AUDITORY MOTION DOES NOT MODULATE SPIKING ACTIVITY IN VISUAL MOTION PROCESSING AREAS MT AND MST

Abstract

The integration of multiple sensory modalities is one of the key aspects of brain function, allowing animals to take advantage of concurrent sources of information to make more accurate perceptual judgments. For many years, it was thought that multisensory integration in the cerebral cortex only occurs in high-level "polysensory" association areas, but recent studies have demonstrated cross-modal influences in regions that were traditionally designated as unimodal. In particular, several human neuroimaging studies have reported that extrastriate areas involved in visual motion perception are also activated by auditory motion, and may integrate audio-visual motion cues. However, the exact nature and extent of the effects of auditory motion on the visual cortex have not been studied at the single neuron level. We recorded the spiking activity of neurons in the middle temporal (MT) and medial superior temporal (MST) areas of anesthetized marmoset monkeys upon presentation of unimodal stimuli (moving auditory or visual patterns), as well as bimodal stimuli (concurrent audio-visual motion). Despite robust, direction selective responses to visual motion, none of the sampled neurons responded to auditory motion stimuli. Moreover, concurrent moving auditory stimuli had no significant effect on the ability of single MT and MST neurons, or populations of simultaneously recorded neurons, to discriminate the direction of motion of visual stimuli (moving random dot patterns with varying levels of motion noise). Our findings do not support the hypothesis that direct interactions between MT, MST and areas low in the hierarchy of auditory areas underlie audiovisual motion integration.

Significance Statement

Many studies have demonstrated that brain regions originally thought to be unisensory may play a role in multisensory processing. For example, some neuroimaging studies have found activity in regions involved in the processing of visual motion can be modified by auditory motion. We tested whether the spiking activity of neurons in two visual motion processing areas of the primate brain, areas MT and MST, can be modulated by moving auditory stimuli. Our results revealed that neurons in these areas neither respond to auditory motion, nor change their responses to visual motion according to auditory motion along the frontoparallel plane. These findings call into question the idea that audio-visual integration occurs at early stages of processing in the extrastriate cortex.

Introduction

The natural environment often produces stimuli that can be perceived by multiple senses, making multisensory integration one of the fundamental aspects of brain function (Stein and Stanford, 2008; Stein et al., 2014). There is evidence that humans and monkeys can integrate multisensory cues in a statistically optimal way (Ernst and Banks, 2002; Gu et al., 2008; Parise et al., 2012), giving a more reliable account than either sense alone.

Many studies of the neurological basis of multisensory integration have been shaped by the traditional model of sensory processing, in which each sensory modality is processed independently and only integrated in higher-level cortical areas (Felleman and Van Essen, 1991; Wallace et al., 2004) or in subcortical structures (Meredith and Stein 1983; Meredith et al. 1987; Reig and Silberberg 2014). However, studies have now shown that low level sensory cortical areas, historically considered unisensory, can be influenced by other modalities (for a review see Ghazanfar and Schroeder, 2006). In particular, these studies have shown cross-modal influences in the auditory and visual systems (awake primates: Schroeder and Foxe, 2002; Lakatos et al., 2007; Wang et al., 2008; awake mice: Iurilli et al., 2012; Ibrahim et al., 2016; Meijer et al., 2017; anesthetised mice Olcese et al., 2013; anesthestised ferrets: Bizley et al., 2007; Bizley and King, 2008, 2009).

Studies of visual motion perception have provided some of the most in-depth insights into the relationship between sensory neurons and behavior (Parker and Newsome, 1999), and motion stimuli have also proven to be a useful tool for understanding multisensory integration in low level visual areas (awake primates: Gu et al., 2008; Fetsch et al., 2013). Although it is well know that visual motion processing areas can integrate self-motion cues from the vestibular system in awake primates (Duffy, 1998; Gu et al., 2006), the integration of moving audio-visual cues has remained more controversial. Some human imaging studies have reported that the addition of a moving sound to a moving visual stimulus can modulate the responses in the human visual motion processing complex (hMT+, Alink et al., 2008; Lewis and Noppeney, 2010; von Saldern and Noppeney, 2013). Two imaging studies have also found evidence of responses to auditory motion alone in hMT+ (Poirier et al., 2005; Strnad et al., 2013 but see Jiang et al., 2014). The integration of audio-visual cues in hMT+ is an attractive proposition, given that psychophysical studies have shown that moving auditory cues can be integrated with visual stimuli to improve visual motion detection (Meyer and Wuerger, 2001; Kim et

al., 2012), that auditory motion can improve learning in visual motion tasks (Seitz et al., 2006), and that auditory stimuli can influence visual motion perception (Sekuler et al., 1997; Meyer and Wuerger, 2001; Kitagawa and Ichihara, 2002; Beer and Roder, 2004; Soto-Faraco et al., 2005; Freeman and Driver, 2008; Alink et al., 2012a; Stoner, 2012; Kafaligonul and Oluk, 2015).

In contrast, other studies suggest that integration of audio-visual motion cues may only occur in higher-level brain areas. Some psychophysical studies (Wuerger et al., 2003; Alais and Burr, 2004) have found that observer performance is consistent with probability summation, i.e. improved performance results from the fact that observers have two chances (i.e. visual or auditory) to answer correctly. This type of integration is likely to occur in a high-level brain region (Bizley et al., 2016), and human imaging studies often report activity in such regions (Lewis et al., 2000; Baumann and Greenlee, 2007; von Saldern and Noppeney, 2013).

To test if neurons in visual motion processing areas of the primate cerebral cortex integrate auditory motion cues without the influence of top-down pathways, we performed extracellular recordings in the middle temporal (MT) and medial superior temporal (MST) areas (Figure 1A), which together comprise the homolog of hMT+ (Zeki et al., 1991; Huk et al., 2002), in anaesthetized marmosets. In marmosets, these areas receive sparse connections from auditory cortex (Palmer and Rosa, 2006a, 2006b), and their human homologs have been implicated in the perception of moving auditory stimuli (Baumgart and Gaschler-Markefski, 1999; Pavani et al., 2002; Warren et al., 2002; Ducommun et al., 2004; Alink et al., 2012a). We tested two hypotheses: first, that neurons in MT and MST show responses to unimodal auditory stimuli, and second, that concurrent auditory and visual stimulation in the same direction of motion facilitates neuronal responses, in comparison with visual motion only. The second hypothesis was tested with levels of visual noise, as stimulus conditions that are more difficult to discriminate often reveal more robust multisensory integration effects (Meredith and Stein, 1983; Deneve et al., 2001; Ma et al., 2006; Gu et al., 2008; Fetsch et al., 2011).

