Visually-evoked activity and variable modulation of auditory responses in the macaque inferior colliculus

Abbreviated title:

How visual cues affect auditory activity in the IC

Meredith N. Schmehl[1-3], Jesse L. Herche[1-3], and Jennifer M. Groh[1-5]

1. Department of Neurobiology, Duke University, Durham, NC 27708
2. Center for Cognitive Neuroscience, Duke University, Durham, NC 27708
3. Duke Institute for Brain Sciences, Duke University, Durham, NC 27708
4. Department of Psychology & Neuroscience, Duke University, Durham, NC 27708
5. Department of Biomedical Engineering, Duke University, Durham, NC 27708

Corresponding Author: Meredith N. Schmehl (meredith.schmehl@duke.edu)

The authors declare no competing financial interests.

Abstract

How multisensory cues affect processing in early sensory brain areas is not well understood. The inferior colliculus (IC) is an early auditory structure that is visually responsive (Porter et al., 2007; Bulkin & Groh, 2012b, 2012a), but little is known about how visual signals affect the IC's auditory representation. We explored how visual cues affect both spiking and local field potential (LFP) activity in the IC of two monkeys performing a task involving saccades to auditory, visual, or combined audiovisual stimuli. We confirm that LFPs are sensitive to the onset of fixation lights as well as the onset of visual targets presented during steady fixation. The LFP waveforms evoked by combined audiovisual stimuli differed from those evoked by sounds alone. In single-unit spiking activity, responses were weak when visual stimuli were presented alone, but visual stimuli could modulate the activity evoked by sounds in a stronger way. Such modulations could involve either increases or decreases in activity, and whether increases or decreases were observed was variable and not obviously correlated with the responses evoked by visual or auditory stimuli alone. These findings indicate that visual stimuli shape the IC's auditory representation in flexible ways that differ from those observed previously in multisensory areas.

Significance Statement

We find that the inferior colliculus, a primarily auditory brain area, displays distinct population-level responses to visual stimuli. We also find that visual cues can influence the auditory responses of individual neurons. Together, the results provide insight into how relatively early sensory areas may play a role in combining multiple sensory modalities to refine the perception of complex environments.

Introduction

Visual and auditory processing streams were once thought to remain initially separate in the brain. Studies of how visual and auditory stimuli *interact* focused chiefly on late stages of processing: association areas such as intraparietal cortex or motor output stages such as the superior colliculus (SC) (e.g., (Witten & Knudsen, 2005; Chandrasekaran, 2017; Schmehl & Groh, 2021)). At such late processing stages, two key features characterize response patterns: a) both modalities can lead to strong responses, i.e., spiking activity increases in response to either visual or auditory stimuli; and b) combinations of visual and auditory stimuli lead to greater increases in spiking activity than do either of the two stimuli when presented alone. Indeed, in the SC, combined visual and auditory stimuli are thought to lead to nearly linear summation of the responses evoked by each stimulus alone (Meredith & Stein, 1986; Wallace et al., 1998; Stanford et al., 2005; Alvarado et al., 2007).

However, interactions between these sensory systems are now known to begin earlier than was previously appreciated, including at pre-cortical stages (Bulkin & Groh, 2006; Ghazanfar & Schroeder, 2006; Schmehl & Groh, 2021). What *kinds* of interactions between visual and auditory signals occur in “early” areas is unknown, and may be different from those observed at later stages. Early sensory areas are typically generalists, playing a role in many aspects of sensory perception, whereas later areas such as the SC are thought to be focused on specific roles such as generating orienting movements of the eyes to the location of a sensory stimulus regardless of its modality. And likely only some of those perceptual functions are under the influence of other sensory modalities. It is also not clear that sensory signals should necessarily add linearly; rather, a variety of different types of modulation may occur.

To gain insight into the properties of multimodal interactions in an early sensory brain region, we turned to a predominantly auditory area, the inferior colliculus (IC). Building on

previous work establishing the presence of visual inputs to the IC (retina: Itaya & Van Hoesen, 1982; Yamauchi & Yamadori, 1982; Zhang, 1984; Paloff et al., 1985; superior colliculus: Harting, 1977; visual cortex: Cooper & Young, 1976) as well as visual responses in both spiking and local field potential (LFP) activity (Porter et al., 2007; Bulkin & Groh, 2012b), we evaluated both single-unit and LFP signals in the IC during an audiovisual localization task. Visually-evoked activity was robust in the LFP, but much weaker in spiking activity. Nevertheless, visual stimuli could alter auditory-evoked responses in both measures of activity. Intriguingly, the addition of a visual stimulus to an auditory stimulus could either increase or decrease the response, and the direction of the effect was not predictable from the unimodal responses. Together, these results indicate that the cross-modal interactions in this early auditory brain area are considerably different from those reported in the adjacent SC, suggesting that the computations supporting multisensory interactions may vary widely in the brain.

Materials and Methods

Subjects

Subjects were two rhesus macaques (*Macaca mulatta*), one female age 13 years (Monkey J) and one male age 4 years (Monkey D) at the start of training. All procedures were conducted at Duke University and approved by the Duke University Institutional Animal Care and Use Committee.

Surgical Preparation

Each monkey underwent two stereotactic surgeries using isoflurane anesthesia and aseptic techniques, and post-surgical analgesics were administered. The first surgery served to implant a device to limit head motion during experiments. The second surgery consisted of implanting a recording cylinder (Crist Instruments, 2 cm in diameter) over a craniotomy to make

both the right and left inferior colliculi (IC) accessible at an angle of 26 degrees from vertical in the coronal plane (approaching from the right for Monkey J and from the left for Monkey D).

Equipment

During experiments, monkeys were seated comfortably in custom primate chairs (Crist Instruments) with heads restrained. The chair was situated in a dark single-walled sound attenuation chamber (Industrial Acoustics Company) lined with sound attenuating foam (Sonex). Visual stimuli were presented via green light-emitting diodes (LEDs), and sounds were presented from loudspeakers (Cambridge SoundWorks) calibrated to equal volume across speakers (see “Stimuli”). Monkeys performed a saccade task involving looking at the lights and/or sounds as described in detail below (see “Task”). Eye position was monitored using an Oculomatic Pro eye tracking camera and software (500 Hz sampling rate). On correctly-performed trials, juice rewards were delivered using a custom mouth piece and a PVC tube (ClearFLEX, Finger Lakes Extrusion) controlled with a solenoid valve (Parker Hannifin Corporation).

Neural recordings were conducted with 125 mm tungsten microelectrodes (FHC). A 19 mm diameter grid of holes (Crist Instruments) spaced 1 mm apart was placed into the recording cylinder to allow electrode penetrations, which were controlled by a microdrive (NAN Instruments LTD). Neural activity was collected with a Multichannel Acquisition Processor (MAP system, Plexon Inc.), passed through appropriate amplifiers, and sent to SortClient software for online spike sorting. Single unit spiking activity was extracted from signals bandpass filtered between 150 Hz and 8 kHz. Local field potential (LFP) signals were filtered from 0.7 to 300 Hz and stored at a sampling rate of 20 kHz (Plexon, Inc.). Spike times, eye positions, and task parameters were also stored for offline analysis (Plexon, Inc.; Ryklin Software, Inc.).

Stimuli

As noted above, stimuli consisted of sounds and lights originating from loudspeakers and LEDs located on a horizontal row at the monkey's eye level. Speakers and lights were located at -12 and +12 horizontal degrees in the monkey's visual field. Sounds were bandpass filtered noise at center frequencies of 420 or 2000 Hz (11 kHz sampling rate, 55 dB SPL, ±100 Hz bandwidth, 1000 ms on ramp). Sounds were frozen for a given neuron, meaning a single bandpass stimulus was generated for each frequency at the beginning of the session and the same pattern was used throughout the session. Lights were green light-emitting diodes (LEDs) placed on the front face of each speaker and were amplified with a 5.2 volt power source.

