V.S., Schreiner, C.E., 2018. Coordinated neuronal ensembles in primary auditory cortical columns. *Elife* 7, e35587.
- Shu, Y., Hasenstaub, A., Badoual, M., Bal, T., McCormick, D.A., 2003. Barrages of synaptic activity control the gain and sensitivity of cortical neurons. *J. Neurosci.* 23, 10388–10401.
- Slee, S.J., David, S.V., 2015. Rapid task-related plasticity of spectrotemporal receptive fields in the auditory midbrain. *J. Neurosci.* 35, 13090–13102.
- Smiley, J.F., Falchier, A., 2009. Multisensory connections of monkey auditory cerebral cortex. *Hear. Res.* 258, 37–46.
- Stehberg, J., Dang, P.T., Frostig, R.D., 2014. Unimodal primary sensory cortices are directly connected by long-range horizontal projections in the rat sensory cortex. *Front. Neuroanat.* 8, 93.
- Stein, B.E., Stanford, T.R., 2008. Multisensory integration: current issues from the perspective of the single neuron. *Nat. Rev. Neurosci.* 9, 255–266.
- Sugihara, T., Diltz, M.D., Averbeck, B.B., Romanski, L.M., 2006. Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *J. Neurosci.* 26, 11138–11147.
- Thorson, L.L., Liénard, J., David, S.V., 2015. The essential complexity of auditory receptive fields. *PLoS Comput. Biol.* 11, e1004628.
- Tremblay, R., Lee, S., Rudy, B., 2016. GABAergic interneurons in the neocortex: from cellular properties to circuits. *Neuron* 91, 260–292.
- Wallace, M.T., Ramachandran, R., Stein, B.E., 2004. A revised view of sensory cortical parcellation. *Proc. Natl. Acad. Sci. U.S.A.* 101, 2167–2172.
- Wang, X., Lu, T., Snider, R.K., Liang, L., 2005. Sustained firing in auditory cortex evoked by preferred stimuli. *Nature* 435, 341–346.
- Ward, L.M., MacLean, S.E., Kirschner, A., 2010. Stochastic resonance modulates neural synchronization within and between cortical sources. *PLoS One* 5, e14371.
- Warton, D.I., Wright, L.J., Falster, D.S., Westoby, M., 2006. Bivariate line-fitting methods for allometry. *Biol. Rev.* 81, 259–291.
- Wiesenfeld, K., Moss, F., 1995. Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDS. *Nature* 373, 33–36.
- Wu, G.K., Arbuckle, R., Liu, B.H., Tao, H.W., Zhang, L.I., 2008. Lateral sharpening of cortical frequency tuning by approximately balanced inhibition. *Neuron* 58, 132–143.
- Wu, M.C., David, S.V., Gallant, J.L., 2006. Complete functional characterization of sensory neurons by system identification. *Annu. Rev. Neurosci.* 29, 477–505.
- Yu, J., Hu, H., Agmon, A., Svoboda, K., 2019. Recruitment of GABAergic interneurons in the barrel cortex during active tactile behavior. *Neuron* 104, 412–427.