Methods

Animals and surgical preparation

Single-unit and multi-unit extracellular recordings in areas MT and MST were obtained from 5 (2 male and 3 female, between 1.5 and 3 years of age, with no history of veterinary complications) marmoset monkeys. These animals were also used for unrelated anatomical tracing and visual physiology experiments. Experiments were conducted in accordance with the [Author Country] Code of Practice for the Care and Use of Animals for Scientific Purposes, and all procedures were approved by the [Author University] Animal Ethics Experimentation Committee.

Anesthesia was induced with alfaxalone (Alfaxan, 8 mg/kg), allowing a tracheotomy, vein cannulation and craniotomy to be performed. After all surgical procedures were completed, the animal was administered an intravenous infusion of pancuronium bromide (0.1 mg⁄kg⁄h) combined with sufentanil (6-8 μg⁄kg⁄h, adjusted to ensure no physiological responses to noxious stimuli) and dexamethasone (0.4 mg／kg／h), and was artificially ventilated with a gaseous mixture of nitrous oxide and oxygen (7:3). The electrocardiogram and level of cortical spontaneous activity were continuously monitored. Administration of atropine (1%) and phenylephrine hydrochloride (10%) eye drops was used to produce mydriasis and cycloplegia. Appropriate focus and protection of the corneas from desiccation were achieved by means of hard contact lenses selected by retinoscopy. This preparation has been used many times to record spiking activity in the visual cortex including area MT (e.g. Lui et al., 2007) and robust spiking responses in the auditory cortex (Rajan et al., 2013) including sustained responses and both “on” and “off” components of vocalizations, similar to those found in awake preparations, and sensitivity to interaural level differences (Lui et al. 2015).

Electrophysiology, data acquisition and pre-processing

We recorded neural activity with single shaft linear arrays (NeuroNexus) consisting of 32 electrodes separated by 50 μm. MT and MST recording sites were identified during experiments using anatomical landmarks, receptive field progression and size (Rosa and Elston, 1998), and direction selectivity, and were confirmed post-mortem by histological examination.

Electrophysiological data were recorded using a Cereplex system (Blackrock Microsystems) with a sampling rate of 30 kHz. For online analysis of spiking activity, each channel was high-pass filtered at 750 Hz and spikes were initially identified based on threshold crossings (-4 standard deviations of the root mean square). Threshold crossings were sorted for offline analysis using Offline Sorter (Plexon Inc.). Threshold crossings were classified as single-units if they showed good separation on the (2 component) principal component analysis plot, and were confirmed by inspection of the inter-spike interval histogram and consistency of waveform over time. Any remaining threshold crossings were classified as multi-unit activity. We excluded five single-units from adjacent channels since it was apparent they were duplicated across two channels, based on their sharp cross correlogram peak and high signal correlation (Bair et al., 2001). The median spontaneous firing rate (to a blank, black screen) of multi-units was 1.9 spikes/s, indicating that the multi-units consisted of relatively few single-units. For analysis of evoked potentials, each channel was low-pass filtered at 250Hz and then filtered with a 50Hz notch filter to remove line noise.

Visual stimuli

Visual stimuli were presented on a VIEWPixx3D monitor (1920 x 1080 pixels; 520 x 295 mm; 120 Hz refresh rate, VPixx Technologies) positioned 0.35 to 0.45 m from the animal on an angle to accommodate the size and eccentricity of the receptive field(s), typically subtending 70° in azimuth, and 40° in elevation. All stimuli were generated with MATLAB using Psychtoolbox-3 (Brainard, 1997).

The main visual stimulus consisted of random dots presented full screen. White dots (106 cd/m^2) of 0.2° in diameter were displayed on a black (0.25 cd/m^2) background (full contrast). The density was such that there were on average 0.5 dots per °^{2}; this was chosen because these parameters elicit good responses from marmoset MT when displayed on LCD monitors (Solomon et al., 2011; Zavitz et al., 2016). Dot coherence was controlled using the white noise method (i.e. Britten et al., 1992, 1996; see Pilly and Seitz 2009) by randomly choosing a subset of “noise” dots on each frame, which were displaced to random positions within the stimulus aperture. The remaining “signal” dots were moved in the same direction with a fixed displacement. We also presented an 8ms full screen flash stimulus at 1Hz (100 repeats) to test for visually evoked potentials.

Auditory stimuli

The ear canals were surgically exposed to allow the insertion of sound delivery tubes connected to speakers (MF1, Tucker-Davis Technologies). This method has the advantage of bypassing the outer ear, where certain frequencies are attenuated, ensuring that auditory stimuli will be delivered to the inner ear with greater precision and reliability. It has been used previously for evoking auditory spiking responses in the cortex of anaesthetized marmosets (Rajan et al., 2013; Lui et al., 2015) as well as cats (Rajan, 2000). The sound pressure level (SPL) and frequency response function of each speaker was calibrated using a Type 2673 microphone (Bruel & Kjaer) connected to a sound level calibrator type 4230 (94 dB–1000 Hz, Bruel & Kjaer), powered by a type 2804 microphone power supply. Experiments were conducted in a sound attenuating room.