Task

Monkeys performed tasks involving saccade targets at either one or two locations; the current paper reports results from the single location trials. To initiate a trial, the monkey fixated on a central light (0 horizontal degrees from the midline, at eye level). After 600-700 ms, a visual, auditory, or combined visual and auditory stimulus was presented at either -12 or 12 degrees while the fixation light remained on. After a further 600-800 ms of fixation, the fixation light was extinguished to cue the monkey to make a saccade to the target (Figure 1). The reinforcement window around the target was 16 degrees wide by 60 degrees tall, to account for low precision in sound localization in the vertical domain (Gnadt et al., 1991; Barton & Sparks, 2001). Monkeys then maintained fixation on the target for 200-250 ms to receive a fluid reward (typically Welch's 100% grape juice diluted 30% in water). The additional two-location trials not analyzed here were very similar, except that two target locations were used simultaneously and the monkeys were required to make sequential saccades to both target locations. All conditions were randomly interleaved.

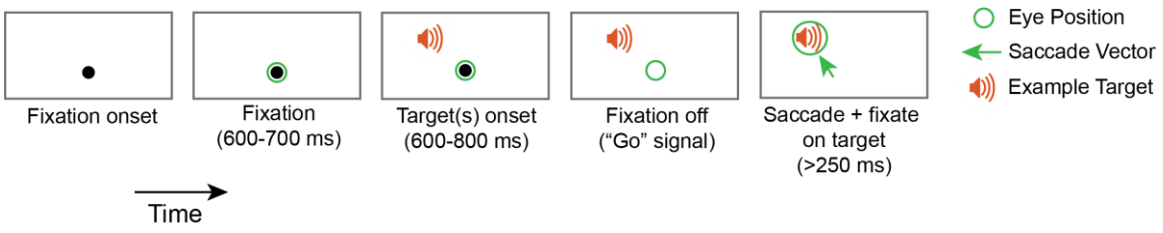


Figure 1: Task structure. Each trial began with the onset of a central light, and the monkey initiated the trial by fixating on the light for a variable period between 600 and 700 ms. A target was then presented (auditory, visual, or both), and the monkey waited 600-800 ms for fixation offset before making a saccade to the target (green arrow). After localizing the target, the monkey received a fluid reward. Trials continued in succession, and conditions were randomly interleaved. Conditions involving two target locations (two sounds with varying numbers of accompanying lights) were also included but are not analyzed here.

Modalities were weighted such that 30% of the trials were visual, 30% of the trials were auditory, and 40% of the trials were audiovisual. Performance was >74% correct for both monkeys for all modalities, with performance on visual-only and audiovisual trials slightly exceeding that for auditory-only trials (Supplementary Figure 1).

Neural Recordings

The left and right IC in each monkey were localized using magnetic resonance imaging scans conducted at Duke University Hospital: sagittal localizing scan, coronal MPRAGE (T1 weighted 3D image, 0.5 mm), coronal T2 1.0 mm 55 slices with phase set right to left, coronal T2 0.5 mm 55 slices with phase set right to left (through the cylinder), and coronal T2 0.5 mm 55 slices with phase set anterior to posterior (through the cylinder). The location of the IC could be reconstructed accurately in reference to the coordinate system of a grid for holding electrodes in the recording cylinder (Crist Instruments). We recorded from the IC using a single tungsten microelectrode (FHC) while the monkey performed the task and neural data were collected as described previously. Most recording sites were found in the low frequency-responsive region of the IC (Bulkin & Groh, 2011).

Data Analysis

All analyses were conducted on correctly performed trials, in which the monkey successfully localized the target. Only recording sites where the neuron was responsive to at least one modality (outlined below) and that had a minimum of 250 correct trials were included for analysis.

Single Unit Activity

We recorded a total of 106 well-isolated single neurons, 82 from the right IC (Monkey J), and 24 from the left (Monkey D). Results were similar across monkeys and were combined for subsequent analysis.

Auditory responsiveness was assessed by comparing the activity during a time window after sound onset to a corresponding period of steady fixation immediately prior to sound onset. Two windows were used, either 0-100 ms or 0-500 ms after sound onset, to account for the presence of primarily transient vs. transient+sustained response patterns, which are known to occur in IC neurons (Ryan & Miller, 1978; Willott & Urban, 1978; Aitkin et al., 1994; Syka et al., 2000; Bulkin & Groh, 2011). Units that exhibited a significant change in activity relative to baseline in either of these periods, or that passed visual inspection of peri-stimulus time histograms pooled across all auditory conditions, were considered auditory-responsive (pooled across all sound-only trials regardless of sound location or frequency, paired one-tailed t-test comparing firing rates before and after stimulus onset, $p<0.025$ for each tail, equivalent to $p<0.05$ for a two-tailed test). 106 neurons met these criteria.

Visual responsiveness was also assessed in several ways. Since each trial involved a visual fixation stimulus, activity was analyzed in a time window 50-250 ms after the onset of that fixation stimulus in comparison to a comparable baseline prior to fixation stimulus onset. This

time window was chosen based on prior work, which showed that visual responses in the IC occur with a slight delay relative to stimulus onset (Gutfreund et al., 2002; Porter et al., 2007). The position of this fixation stimulus on the retina naturally varied based on where the eyes happened to be when the fixation stimulus appeared. We also examined responsiveness in a similar time period following entry of the eye into the fixation window, again in comparison to baseline (50-250 ms prior to fixation). Finally, on trials in which a visual saccade target was presented, we examined responsiveness 50-250 ms following target onset. As with the tests of auditory responsiveness, we used a paired one-tailed t-test, $p<0.025$ for each tail, deployed on all the trials involving visual stimuli regardless of their locations. Detailed results will be presented further below.

Local Field Potentials

Local field potential (LFP) signals were successfully analyzed from 83 of the 106 recording sites (the remainder were excluded due to technical problems affecting alignment of these data with the trial event data). Signals were normalized by subtracting the mean voltage during a 200 ms baseline period prior to fixation light onset across all trials for each recording site. Recording sites were included based on the significance of the associated single unit activity rather than LFP responsiveness.

Results

Local Field Potential

Given that the IC is fundamentally an auditory structure, visually-evoked activity is expected to be weaker and rarer than auditory-evoked activity. Accordingly, the LFP is a

particularly useful measure because it represents aggregate activity from a population of neurons, and can include pre- and post-synaptic activity in that population regardless of whether the threshold for triggering a spike was crossed. Therefore, we first probed the LFP, and we considered three points in time during the trial: the onset of the visual fixation light, the time of entry of the eyes into the reinforcement window around the fixation light, and the onset of the visual target (for trials with visual saccade targets).

Figure 2 shows LFP signals aligned to each of these time points. Figure 2a shows results aligned on fixation light onset for each recording site (individual gray traces). A distinct wave-like response occurs after fixation light onset for nearly every recording site. This wave exhibits a positive, then negative, peak before settling at a slightly negative plateau. In Figure 2b, the traces are aligned on the eye's entry into the window around the fixation light, and a fresh visual response can be observed as the saccade brings the fixation light to a new (more foveal) retinal location. This response is most evident in the mean across sites (red trace). Finally, Figure 2c shows the responses to a visual target presented at either -12° or 12° during steady fixation (data are combined across both locations). These responses appear slightly weaker than those seen in connection to the onset of the visual fixation light. This could be because (a) they are added on top of any activity evoked by that target, so the deviation from the voltage level prior to target onset is modest, (b) they are presented from only 2 retinal locations, only 1 of which is in the contralateral hemifield, or (c) visually-evoked activity in the IC is context dependent and depends on the role or timing of the stimulus in the task. We will discuss these possibilities further in the Discussion.

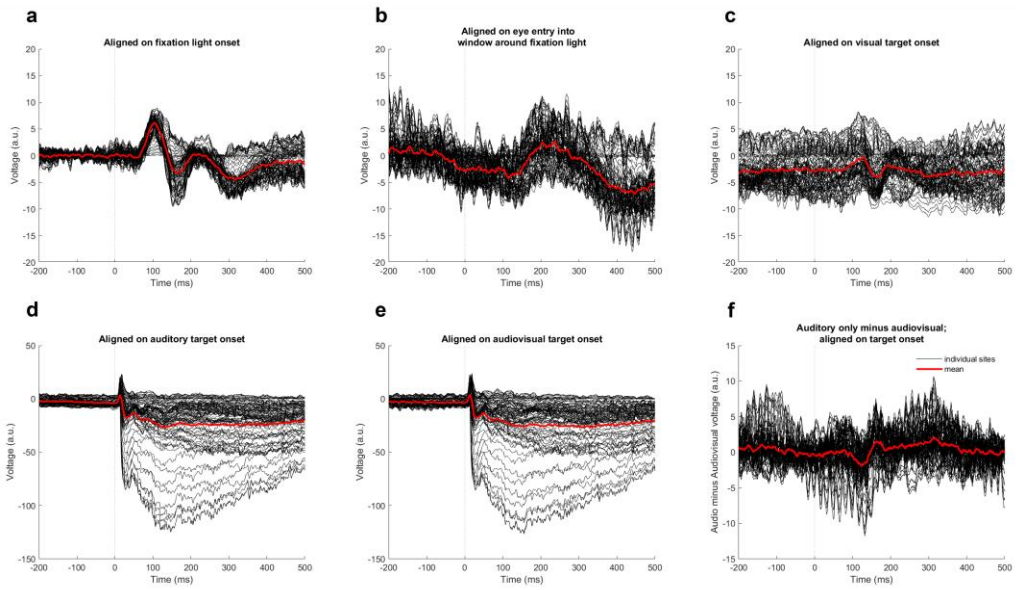


Figure 2: Local field potentials display responses to fixation and target lights in addition to sounds. Local field potential (LFP) aligned to various stages of the trial. a) LFP aligned on the onset of a fixation light. b) LFP aligned on the time of the monkey's eye entry into a spatial window surrounding the fixation light. Note that sometimes the monkey anticipated the onset of the fixation light and no saccade was required; such trials are excluded from this analysis. c) LFP aligned on the onset of a visual target during steady fixation elsewhere. d) LFP aligned on the onset of an auditory target. e) LFP aligned on the onset of an audiovisual target. f) Difference in LFP shown in (d-e). In all subpanels, gray traces represent individual recording sites, and thick red traces represent the mean across all recording sites. Traces are averaged across all relevant stimulus conditions (e.g., traces aligned on auditory target onset in panel d include all sound frequencies and locations; those in e include all sound frequencies and locations that were paired with visual stimuli). Note the change in y-axis scale in (d-e) compared to (a-c), reflecting stronger auditory and audiovisual than visual responses. See Supplementary Figure 2 for a depiction of the mean traces accompanied by standard error for easier comparison to baseline responses.

In contrast, the sound-evoked LFP is quite strong. Figure 2d shows the signals on auditory trials (pooled across both frequencies and locations). Note the y-axis is adjusted to accommodate these much larger signals. Other than the greater magnitude, their character was similar to the visually-evoked LFPs, again usually beginning with a short-latency and short-duration positive peak, followed by a steadier and longer-lasting negative plateau. In other

words, the peak sequences were similar, but occurred much faster on auditory than on visual trials.

The LFP evoked on audiovisual trials (again pooled across all stimulus combinations) is grossly similar to that observed on auditory-only trials (Figure 2e). However, when the mean audiovisual LFP for each site is subtracted from the corresponding auditory one (Figure 2f), a modest influence of the inclusion of the visual stimulus appears, particularly when averaged across sites (red trace deflects away from zero, first negative, then positive peaks).

Together, these findings reveal that visual signals are indeed reaching the IC in the context of this task, and that, as previously shown (Porter et al., 2007; Bulkin & Groh, 2012b), the most obvious visual influences can be seen at the onset of the fixation light. Even so, the presentation of a visual target later in the trial does cause a measurable signal and a measurable modulation of auditory signals.

Spiking Responses

We next considered individual neurons' spiking responses to the onset of the fixation light, as previously reported by (Porter et al., 2007). Generally, these responses were quite weak. Figure 3a shows the individual (gray) and average peri-stimulus time histograms (PSTHs) (red) across sites. A few sites show distinct excitatory or inhibitory responses, but, in general, any visual effect on actual spiking was relatively uncommon in our dataset: the difference in the number of spikes observed 50-250 ms after fixation light onset was usually not very different from the number of spikes observed in the same window before fixation light onset (Figure 3b, Z-scores calculated by dividing the difference between the mean light-evoked and baseline firing rates by the standard deviation of baseline firing rates). In total, 8% of IC neurons showed significant excitatory responses (green bars) and 17% showed significant inhibitory responses

(red bars; t-test, $p<0.025$ for each tail, equivalent to $p<0.05$ for a two-tailed test). One possible explanation for the low prevalence of visual responses is that the visual receptive fields found in this population of IC neurons may be poorly probed by “chance” variation in eye position at the time of fixation stimulus onset. Figure 3c-f show the individual PSTHs (c,d) and receptive field maps (e,f) for the two most responsive neurons in the sample – while they are strongly responsive, their receptive fields are quite small and centered on the fovea. The fixation light only appeared at one location, and the monkeys’ eyes often lingered at or near that location while waiting for the next trial to start. This may have privileged detecting spike responses in a small set of foveally-sensitive IC neurons while disadvantaging the detection of spiking responses at other receptive field locations. The LFP signals may have been less sensitive to this limitation, by virtue of pooling over a larger population of neurons and/or inclusion of subthreshold signals.

Similarly, Figure 3g shows the individual (gray) and average PSTHs (red) across sites for spiking responses to the onset of the target light during fixation. Like responses to the fixation light, these responses show visual effects on spiking activity in about 17% of the neurons (Figure 3h; 4% excitatory, green bars; 13% inhibitory, red bars; t-test, $p<0.025$, equivalent to two-tailed t-test $p<0.05$). In total, 37% of neurons exhibited significant responses in one or the other of these two time periods, with 6% responsive to both. When a neuron was responsive during both time periods, the direction of modulation (excitatory vs. inhibitory) was typically the same in these two cases. When separating out by target location, slightly more neurons were responsive to the contralateral stimulus (4% excited, 4% inhibited, or 8% overall) than to the ipsilateral target (2% excited, 4% inhibited, or 6% overall). (The lower proportion of responsive neurons when limiting to just one target location may be due to the impact of cutting the data in half on the odds of reaching statistical significance.) The prevalence of visual

responsiveness was also similar in both monkeys: 23% of neurons in Monkey J and 33% of neurons in Monkey D were responsive to the fixation light, while 13% of neurons in Monkey J and 29% of neurons in Monkey D were responsive to the visual target. In summary, spiking modulation occurred in response to both fixation and target lights, but was rarer in this IC dataset in comparison to previous reports (Porter et al., 2007; Bulkin & Groh, 2012b), perhaps due to the more sparse sampling of visual space.