The main auditory stimulus was a 6-12kHz band pass noise stimulus, chosen to match the common marmoset vocalization range (Agamaite et al., 2015), which was created with an 11th order Butterworth filter, presented at 70dB average binaural intensity (ABI). At this intensity and frequency range, auditory stimuli can be easily detected by humans and marmosets (Osmanski and Wang, 2011). Similar intensity levels have been used to investigate audiovisual integration in humans (Meyer and Wuerger, 2001; Kim et al., 2012) and induced BOLD responses hMT+ in fMRI studies (von Saldern and Noppeney, 2013). This stimulus was randomly generated for each trial. We used this type of stimulus as both a stationary (200 ms) and moving (500-1000 ms) sound to test for spiking responses to auditory stimuli. We varied the apparent spatial position of the static stimuli, and the direction of the moving stimuli, by manipulating the interaural level difference (ILD), the main cue for high frequency sounds in the azimuth in marmosets (Slee and Young, 2010). This approach allowed us to present moving auditory and visual stimuli concurrently, as a moving speaker would either block the visual stimulus on the monitor, or the monitor would cast an acoustic shadow on the speaker.

The stationary stimuli were presented at ILDs ranging from -25 dB to +25 dB in 5 dB increments. We used published data on the head related transfer function (Slee and Young, 2010) to guide the modulation of the ILD for the moving stimulus. For this frequency range, the ILD is approximately 5 dB at 15° azimuth, 10 dB at 30° azimuth and 15 dB at 90° azimuth. The moving auditory stimulus moved at a speed of 60°/s from the

midline (0 dB ILD) to 36° azimuth (~10dB ILD, Slee & Young 2010) and vice versa for the opposite direction of motion. Cells in the marmoset auditory cortex under the same preparation are sensitive to ILD differences as little as 5 dB for pure tones and broadband stimuli at a range of intensity levels (30-70 dB ABI), including levels used in the current experiment (Lui et al., 2015). We also presented auditory clicks (0.1 ms square wave, 70dB SPL) at 1Hz to test for auditory evoked potentials (100 repeats).

Audio-visual motion stimuli

Given that the auditory motion stimuli were produced by modulating the ILD, we could only simulate auditory motion in the horizontal plane. Thus we presented visual and auditory motion moving leftwards or rightwards (Figure 1B). For visual stimuli, we presented dots at coherences of 100, 82, 64, 46, 28, 10 and 0%. For auditory stimuli, there were no such coherence conditions, only leftwards or rightwards motion. In audio-visual conditions, the direction of auditory motion was the same as the visual motion, except in the case of 0% coherence visual motion, which we presented with both left and right auditory motion. All stimuli were presented for 600ms, based on previous psychophysical results (Kim et al., 2012), with 60 repeats per condition.

Auditory stimuli were synchronized to the visual stimulus using an AudioFile (Cambridge Research Systems). The AudioFile maintained a set of auditory stimuli as files on internal storage, and was triggered to play a given auditory stimulus when it received a digital signal. We configured the monitor to send a digital signal when the first frame of the visual stimulus appeared, and the AudioFile played the corresponding auditory stimulus within a few hundred microseconds, thereby producing highly synchronized audio-visual stimuli. The selection of the audio stimulus to be played was communicated from the stimulus generation PC to the AudioFile by a USB based DAQ device (USB1208FS, Measurement Computing) before the trial began.

Determination of receptive fields and basic direction tuning

Visual receptive fields were quantitatively mapped using a grid of either static flashed squares or small apertures of briefly presented moving dots. Visual stimuli were presented full screen, so as to cover as many neurons' receptive fields as possible. We also conducted visual direction tuning tests, which aided in

identifying the location of MT and MST. Direction selectivity was determined using a Rayleigh test ($p<0.05$) (Berens, 2009).

Data Analysis

Time windows and inclusion criteria: Mean firing rates were calculated using a time window starting 10 ms after stimulus onset (to avoid a potential noise artifact from the speakers) and finishing at the stimulus offset. Units were deemed responsive if they passed a Wilcoxon rank sum test ($p<0.01$) and if the firing rate was at least 2 spikes/s above the spontaneous rate. Units were considered left-right selective if the firing rate to the best direction of motion (left or right) at 100% coherence was significantly greater than that to the other direction (Wilcoxon rank sum test, $p<0.05$). Spike rate density functions were calculated using the Chronux software package (Mitra and Bokil, 2008, chronux.org), using a Gaussian kernel with a standard deviation of 25ms. For the population averages analysis, the spike rate density function of each unit was normalized by subtracting the spontaneous firing rate and dividing by the peak spike rate produced during the preferred direction of visual motion at 100% coherence (i.e. the maximum firing rate produced by the unit). These normalized spike rate density functions were then averaged to produce a population normalized spike rated density function for each stimulus type.

Evoked potentials: Evoked potentials were calculated by averaging the voltage across trials. We also calculated an average evoked potential for each penetration (across electrodes) to confirm the presence of auditory and visual evoked potentials.

Neurometric thresholds: We employed ideal observer analysis to test whether the addition of a moving auditory stimulus improved the ability of neurons to discriminate the direction of motion in a visual left-right discrimination task (Britten et al., 1992). For each level of coherence, we calculated the area under the Receiver Operator Characteristic (aROC) curve from the distributions of responses to the preferred and null directions, the former of which was determined by the direction, either left or right, which elicited the best response at 100% coherence. The aROC values were fitted using least squares regression with two variants of the Weibull function, resulting in a neurometric curve that described the neuron's performance with respect to coherence (an aROC plot of an example neuron is shown in Figure 5A):

$$p = 1 - 0.5 \exp\left[-(c/\alpha)^{\beta}\right] \quad (1)$$

$$p = \delta - (\delta - 0.5) \exp\left[-(c/\alpha)^{\beta}\right] \quad (2)$$

where p was the probability of correctly discriminating the direction of motion at coherence c, α was the coherence of threshold performance (p=0.82, convention established by Britten et al., 1992), β controlled the slope and δ was the asymptotic level of performance (less than 1). As Equation 2 has an extra free parameter, we used an F-test to decide whether to reject the use of Equation 2 over Equation 1. The α was limited to between 0 and 3, β was limited to lie between 0 and 10, and δ was limited to lie between 0 and 1. Units that did not have an aROC of at least 0.82 at 100% coherence could not have a threshold (i.e. $p(c=100)<0.82$), were excluded from analyses of thresholds, as was any neuron whose threshold exceeded 100% (given that curving fitting does not guarantee the function will fit all data points perfectly). Statistical testing for differences in the visual and the audio-visual condition was performed with a permutation test (1000 iterations). Each iteration, the visual and audio-visual spike rates were randomly shuffled, and the aROCs and thresholds were recalculated. A distribution of differences in the visual and audio-visual thresholds was constructed from these iterations. A unit was considered to have a statistically significant difference in visual and audio-visual thresholds if the 95% interval of this distribution did not overlap with zero.