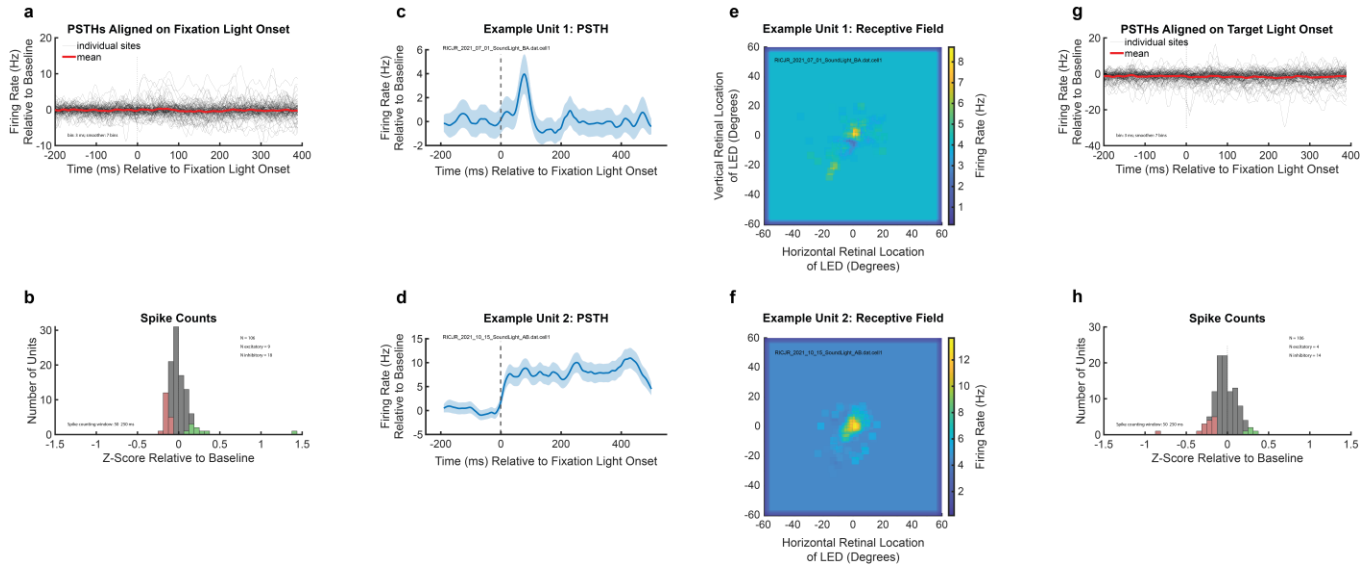


Figure 3: Spiking activity in the IC reflects weak and rare responses to fixation and target lights during a localization task. a) PSTHs of individual units (gray) and mean PSTH (red) for responses aligned to the onset of a fixation light presented at the horizontal midline and at eye level. b) Z-scores calculated as a difference between the mean light-evoked and baseline firing rates divided by the standard deviation of baseline firing rates. Green and red bars indicate units with a score significantly different from zero. c-d) Example responses of two individual neurons to the fixation light. Baseline firing rates were subtracted, and PSTHs were smoothed by convolving with a 5-point sliding triangle window and cubic spline interpolation. Shaded areas represent standard error. e-f) Spatial receptive fields of the same neurons as in (c-d), showing each neuron's response to an LED fixation light falling on various horizontal and vertical locations on the retina during passive looking at the start of a trial. Baseline firing rates are indicated by the background color. Activity patterns were smoothed both horizontally and vertically using a 5-degree window. g) PSTHs of individual units (gray) and mean PSTH (red) for responses aligned to the onset of a target light. h) Z-scores calculated as a difference between the mean light-evoked and baseline firing rates divided by the standard deviation of baseline firing rates, using a spike counting window 50-250 ms after stimulus onset. Green and red bars

indicate units with a score significantly different from zero. Results in panels g, and h are pooled across both visual target locations for individual neurons. Responses were also evaluated separately for the contralateral and ipsilateral targets (significant responses to the contralateral target: 4% excitatory, 4% inhibitory; for the ipsilateral target: 2% excitatory, 4% inhibitory; results not shown).

We next asked how visual stimuli affect IC spiking activity when presented in combination with sounds. Specifically, we examined responses aligned to the onset of visual, auditory, and audiovisual targets following fixation onset during the task. Figure 4 shows the responses of four example neurons to all combined presentations of visual (blue), auditory (orange), and audiovisual (purple) targets at single locations. Although these neurons were not responsive to visual input (blue), pairing that same visual cue with a sound (orange) produced a different auditory response (purple) – either by enhancing (Figure 4a) or suppressing (Figure 4b,c) the neuron's auditory response. However, this was not always the case: Figure 4d shows an example of a neuron whose auditory responses appeared to be unaffected by the simultaneous presence of a visual stimulus.

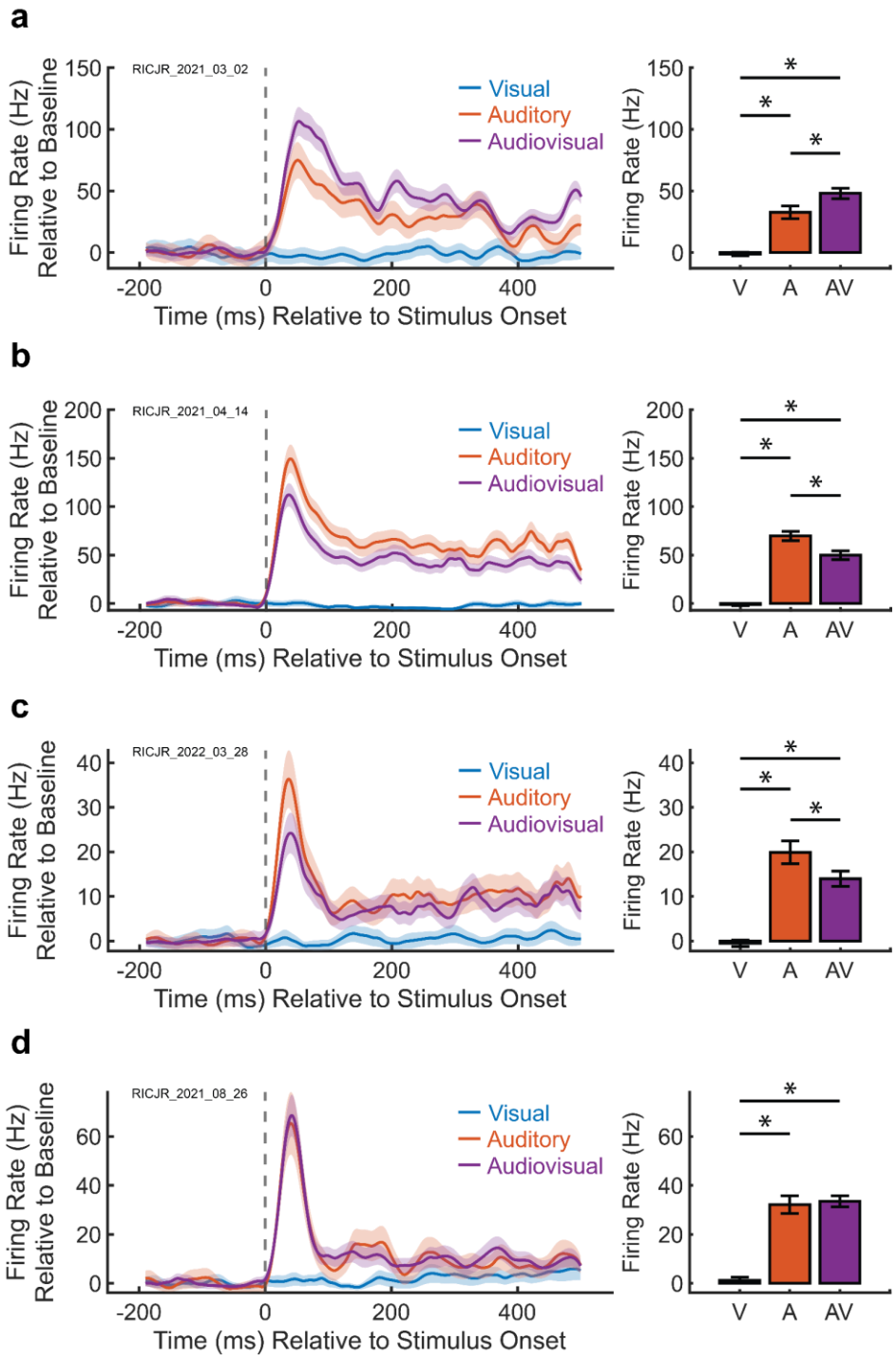


Figure 4: A visual cue can modulate a neuron's auditory response. Example responses of four individual neurons to pooled combinations of all visual (blue), auditory (orange), and audiovisual (purple) conditions presented at a single location. Within each panel, the left subpanel is a peri-stimulus time histogram (PSTH) of each modality, and the right subpanel is a quantification of the average responses either in the first 500 ms after stimulus onset (0-500 ms)