Neurometric performance: For units that did not meet the above criteria but were still left-right selective, we calculated a "neurometric performance", which is the integral of the fitted Weibull function from 0 to 100% coherence. This produces a value ranging from 0.5 (uninformative for all coherences) to 1 (perfectly informative across all coherences). Statistical significance testing was performed in the same way as for the neurometric thresholds.

Population decoding: For each penetration with at least two left-right selective units, we trained a classifier using Linear Discriminant Analysis to decode the direction of motion at each coherence. We estimated the accuracy and variability of the classifier by training on a randomly sampled subset of 80% (48/60) of trials and testing on the remainder, and repeating this process 1000 times. For each iteration, the population neurometric threshold was calculated in the same way as for the individual units, and the final threshold was calculated as

the average threshold across iterations. Differences in visual and audio-visual thresholds were deemed statistically significant if the confidence interval of the distribution of threshold differences did not overlap zero.

Histology

At the end of the recordings, the animals were given an intravenous overdose of sodium pentobarbitone and, following cardiac arrest, were perfused with 0.9% saline, followed by 4% paraformaldehyde in 0.1 M phosphate buffer pH, 7.4. The brain was post-fixed for approximately 24 hours in the same solution, and then cryoprotected with fixative solutions containing 10%, 20%, and 30% sucrose. The brains were then frozen and sectioned into 40 µm coronal slices. Alternate series were stained for Nissl substance and myelin (Gallyas, 1979). The location of recording sites was reconstructed by identifying electrode tracks and depth readings recorded during the experiment. Additionally, each electrode array was coated in DiI, allowing visualization under fluorescence microscopy prior to staining of the sections (Figure 1C). In coronal sections, MT is clearly identifiably by heavy myelination in the granular and infragranular layers (Figure 1C) (Rosa and Elston, 1998), whereas MST is more lightly myelinated and lacks clear separation between layers (Palmer and Rosa, 2006a). The majority of units reported here were histologically confirmed to be in MT or MST, but for some penetrations in which the histology was unclear (12% of units), units were included on the basis of their receptive field size and progression, and their direction tuning.

Results

Sample size

We made 27 electrode array penetrations in areas MT and MST (MT: n=18; MST: n=7; MT/MST: 2), and recorded 314 visually responsive units (MT: n=223; MST: n=91), of which 11% were classified as single-units. We did not find any significant difference between single and multi-units in the following analyses, so they were grouped together. In agreement with previous reports (Zeki, 1974; Maunsell and Van Essen, 1983; Albright, 1984; Celebrini and Newsome, 1994; Born and Bradley, 2005; Lui and Rosa, 2015), we found that most units showed visual direction selectivity (MT 76% MST 60%, Rayleigh test, $p<0.05$). To test for audio-visual integration, we only presented leftwards or rightwards motion, but did not restrict our analyses to units which preferred one of these directions. We found that 238 (76%) of the units we recorded showed significantly different responses to visual leftwards versus rightwards motion (Wilcoxon rank sum test, $p<0.05$; MT: n=180; MST: n=58).

Auditory stimuli do not elicit spiking responses in MT/MST

In all penetrations, we presented moving visual stimuli as well as moving and stationary auditory stimuli. Although we recorded a large number of visually responsive units, we did not observe any spiking responses to any of the auditory stimuli. Figure 2 shows the spiking activity of 4 example units that were selective for visual motion on the left-right axis, and one unit which was not left-right selective. Visual and audiovisual stimuli evoked strong spiking responses over time for at least one direction of motion at 100% coherence (blue and red lines, bottom and middle rows of the rasters). In contrast, moving and stationary auditory stimuli did not evoke responses at all (yellow lines, top row of rasters).

Figure 3 shows the distributions of the mean firing rates of the best moving visual and auditory stimuli (panels A and B, respectively) in 3 animals in which we interleaved visual and auditory motion trials (n=203 units). None of the visually responsive units showed statistically significant responses to auditory stimuli (determined by a Wilcoxon Rank Sum test, $p<0.01$). To confirm that the auditory stimulus which we used could evoke spiking activity elsewhere in the brain, we recorded from auditory areas adjacent to MST as a positive control (Figure 1A, dark green region) and found units that produced clear responses to the moving auditory stimulus

(e.g. Figure 3C). There were also small but discernable evoked potentials in response to auditory click stimuli, even in MT and MST (Figure 3D), confirming that auditory stimulation evoked neuronal activity in all animals. However, the MT/MST auditory evoked potential was small in comparison to the visual (Figure 3D, red vs blue trace). The auditory cortex is within 4 mm of our recording sites in areas MT and MST in the marmoset (Figure 1A), so it is likely that this auditory evoked potential was the result of activity in distal auditory brain regions (most likely caudal auditory areas: Palmer and Rosa, 2006a; Kajikawa and Schroeder, 2011) rather than being generated by neurons in MT/MST.

Moving auditory stimuli do not modulate firing rates in MT/MST

As moving and static auditory stimuli did not elicit spiking activity, we investigated if the addition of a moving auditory stimulus modulated the firing rate of neurons to a moving visual stimulus in the same direction. We tested this at different levels of motion coherence, given that multisensory integration is more likely to occur when the unimodal signal strength (or neural response) is weak (Meredith and Stein, 1983; Meredith et al., 1987; Deneve et al., 2001; Ma et al., 2006; Gu et al., 2008; Fetsch et al., 2011). Neurons in areas MT and MST are known to be direction selective (Dubner and Zeki, 1971; Felleman and Kaas, 1984; Celebrini and Newsome, 1994), but as we only presented motion along the horizontal axis, we observed a range of responses to changes in coherence (Figure 2). Some units showed a response pattern consistent with a clear preference for one direction (e.g. Figure 2A-C) (Britten et al., 1992), some had only a moderate degree of left-right selectivity (e.g. Figure 2D), and others showed no preference for leftwards or rightwards motion (e.g. Figure 2E). The majority of the last grouping (68%) were direction selective for visual motion but preferred directions of motion along the vertical axis.

To analyze the full population, each unit was designated as right-preferring or left-preferring based on which direction elicited the highest mean spike rate at 100% coherence, with the preferred direction being designated as positive 100% coherence, and the null direction as negative 100% coherence. To visualize the population response, the responses of each unit were first normalized to lie between 0 and 1, so that 0 is the spontaneous rate and 1 is the response to the preferred direction, and then averaged to produce a population-average response to motion in the preferred and null directions, at different levels of coherence, for visual-only and

audio-visual conditions (Figure 4). A 2-way repeated measures ANOVA did not find a significant effect of modality (visual or AV; p=0.983), only coherence (p<0.001), on firing rates, and did not find a significant interaction effect between modality and coherence (p=0.99). We also found that there was no difference in firing rates due to modality when units from MT and MST were analyzed separately (2-way ANOVA, p=0.998 and p=0.956 respectively for main effect of modality). To test if auditory stimuli had any effect on the 0% coherence stimulus, we used a 1-way ANOVA to test if there was any difference in firing rate in the visual-only condition compared to the audio-visual conditions with auditory motion in the preferred or null visual direction, but again found no effect (p=0.660). In summary, there was no evidence to suggest that auditory motion modulates the spike rates of MT/MST neurons, and this was consistent for all levels of visual motion noise tested.

Moving auditory stimuli do not improve visual neurometric thresholds

The analysis in the previous section included all visually responsive units, even those that do not carry information regarding the direction of motion (e.g. Figure 2E). Working on the hypothesis that auditory stimuli may enhance motion discriminability or sensitivity, we analyzed just the responses of units that show strong left-right selectivity (e.g. Figure 2A-C). We used aROC analysis to determine how well an ideal observer could determine the direction of motion (left or right) from the firing rate of a neuron (Newsome et al., 1989; Britten et al., 1992) at different levels of motion coherence, to determine whether the addition of the auditory stimulus resulted in improved neurometric performance compared to the visual only stimulus. We calculated the aROC at each coherence and fitted a Weibull function (e.g. Figure 5A) to determine the neurometric threshold - the coherence level at which the aROC reaches 0.82 (by convention, Britten et al., 1992, see inset of Figure 5A). Of the 238 left-right direction selective units recorded, 135 (57%; MT: n=103; MST: n=32) reached a maximum aROC of at least 0.82, and thus had a defined neurometric threshold. The mean change in neurometric threshold for the audio-visual versus visual conditions was not statistically significant different from zero (Figure 5B, 0.3% coherence impairment for the audio-visual condition, paired t-test p=0.88). We did not find any difference when units from MT and MST were analyzed separately (p=0.83 and p=0.94 respectively, paired t-tests). Only 3 units (2.2%, permutation test p<0.05) showed statically significant differences (Figure 5B, black bars); which were most likely false positives.

Of the left-right selective units, 77 (24%) did not meet the criteria for threshold analyses. To investigate the contribution of all left-right selective units, we measured the neurometric performance as the area under the fitted Weibull curve, as this does not require that a unit achieves any particular threshold aROC. Larger areas correspond to better discrimination performance across coherences (e.g. Figure 5A vs Figure 5C). In line with the neurometric threshold analyses, we found that the mean change in neurometric performance for the audio-visual versus visual condition was not significantly different from zero (Figure 5D, 0.002 unit impairment for the audio-visual condition, paired t-test p=0.86). Again, there was no difference when units from MT and MST were analyzed separately (p=0.75 and p=0.9 respectively, paired t-tests). Only 11 (4.6%) of the units showed a statistically significant difference between the visual and audio-visual conditions (permutation test, Figure 5D, black bars), and the effects were balanced in terms of whether the audio-visual stimulation improved or decreased performance.

Trial to trial correlations between neurons can have profound effects on population coding (Bair et al., 2001; Cohen and Newsome, 2009; Quian Quiroga and Panzeri, 2009; Ruff and Cohen, 2016; Panzeri et al., 2017) and so we tested if auditory stimulation affects the encoding of stimuli at the population level. To investigate this, we trained a classifier to decode the direction of motion (left or right) for each of the 25 array penetrations that contained at least 2 left-right selective units. Of these, 21 penetrations had defined population neurometric thresholds (analogous to the individual unit neurometric thresholds - performance of at least 82% at 100% coherence). Figure 5E shows the performance of one such penetration, containing 25 units from area MT. We measured the neurometric performance of each penetration under visual and audio-visual conditions (Figure 5F), but found none that showed significant differnces in visual and audio-visual thresholds. Moreover, the mean difference in visual and audio-visual thresholds across penetrations was not significantly different from zero (Figure 5F; 0.11% coherence improvement in the audio-visual condition, p=0.98, paired t-test). In summary, the addition of the moving auditory stimli in a congruent direction to random dot visual stimuli did not increase or decrease the amount of information carried by single neurons or populations of simultaneously recorded neurons.

Temporal response profile for visual, auditory and audiovisual stimuli

To investigate whether there were any modulations of spiking activity in specific time ranges that may not have been detected in the previous analysis (which counted spikes over the full stimulus duration), we examined the temporal profile of spiking activity in response to visual, auditory and audio-visual stimuli. To test whether the addition of auditory stimulus enhanced (or diminshed) directional information contained in spiking neurons, we analysised the the subset of units (n=135) that had defined neurometric thresholds and were therefore strongly left-right selective (Figure 5A,B), we calculated normalized spike rate density functions for each direction and level of motion coherence (Figure 6). Confirming the findings from Figure 3, the mean firing rate in response to auditory stimuli was not significantly different to to zero for all time bins during the stimulus presentation (25ms time bins, t-test, $p < 0.01$), demonstrating there was no brief or transient response to the auditory stimulus at any stage of the stimulus presentation. In agreement with the analysis in Figures 4 and 5, the mean firing rate of the visual and audiovisual conditions was not significantly different at any time bin during the stimulus presentation (25ms time bins, t-test, $p < 0.01$). Therefore auditory responses and audio-visual modulations were not present at any time point in the population of the strongly left-right selective units.