for sustained responses, or in the first 100 ms after stimulus onset (0-100 ms) for transient responses. Baseline firing rates were subtracted, and PSTHs were smoothed by convolving with a 5-point sliding triangle window and cubic spline interpolation. Shaded areas and error bars represent standard error. a) The presence of a visual cue augments the auditory response, and both the auditory and audiovisual responses are sustained throughout the stimulus presentation. b) The presence of a visual cue dampens the auditory response, and both the auditory and audiovisual responses are sustained throughout the stimulus presentation. c) The presence of a visual cue dampens the auditory response only immediately after stimulus onset, and the auditory and audiovisual responses are similar later in the stimulus presentation window. d) Auditory and audiovisual responses are similar, indicating no effect of the visual cue on auditory responsiveness. Note low visual responsiveness in all panels, indicating that the presence of a visual cue can modulate a neuron's auditory response even if the neuron is unresponsive to visual input alone.

Given the low magnitude and low prevalence of visually-driven responses in these neurons (Figure 3), the visually-induced modulation of auditory responsiveness (Figure 4) is quite striking. These results suggest that although visual input may not drive a neuron's spiking activity individually, pairing the same visual cue with a driving auditory input can be sufficient to modulate a neuron's response, potentially to inform perceptual output.

We next quantified these effects at the population level. Figure 5a shows the differences between the responses to an audiovisual stimulus and the corresponding auditory stimulus alone. About 15% of conditions showed a significant decrease in responsiveness (red/green bars, t-test, $p<0.025$ for each tail, equivalent to $p<0.05$ for a two-tailed test), whereas 7.5% showed a significant increase in responsiveness.

The direction of the effect was not strongly related to whether the visual response was excitatory on its own, nor was it strongly related to the magnitude of the auditory response on its own. Figure 5b shows the mean audiovisual response as a function of the mean auditory response, with color indicating whether the visual response was excitatory (red) or inhibitory (blue) relative to baseline. Both red and blue points lie both above and below the diagonal line of slope one, indicating that responses could be either increased or decreased when the visual and auditory stimuli were combined. Figure 5c shows the same pattern but from a different

perspective – here the differences between the audiovisual and auditory responses are plotted as a function of the visual responses, with color indicating the magnitude of the auditory responses. Points lie in all 4 quadrants, indicating that both enhanced and reduced responses to audiovisual stimuli can occur regardless of whether the visual input is slightly excitatory or slightly inhibitory when presented on its own. There is little evident relationship to the strength of the auditory response: darker red points lie in all four quadrants. These results further support the finding that the presence of a visual cue can modulate auditory responsiveness in the IC, but reveal that the direction of the effect – increasing vs. decreasing activity – can vary across the population and is not easily predicted from either the response to sound alone or the response to the visual stimulus alone.

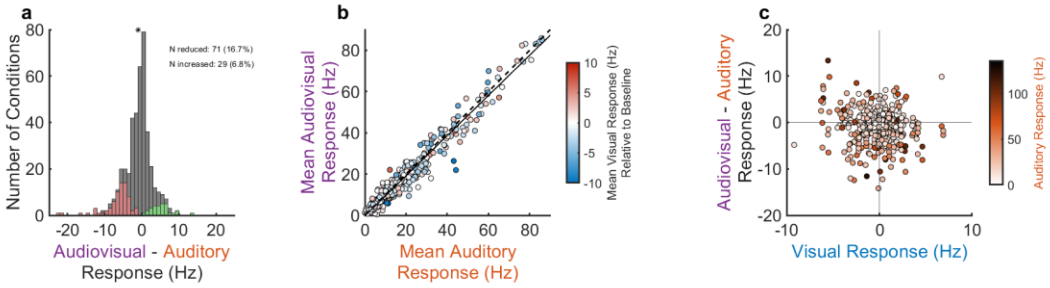


Figure 5: Visual influences on auditory responses are varied. a) Distribution of audiovisual minus auditory responses. Visual stimuli added to sounds reduced activity in about 15% of conditions (red bars) whereas they increased the activity in about 7% of conditions (green bars; two-tailed t-test, $p<0.025$). In net, the population activity on audiovisual trials was slightly lower than on their auditory counterparts ($p = 0.0383$, paired sample t-test). Asterisk indicates distribution mean. Green and red bars indicate conditions with a value significantly different from zero. b) Comparison of mean auditory and audiovisual firing rates for each neuron, with colors

indicating the corresponding visual firing rate for a visual cue presented at the same location. No strong relationship arises between visual firing rates and auditory or audiovisual firing rates, with both red and blue points falling both above and below the unity line. For clarity, visual responses are depicted relative to baseline, such that positive values indicate an excitatory response and negative values indicate an inhibitory one. No such baseline subtraction was applied to the auditory or audiovisual trials, but most were excitatory. c) Comparison of mean difference between audiovisual and auditory firing rates and baseline-subtracted visual firing rates, with colors indicating the corresponding auditory firing rate. Points arise in all four quadrants, indicating a wide degree of visual influences on auditory firing rates. In all subpanels, each neuron contributes 4 data points as sound frequencies and visual cue locations are counterbalanced (4 frequency-location pairs per neuron).

Discussion

Visual cues influence auditory perception at the behavioral level, but the role of early, predominantly unisensory areas in this process has remained relatively unexplored. Here we investigated how visual cues modulate activity in the IC at population and single-unit scales and in response to visual stimuli presented alone or with sounds. We found that fixation lights and visual targets can evoke distinct wave-like responses in local field potential recordings, and that single units exhibit visual responses that are weaker but statistically evident in a modest subpopulation of neurons. Intriguingly, individual units' responses to sounds can be modulated by the presence of a simultaneous visual cue, but in directions that are not easily accounted for by either the auditory or the visual responses alone. Indeed, modulation of auditory signals could be seen even when the visual stimulus did not itself evoke an overt spiking response. This suggests a complex form of interaction between visual and auditory signals in this nominally auditory structure.

These findings extend previous work that identified visual and eye movement-related responses (Porter et al., 2007; Bulkin & Groh, 2012b) and eye position-dependent auditory responses (Groh et al., 2001; Zwiers et al., 2004; Porter et al., 2006) in the IC. The chief difference between our current study and the earlier studies from our group (Porter et al., 2007; Bulkin & Groh, 2012b) is that the driven visual responses appeared to be less prevalent here. It

is not yet clear what may account for this difference. One possibility is that the current study sampled visual space in a more limited way – there was only a single fixation location and only two visual stimulus locations, limiting the sampling of retinal receptive fields. Another possible factor is that the recordings in the current study were restricted to particular spatial areas within the IC, whereas previous work intentionally sampled a wide range of locations spanning a large portion of the IC (Bulkin & Groh, 2012b). The difference in the variety of recording sites sampled may mean that the recording sites in the current study had a lower overall degree of visual input than in prior studies. It is known that the IC's subregions differ in the prevalence of visual receptive fields with lower incidence in the low frequency-responsive region that we tested here (Bulkin & Groh, 2012b). Other attributes of the stimuli, tasks, or research subjects may also contribute to differences in visual responsiveness across studies.