Discussion

We investigated the extent to which neurons in the visual motion processing areas MT and MST respond to auditory motion stimuli, and whether congruent auditory stimuli modulate the activity of these neurons in conjunction with visual motion. Using an anaesthetized preparation in which both auditory and visual responses can be robustly elicited, we found no evidence of cross-modal interactions in spiking activity in either MT or MST neurons. Neither stationary nor moving auditory stimuli elicited spiking when presented alone, and moving auditory stimuli did not modulate firing rates when paired with visual stimuli. We investigated this using methods for characterizing the information carried by both individual neurons (Britten et al., 1992; Gu et al., 2008) and for populations of neurons. We found no evidence of multi-sensory integration at either near-threshold or supra-threshold levels of stimulus noise (Meredith and Stein, 1983; Deneve et al., 2001; Ma et al., 2006; Gu et al., 2008; Fetsch et al., 2011).

Experimental conditions affecting the lack of audio-visual integration.

As our study was conducted using an anesthetized preparation, it could be argued that we did not observe cross-modal effects because they are only present in the conscious state, or when the animal is performing an audio-visual motion discrimination task. Although we cannot rule out this explanation, it should be noted that multisensory responses have been widely reported under anesthesia in both high and low level areas of the cortex (primates: Bruce et al., 1981; Hikosaka et al., 1988; cats: Wallace et al., 1992, 2004; ferrets: Bizley et al., 2007), as well as in subcortical structures (cats: Meredith and Stein, 1983; mice: Reig and Silberberg, 2014). Therefore multisensory integration can be supported, at least in part, by bottom-up processes which are still present in the anesthetized state (Alkire et al., 2008). Our results indicate that the integration of audio-visual motion cues is not likely to be supported by such processes in areas MT and MST, although the possibility remains that they may be evident following feedback connections from high-order polysensory association areas such as the superior temporal polysensory cortex (area TPO; Baylis et al., 1987).

It is also possible that audio-visual integration would have been evident if different types of stimuli were used. However, MT neurons respond well to many types of visual motion (bars: Maunsell and Van Essen, 1983; Albright, 1984; gratings: Movshon et al., 1985; dot patterns: Albright, 1984; Britten et al., 1992; Solomon et

al., 2015); it is therefore unlikely that MT cells would require very specific auditory stimuli to activate, or integrate, which were not covered within the frequency and ILD range of our broadband stimuli. Our experiments required the use of headphones to simulate auditory motion by manipulating the ILD, but it could be argued that demonstrating multisensory responses in MT and MST necessitates other auditory motion cues, such as interaural time differences and pinnae-based spectral cues. However, similar set-ups using headphones have been used in human studies that have shown audio-visual integration for motion (Kayser et al., 2017), and ILDs have been shown to be the dominant cue for sound localization in the azimuth for marmosets for high frequency sounds (Slee and Young, 2010). Yet, we observed no auditory responses (Figure 2) or modulations (Figures 3 and 4) when manipulating the ILD. In summary, the auditory stimuli used in the present study contained strong cues for sound location and motion that would be expected to elicit at least observable changes in neural activity.

Comparison to human studies

We did not observe any spiking responses to moving auditory stimuli in MT and MST neurons, in agreement with several fMRI studies in humans (Bedny et al., 2010; Alink et al., 2012b; Jiang et al., 2014). Two other studies have reported activation in hMT+ in response to moving auditory stimuli (Poirier et al., 2005; Strnad et al., 2013), but it has been suggested that this may be explained by the methodology that was used to localize hMT+, and that the auditory-related effects may arise from a cortical area outside hMT+ (Jiang et al., 2014). Our observations support this point of view.

Other fMRI studies have shown that moving auditory stimuli can modulate hMT+ activity when paired with visual stimuli (Lewis and Noppeney, 2010; Strnad et al., 2013; von Saldern and Noppeney, 2013). One key difference between these studies and ours is that the human participants are conscious and often performing a task. Therefore the audio-visual related activity may not correspond to audio-visual cue integration itself, but may instead reflect task-related activity (Alink et al., 2012b; Bizley et al., 2016; Kayser et al., 2017), such as the binding of the two modalities to form a unified percept (Nahorna et al., 2012, 2015; Bizley and Cohen, 2013), attentional effects (Beer and Roder, 2004; Beer and Röder, 2005; Lakatos et al., 2008), or choice-related signals from the decision making process (Cumming and Nienborg, 2016). Indeed, it is well known

that top-down attention can affect spiking activity in MT and MST (Treue and Maunsell, 1996) and this extends to feature based attention, which is selective for direction of motion (Treue and Martinez-Trujillo, 1999).

Another key difference is that the hemodynamic response observed in fMRI, as well as the evoked potentials observed in EEG, are likely to reflect synaptic (input) activity rather than spiking (output) activity (Logothetis et al., 2001; Buzsáki et al., 2012). We also observed auditory evoked potentials in MT and MST to auditory click stimuli (Figure 2D). Intracranial auditory evoked potentials have been reported in the visual cortex of mice (awake: Iurilli et al., 2012; Ibrahim et al., 2016; anesthetised: Olcese et al., 2013), and recently in humans (Brang et al., 2015). Thus, it may be possible that areas MT and MST do receive auditory information, as suggested by studies which demonstrate that they receive sparse inputs from the auditory cortex (Palmer and Rosa, 2006a, 2006b). However, we could not rule out the possibility that these evoked potentials we observed in MT and MST originated from nearby auditory cortex, which is only a few millimeters away from our recording sites (Palmer and Rosa, 2006a; Kajikawa and Schroeder, 2011). As we did not find any evidence of auditory modulation of visually evoked spiking in response to motion, it is unlikely that that any auditory inputs to MT or MST lead directly to audio-visual motion cue integration.

Implications for multisensory integration in the cerebral cortex

Our results suggest that the integration of audio-visual motion cues in MT and MST is not likely to be an automatic, bottom-up process, unlike the clear multisensory responses observed in other structures, such as the superior colliculus, even when anesthetized (Meredith and Stein, 1983). If neural activity in these areas does reflect audio-visual integration in the awake behaving state, it may be the result of feedback from higher-level areas, potentially from the superior temporal cortex (Lewis et al., 2000; Baumann and Greenlee, 2007; von Saldern and Noppeney, 2013), or it may develop overtime through multi-sensory perceptual learning (e.g. Seitz et al., 2006). Overall, our results favor the traditional model of cortical multisensory processing, with separate cortical domains for each modality, and multisensory neurons being restricted to transition zones between these domains (Wallace et al., 2004). Areas MT and MST have been implicated in the multisensory processing of several different modalities, and it is well established that area MST (but not MT) responds to

and integrates vestibular motion cues, with studies in awake primates showing clear evidence of spiking responses (Duffy, 1998; Gu et al., 2006) and multisensory integration (Gu et al., 2008; Fetsch et al., 2011). Other studies have indicated that hMT+ is involved in processing tactile stimuli (Hagen et al., 2002; Beauchamp et al., 2007; Basso et al., 2012; Pei and Bensmaia, 2014), but much like auditory processing (Jiang et al., 2014), this finding remains controversial (Jiang et al., 2015).

Most of the physiological evidence for auditory influences on spiking activity in visual cortex have come from studies in mice (awake: Iurilli et al., 2012; Ibrahim et al., 2016; Meijer et al., 2017; anesthetised: Olcese et al., 2013). To our knowledge there is only one study which has shown that sounds can modulate the spiking activity in the visual cortex of primates (Wang et al., 2008), producing a modest decrease in response latency in awake animals. Despite the lack of physiological data, it was been well established that there are direct anatomical connections between the visual and auditory cortex in primates (Falchier et al., 2002; Rockland and Ojima, 2003; Cappe and Barone, 2005; Palmer and Rosa, 2006a, 2006b). However, cross-modal connections do not guarantee the presence of multisensory neurons (Allman et al., 2008), so their role may be to facilitate more subtle task related modulations rather than multisensory cue integration.

References

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

Alais D, Burr DC (2004) No direction-specific bimodal facilitation for audiovisual motion detection. Cogn Brain Res 19:185–194.

Albright TD (1984) Direction and orientation selectivity of neurons in visual area MT of the macaque. J Neurophysiol 52:1106–1130.

Alink A, Euler F, Galeano E, Krugliak A, Singer W, Kohler A (2012a) Auditory motion capturing ambiguous visual motion. Front Psychol 3:1–8.

Alink A, Euler F, Kriegeskorte N, Singer W, Kohler A (2012b) Auditory motion direction encoding in auditory cortex and high-level visual cortex. Hum Brain Mapp 33:969–978.

Alink A, Singer W, Muckli L (2008) Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. J Neurosci 28:2690–7.

Alkire MT, Hudetz AG, Tononi G (2008) Consciousness and anesthesia. Science (80-) 322:876–880.

Allman BL, Bittencourt-Navarrete RE, Keniston LP, Medina AE, Wang MY, Meredith MA (2008) Do cross-modal projections always result in multisensory integration? Cereb Cortex 18:2066–2076.

Bair W, Zohary E, Newsome WT (2001) Correlated firing in macaque visual area MT: time scales and relationship to behavior. J Neurosci 21:1676–1697.

Basso D, Pavan A, Ricciardi E, Fagioli S, Vecchi T, Miniussi C, Pietrini P (2012) Touching motion: RTMS on the human middle temporal complex interferes with tactile speed perception. Brain Topogr 25:389–398.

Baumann O, Greenlee MW (2007) Neural correlates of coherent audiovisual motion perception. Cereb Cortex 17:1433–43.

Baumgart F, Gaschler-Markefski B (1999) A movement-sensitive area in auditory cortex. Nature 400:1997–1999.

Baylis GC, Rolls ET, Leonard CM (1987) Functional subdivisions of the temporal lobe neocortex. J Neurosci 7:330–342.

Beauchamp MS, Yasar NE, Kishan N, Ro T (2007) Human MST but not MT responds to tactile stimulation. J Neurosci 27:8261–8267.

Bedny M, Konkle T, Pelphrey K, Saxe R, Pascual-Leone A (2010) Sensitive period for a multimodal response

in human visual motion area MT/MST. Curr Biol 20:1900–1906.

Beer AL, Roder B (2004) Unimodal and crossmodal effects of endogenous attention to visual and auditory motion. Cogn Affect 4:230–240.

Beer AL, Röder B (2005) Attending to visual or auditory motion affects perception within and across modalities: An event-related potential study. Eur J Neurosci 21:1116–1130.

Berens P (2009) CircStat□: A MATLAB toolbox for circular statistics. J Stat Softw 31:1–21.

Bizley JK, Cohen YE (2013) The what, where and how of auditory-object perception. Nat Rev Neurosci 14:693–707.

Bizley JK, Jones GP, Town SM (2016) Where are multisensory signals combined for perceptual decision-making? Curr Opin Neurobiol 40:31–37.

Bizley JK, King AJ (2008) Visual-auditory spatial processing in auditory cortical neurons. Brain Res 1242:24–36.

Bizley JK, King AJ (2009) Visual influences on ferret auditory cortex. Hear Res 258:55–63.

Bizley JK, Nodal FR, Bajo VM, Nelken I, King AJ (2007) Physiological and anatomical evidence for multisensory interactions in auditory cortex. Cereb Cortex 17:2172–89.

Born RT, Bradley DC (2005) Structure and function of visual area MT. Annu Rev Neurosci 28:157–189.

Brainard DH (1997) The psychophysics toolbox. Spat Vis 10:433–6.

Brang D et al. (2015) Peripheral sounds rapidly activate visual cortex□: evidence from electrocorticography. J Neurophysiol 114:3023–3028.