Our results are similar to those in auditory cortex, where a variety of multisensory interactions have been identified. In auditory cortex, visual cues can elicit their own responses as well as modulate auditory responses in both excitatory and inhibitory ways (Kayser et al., 2007, 2008). Non-auditory cues can also modulate current source density in auditory cortex, either by eliciting direct visual responses (Schroeder & Foxe, 2002) or by modulating an auditory response by resetting the phase of neural oscillations (Lakatos et al., 2007). These effects are also evident for stimuli with high behavioral relevance; images or videos of faces can elicit their own independent responses as well as modulate LFP activity in response to vocalizations (Ghazanfar, 2005). Finally, visual stimuli can modulate auditory responsiveness in more complicated ways, such as influencing which of two sounds is predominantly represented by individual neurons or the neural population (Atilgan et al., 2018).

Together, our results and those of these other prior studies in auditory cortex differ from related investigations in more multimodal areas such as the superior colliculus (SC). In the SC,

which is reciprocally connected to the IC (Woollard & Harpman, 1940; Moore & Goldberg, 1963; Powell & Hatton, 1969; Harting, 1977; Adams, 1980; Covey et al., 1987; Coleman & Clerici, 1987; Sparks & Hartwich-Young, 1989; Doubell et al., 2000; Hyde & Knudsen, 2000; Caño et al., 2006), spiking responses to both visual and auditory stimuli are common. Combinations of visual and auditory stimuli are generally thought to increase the activity of individual neurons, most often in an additive fashion (Meredith & Stein, 1986; Wallace et al., 1998; Stanford et al., 2005; Alvarado et al., 2007). In contrast, audiovisual responses in the IC can be either slightly higher or slightly lower than auditory-alone responses. These patterns occurred even when a neuron had no visually-driven response, and rarely appeared to be additive.

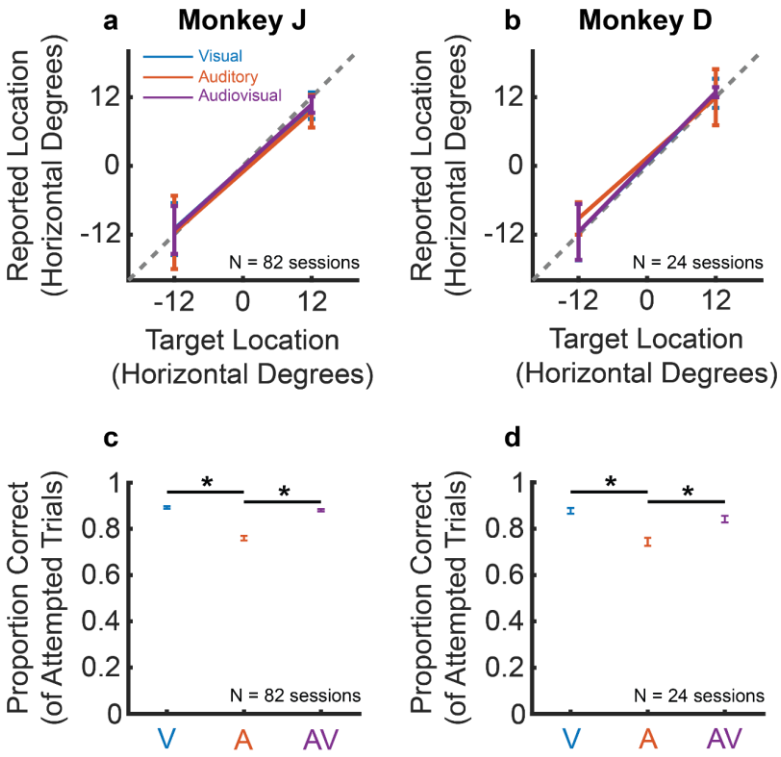
A likely factor in these differences is that IC is thought of as a general-purpose auditory area, in contrast to the multimodal/oculomotor SC. The IC is located only a few synapses along the auditory pathway from the cochlea, and nearly all ascending auditory input passes through this brain region. Thus, it is likely that the overall purpose of this structure is to process sound, and that any visual signals in this area serve a purpose of shaping those responses rather than simply adding to them. Such shaping is likely to take multiple forms and not reduce easily to a simple linear operation in which stronger input signals lead to more output (spiking) activity. Our study provides a first look at the scope of the IC's visual-auditory interactions and the factors that govern them. More work will be needed to explore such factors. For example, full mapping of visual receptive fields should be combined with assessing the role of visual-auditory spatial alignment on IC activity. This is a particularly interesting unexplored domain because the native codes for visual and auditory space are so different. The optics of the eye cause light from different locations in the environment to land at particular locations on the retinal surface, leading neurons throughout the visual pathway to have circumscribed receptive fields. In contrast, sound location is computed in part from binaural difference cues, which are

proportional to the horizontal eccentricity of a sound. Neurons in the primate IC (and SC) do not exhibit circumscribed receptive fields, but rather appear to encode auditory space as a meter or rate code in which a large population of neurons responds with a particular level of activity, depending on where a sound is located along the azimuth (primate IC: Groh et al., 2003; Zwiers et al., 2004; primate SC: Lee & Groh, 2014; see also: Middlebrooks et al., 1994, 1998; McAlpine & Grothe, 2003; Werner-Reiss & Groh, 2008; Salminen et al., 2009). The brain must combine these disparate visual and auditory codes to understand where visual and auditory signals arise in space, and the IC is emerging as a site where such combination may occur.

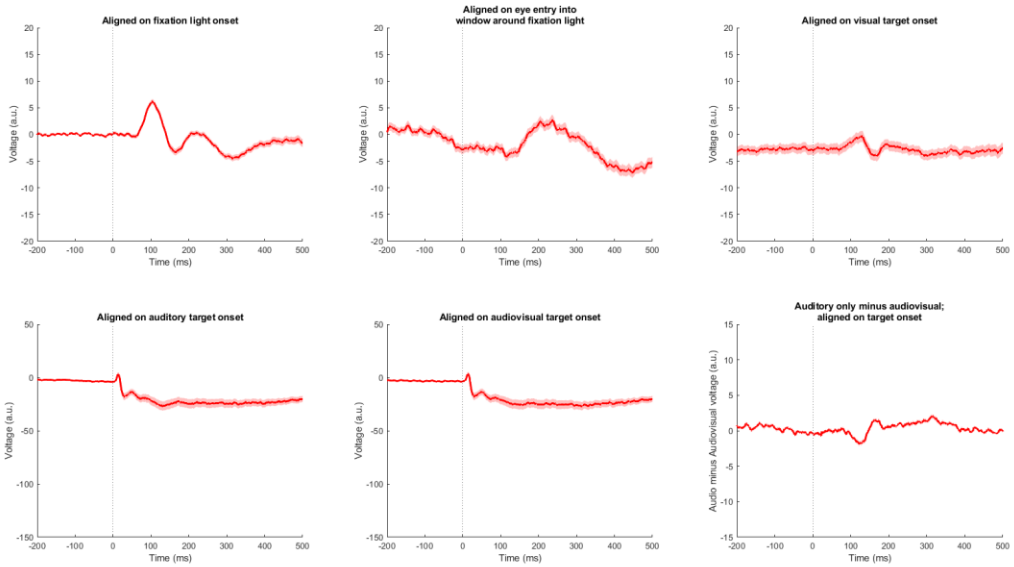
In addition, further work is needed in the frequency domain. For example, at the behavioral level, visual lip reading cues can modulate the perception of speech sounds (e.g., the McGurk effect, (Sams et al., 1991; Wright et al., 2003; Ghazanfar, 2005; Kayser et al., 2007, 2008; Hoffman et al., 2008; Nath & Beauchamp, 2012)). This perceptual phenomenon might emerge from visually-induced changes in the representation of sound frequency within the auditory pathway. Further exploration is needed to ascertain the properties that govern these neural responses and their perceptual correlates.

Collectively, these experiments provide insight into the influence of visual cues on auditory perception. Together with the other studies in related auditory regions, the findings presented here shed new light on the connections between multiple sensory systems and the ways in which context from another modality can inform processing of a brain region's "dominant" modality. As the field of neuroscience moves forward, it will be critical to consider how sensory systems interact to give rise to complex behaviors in varied sensory environments. Growing evidence points to an intriguing possibility: sensory systems may not operate as independently as was first thought.

Supplementary Materials



Supplementary Figure 1: Monkeys accurately perform an audiovisual localization task. a-b) Mean and standard error of saccade endpoints during trials performed by each monkey for visual, auditory, and audiovisual stimuli presented at the two horizontal target locations in the task. Endpoints for all three modalities are largely overlapping and are aligned with the actual target location. c-d) Proportion of attempted trials of each modality in which each monkey correctly localized the targets within a 16-degree wide by 60-degree tall window. Visual and audiovisual conditions had a higher proportion of correct trials than auditory conditions, but all types of conditions were performed at an average accuracy of approximately 74% or greater.



Supplementary Figure 2: Mean local field potentials diverge from baseline in responses to fixation lights, target lights, and sounds. Same mean local field potential (LFP) responses as shown in Figure 2, with shaded areas representing standard error. As in Figure 2, responses are aligned to various stages of the trial. Standard error remains narrow and diverges from baseline in all cases.

References

Adams, J. C. (1980). Crossed and descending projections to the inferior colliculus. *Neuroscience Letters*, *19*(1), 1–5. https://doi.org/10.1016/0304-3940(80)90246-3

Aitkin, L., Tran, L., & Syka, J. (1994). The responses of neurons in subdivisions of the inferior colliculus of cats to tonal, noise and vocal stimuli. *Experimental Brain Research*, *98*(1), 53–64. https://doi.org/10.1007/BF00229109

Alvarado, J. C., Vaughan, J. W., Stanford, T. R., & Stein, B. E. (2007). Multisensory Versus Unisensory Integration: Contrasting Modes in the Superior Colliculus. *Journal of Neurophysiology*, *97*(5), 3193–3205. https://doi.org/10.1152/jn.00018.2007

Atilgan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A. K. C., & Bizley, J. K. (2018). Integration of Visual Information in Auditory Cortex Promotes Auditory Scene Analysis through Multisensory Binding. *Neuron*, *97*(3), 640-655.e4. https://doi.org/10.1016/j.neuron.2017.12.034

Barton, E. J., & Sparks, D. L. (2001). Saccades to remembered targets exhibit enhanced orbital position effects in monkeys. *Vision Research*, *41*(18), 2393–2406. https://doi.org/10.1016/S0042-6989(01)00130-4

Bulkin, D. A., & Groh, J. M. (2006). Seeing sounds: Visual and auditory interactions in the brain. *Current Opinion in Neurobiology*, *16*(4), 415–419. https://doi.org/10.1016/j.conb.2006.06.008

Bulkin, D. A., & Groh, J. M. (2011). Systematic mapping of the monkey inferior colliculus reveals enhanced low frequency sound representation. *Journal of Neurophysiology*, *105*(4), 1785–1797. https://doi.org/10.1152/jn.00857.2010

Bulkin, D. A., & Groh, J. M. (2012a). Distribution of eye position information in the monkey inferior colliculus | Journal of Neurophysiology. *Journal of Neurophysiology*, *107*, 785–795. https://doi.org/10.1152/jn.00662.2011

Bulkin, D. A., & Groh, J. M. (2012b). Distribution of visual and saccade related information in the monkey inferior colliculus. *Frontiers in Neural Circuits*, *6*. https://doi.org/10.3389/fncir.2012.00061

Caño, G. G. D., Gerrikagoitia, I., Alonso-Cabria, A., & Martínez-Millán, L. (2006). Organization and origin of the connection from the inferior to the superior colliculi in the rat. *Journal of Comparative Neurology*, *499*(5), 716–731. https://doi.org/10.1002/cne.21107

Chandrasekaran, C. (2017). Computational principles and models of multisensory integration. *Current Opinion in Neurobiology*, *43*, 25–34. https://doi.org/10.1016/j.conb.2016.11.002

Coleman, J. R., & Clerici, W. J. (1987). Sources of projections to subdivisions of the inferior colliculus in the rat. *The Journal of Comparative Neurology*, *262*(2), 215–226. https://doi.org/10.1002/cne.902620204

Cooper, M. H., & Young, P. A. (1976). Cortical projections to the inferior colliculus of the cat. *Experimental Neurology*, *51*(2), 488–502. https://doi.org/10.1016/0014-4886(76)90272-7

Covey, E., Hall, W. C., & Kobler, J. B. (1987). Subcortical connections of the superior colliculus in the mustache bat, Pteronotus parnellii. *Journal of Comparative Neurology*, *263*(2), 179–197. https://doi.org/10.1002/cne.902630203

Doubell, T. P., Baron, J., Skaliora, I., & King, A. J. (2000). Topographical projection from the superior colliculus to the nucleus of the brachium of the inferior colliculus in the ferret: Convergence of visual and auditory information. *European Journal of Neuroscience*, *12*(12), 4290–4308. https://doi.org/10.1111/j.1460-9568.2000.01337.x

Ghazanfar, A. A. (2005). Multisensory Integration of Dynamic Faces and Voices in Rhesus Monkey Auditory Cortex. *Journal of Neuroscience*, *25*(20), 5004–5012. https://doi.org/10.1523/JNEUROSCI.0799-05.2005

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, *10*(6), 278–285. https://doi.org/10.1016/j.tics.2006.04.008

Gnadt, J. W., Bracewell, R. M., & Andersen, R. A. (1991). Sensorimotor transformation during eye movements to remembered visual targets. *Vision Research*, *31*(4), 693–715. https://doi.org/10.1016/0042-6989(91)90010-3

Groh, J. M., Kelly, K. A., & Underhill, A. M. (2003). A Monotonic Code for Sound Azimuth in Primate Inferior Colliculus. *Journal of Cognitive Neuroscience*, *15*(8), 1217–1231. https://doi.org/10.1162/089892903322598166

Groh, J. M., Trause, A. S., Underhill, A. M., Clark, K. R., & Inati, S. (2001). Eye Position Influences Auditory Responses in Primate Inferior Colliculus. *Neuron*, *29*(2), 509–518. https://doi.org/10.1016/S0896-6273(01)00222-7

Gutfreund, Y., Zheng, W., & Knudsen, E. I. (2002). Gated Visual Input to the Central Auditory System. *Science*, *297*(5586), 1556–1559. https://doi.org/10.1126/science.1073712

Harting, J. K. (1977). Descending pathways from the superior colliculus: An autoradiographic analysis in the rhesus monkey (Macaca mulatta). *Journal of Comparative Neurology*, *173*(3), 583–612. https://doi.org/10.1002/cne.901730311

Hoffman, K. L., Ghazanfar, A. A., Gauthier, I., & Logothetis, N. K. (2008). Category-specific responses to faces and objects in primate auditory cortex. *Frontiers in Systems Neuroscience*, *2*. https://doi.org/10.3389/neuro.06.002.2007

Hyde, P. S., & Knudsen, E. I. (2000). Topographic projection from the optic tectum to the auditory space map in the inferior colliculus of the barn owl. *The Journal of Comparative Neurology*, *421*(2), 146–160.

Itaya, S. K., & Van Hoesen, G. W. (1982). Retinal innervation of the inferior colliculus in rat and monkey. *Brain Research*, *233*(1), 45–52. https://doi.org/10.1016/0006-8993(82)90928-3

Kayser, C., Petkov, C. I., Augath, M., & Logothetis, N. K. (2007). Functional Imaging Reveals Visual Modulation of Specific Fields in Auditory Cortex. *Journal of Neuroscience*, *27*(8), 1824–1835. https://doi.org/10.1523/JNEUROSCI.4737-06.2007

Kayser, C., Petkov, C. I., & Logothetis, N. K. (2008). Visual Modulation of Neurons in Auditory Cortex. *Cerebral Cortex*, *18*(7), 1560–1574. https://doi.org/10.1093/cercor/bhm187

Lakatos, P., Chen, C.-M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex. *Neuron*, *53*(2), 279–292. https://doi.org/10.1016/j.neuron.2006.12.011

Lee, J., & Groh, J. M. (2014). Different Stimuli, Different Spatial Codes: A Visual Map and an Auditory Rate Code for Oculomotor Space in the Primate Superior Colliculus. *PLOS ONE*, *9*(1), e85017. https://doi.org/10.1371/journal.pone.0085017

McAlpine, D., & Grothe, B. (2003). Sound localization and delay lines – do mammals fit the model? *Trends in Neurosciences*, *26*(7), 347–350. https://doi.org/10.1016/S0166-2236(03)00140-1

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, *56*(3), 640–662. https://doi.org/10.1152/jn.1986.56.3.640

Middlebrooks, J. C., Clock, A. E., Xu, L., & Green, D. M. (1994). A panoramic code for sound location by cortical neurons. *Science*, *264*(5160), 842–844. https://doi.org/10.1126/science.8171339

Middlebrooks, J. C., Xu, L., Eddins, A. C., & Green, D. M. (1998). Codes for Sound-Source Location in Nontonotopic Auditory Cortex. *Journal of Neurophysiology*, *80*(2), 863–881. https://doi.org/10.1152/jn.1998.80.2.863

Moore, R. Y., & Goldberg, J. M. (1963). Ascending projections of the inferior colliculus in the cat. *Journal of Comparative Neurology*, *121*(1), 109–135. https://doi.org/10.1002/cne.901210109

Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *NeuroImage*, *59*(1), 781–787. https://doi.org/10.1016/j.neuroimage.2011.07.024

Paloff, A. M., Usunoff, K. G., Hinova-Palova, D. V., & Ivanov, D. P. (1985). Retinal innervation of the inferior colliculus in adult cats: Electron microscopic observations. *Neuroscience Letters*, *54*(2), 339–344. https://doi.org/10.1016/S0304-3940(85)80101-4

Porter, K. K., Metzger, R. R., & Groh, J. M. (2006). Representation of Eye Position in Primate Inferior Colliculus. *Journal of Neurophysiology*, *95*(3), 1826–1842. https://doi.org/10.1152/jn.00857.2005

Porter, K. K., Metzger, R. R., & Groh, J. M. (2007). Visual- and saccade-related signals in the primate inferior colliculus. *Proceedings of the National Academy of Sciences*, *104*(45), 17855–17860. https://doi.org/10.1073/pnas.0706249104

Powell, E. W., & Hatton, J. B. (1969). Projections of the inferior colliculus in cat. *Journal of Comparative Neurology*, *136*(2), 183–192. https://doi.org/10.1002/cne.901360205

Ryan, A., & Miller, J. (1978). Single unit responses in the inferior colliculus of the awake and performing rhesus monkey. *Experimental Brain Research*, *32*(3), 389–407. https://doi.org/10.1007/BF00238710

Salminen, N. H., May, P. J. C., Alku, P., & Tiitinen, H. (2009). A Population Rate Code of Auditory Space in the Human Cortex. *PLOS ONE*, *4*(10), e7600. https://doi.org/10.1371/journal.pone.0007600

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S.-T., & Simola, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, *127*(1), 141–145. https://doi.org/10.1016/0304-3940(91)90914-F

Schmehl, M. N., & Groh, J. M. (2021). Visual Signals in the Mammalian Auditory System. *Annual Review of Vision Science*, *7*(22), 1–23. https://doi.org/10.1146/annurev-vision-091517-034003

Schroeder, C. E., & Foxe, J. J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cognitive Brain Research*, *14*(1), 187–198. https://doi.org/10.1016/S0926-6410(02)00073-3

Sparks, D. L., & Hartwich-Young, R. (1989). The deep layers of the superior colliculus. *Reviews of Oculomotor Research*, *3*, 213–255.

Stanford, T. R., Quessy, S., & Stein, B. E. (2005). Evaluating the Operations Underlying Multisensory Integration in the Cat Superior Colliculus. *Journal of Neuroscience*, *25*(28), 6499–6508. https://doi.org/10.1523/JNEUROSCI.5095-04.2005

Syka, J., Popelář, J., Kvašňák, E., & Astl, J. (2000). Response properties of neurons in the central nucleus and external and dorsal cortices of the inferior colliculus in guinea pig. *Experimental Brain Research*, *133*(2), 254–266. https://doi.org/10.1007/s002210000426

Wallace, M. T., Meredith, M. A., & Stein, B. E. (1998). Multisensory Integration in the Superior Colliculus of the Alert Cat. *Journal of Neurophysiology*, *80*(2), 1006–1010. https://doi.org/10.1152/jn.1998.80.2.1006

Werner-Reiss, U., & Groh, J. M. (2008). A Rate Code for Sound Azimuth in Monkey Auditory Cortex: Implications for Human Neuroimaging Studies. *Journal of Neuroscience*, *28*(14), 3747–3758. https://doi.org/10.1523/JNEUROSCI.5044-07.2008

Willott, J. F., & Urban, G. P. (1978). Response properties of neurons in nuclei of the mouse inferior colliculus. *Journal of Comparative Physiology*, *127*(2), 175–184. https://doi.org/10.1007/BF01352302

Witten, I. B., & Knudsen, E. I. (2005). Why Seeing Is Believing: Merging Auditory and Visual Worlds. *Neuron*, *48*(3), 489–496. https://doi.org/10.1016/j.neuron.2005.10.020

Woollard, H. H., & Harpman, J. A. (1940). The connexions of the inferior colliculus and of the dorsal nucleus of the lateral lemniscus. *Journal of Anatomy*, *74*(Pt 4), 441-458.3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1252612/

Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J., & McCarthy, G. (2003). Polysensory Interactions along Lateral Temporal Regions Evoked by Audiovisual Speech. *Cerebral Cortex*, *13*(10), 1034–1043. https://doi.org/10.1093/cercor/13.10.1034

Yamauchi, K., & Yamadori, T. (1982). Retinal Projection to the Inferior Colliculus in the Rat. *Cells Tissues Organs*, *114*(4), 355–360. https://doi.org/10.1159/000145608

Zhang, A. B. (1984). Retinotectal pathways in rodents: Particularly from the retinal ganglion cells to the inferior colliculus. *Taiwan Yi Xue Hui Za Zhi. Journal of the Formosan Medical Association*, *83*(1), 1–8.

Zwiers, M. P., Versnel, H., & Opstal, A. J. V. (2004). Involvement of Monkey Inferior Colliculus in Spatial Hearing. *Journal of Neuroscience*, *24*(17), 4145–4156. https://doi.org/10.1523/JNEUROSCI.0199-04.2004