Britten KH, Shadlen MN, Newsome WT, Movshon JA (1992) The analysis of visual motion: a comparison of neuronal and psychophysical performance. J Neurosci 12:4745–4765.

Bruce C, Desimone R, Gross CG (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. J Neurophysiol 46:369–384.

Buzsáki G, Anastassiou CA, Koch C (2012) The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. Nat Rev Neurosci 13:407–420.

Cappe C, Barone P (2005) Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. Eur J Neurosci 22:2886–2902.

Celebrini S, Newsome WT (1994) Neuronal and psychophysical sensitivity to motion signals in extrastriate area MST of the macaque monkey. J Neurosci 14:4109–4124.

Cohen MR, Newsome WT (2009) Estimates of the contribution of single neurons to perception depend on timescale and noise correlation. J Neurosci 29:6635–48.

Cumming BG, Nienborg H (2016) Feedforward and feedback sources of choice probability in neural population responses. Curr Opin Neurobiol 37:126–132.

Deneve S, Latham PE, Pouget A (2001) Efficient computation and cue integration with noisy population codes. Nat Neurosci 4:826–831.

Dubner R, Zeki SM (1971) Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. Brain Res 35:528–32.

Ducommun C, Michel C, Clarke S (2004) Cortical motion deafness. Neuron 43:765–777.

Duffy CJ (1998) MST neurons respond to optic flow and translational movement. J Neurophysiol 80:1816–1827.

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429–433.

Falchier A, Clavagnier S, Barone P, Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci 22:5749–5759.

Felleman DJ, Kaas JH (1984) Receptive-field properties of neurons in middle temporal visual area (MT) of owl monkeys. J Neurophysiol 52:488–513.

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1–47.

Fetsch CR, DeAngelis GC, Angelaki DE (2013) Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. Nat Rev Neurosci 14:429–442.

Fetsch CR, Pouget A, DeAngelis GC, Angelaki DE (2011) Neural correlates of reliability-based cue weighting during multisensory integration. Nat Neurosci 15:146–154.

Freeman E, Driver J (2008) Direction of visual apparent motion driven solely by timing of a static sound. Curr Biol 18:1262–6.

Gallyas F (1979) Silver staining of myelin by means of physical development. Neurol Res 1:203–209.

Ghazanfar A a., Schroeder CE (2006) Is neocortex essentially multisensory? Trends Cogn Sci 10:278–285.

Gu Y, Angelaki DE, DeAngelis GC (2008) Neural correlates of multisensory cue integration in macaque MSTd. Nat Neurosci 11:1201–1210.

Gu Y, Watkins P V, Angelaki DE, DeAngelis GC (2006) Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. J Neurosci 26:73–85.

Hagen MC, Franzén O, McGlone F, Essick G, Dancer C, Pardo J V. (2002) Tactile motion activates the human middle temporal/V5 (MT/V5) complex. Eur J Neurosci 16:957–964.

Hikosaka K, Iwai E, Saito H a, Tanaka K (1988) Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. J Neurophysiol 60:1615–1637.

Huk AC, Dougherty RF, Heeger DJ (2002) Retinotopy and functional subdivision of human areas MT and MST. J Neurosci 22:7195–7205.

Ibrahim LA, Mesik L, Ji X, Zingg B, Zhang LI, Ibrahim LA, Mesik L, Ji X, Fang Q, Li H, Li Y, Zingg B, Zhang LI (2016) Cross-modality sharpening of visual cortical processing through layer-1-mediated inhibition and disinhibition. Neuron 89:1–15.

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 73:814–28.

Jiang F, Beauchamp MS, Fine I (2015) Re-examining overlap between tactile and visual motion responses within hMT+ and STS. Neuroimage 119:187–196.

Jiang F, Stecker GC, Fine I (2014) Auditory motion processing after early blindness. J Vis 14:1–18.

Kafaligonul H, Oluk C (2015) Audiovisual associations alter the perception of low-level visual motion. Front Integr Neurosci 9:1–11.

Kajikawa Y, Schroeder CE (2011) How local is the local field potential? Neuron 72:847–858.

Kayser SJ, Philiastides MG, Kayser C (2017) Sounds facilitate visual motion discrimination via the enhancement of late occipital visual representations. Neuroimage 148:31–41.

Kim R, Peters M a K, Shams L (2012) 0 + 1 > 1: How adding noninformative sound improves performance on a visual task. Psychol Sci 23:6–12.

Kitagawa N, Ichihara S (2002) Hearing visual motion in depth. Nature 416:172–174.

Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron 53:279–292.

Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. Science (80-) 320:110–113.

Lewis JW, Beauchamp MS, DeYoe EA (2000) A comparison of visual and auditory motion processing in

human cerebral cortex. Cereb Cortex 10:873–888.

Lewis R, Noppeney U (2010) Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. J Neurosci 30:12329–39.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. Nature 412:150–157.

Lui LL, Bourne JA, Rosa MGP (2007) Spatial and temporal frequency selectivity of neurons in the middle temporal visual area of new world monkeys (Callithrix jacchus). Eur J Neurosci 25:1780–92.

Lui LL, Mokri Y, Reser DH, Rosa MGP, Rajan R (2015) Responses of neurons in the marmoset primary auditory cortex to interaural level differences: comparison of pure tones and vocalizations. Front Neurosci 9.

Lui LL, Rosa MGP (2015) Structure and function of the middle temporal visual area (MT) in the marmoset: Comparisons with the macaque monkey. Neurosci Res 93:62–71.

Ma WJ, Beck JM, Latham PE, Pouget A (2006) Bayesian inference with probabilistic population codes. Nat Neurosci 9:1432–1438.

Maunsell JHR, Van Essen DC (1983) Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. J Neurophysiol 49:1127–1147.

Meijer GT, Montijn JS, Pennartz CMA, Lansink CS (2017) Audiovisual modulation in mouse primary visual cortex depends on cross-modal stimulus configuration and congruency. J Neurosci 37:8783–8796.

Meredith MA, Nemitz J, Stein BE (1987) Determinants of multisensory integration neurons in superior colliculus. I . Temporal factors. J Neurosci 7:3215–3229.

Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. Science (80-) 221:389–391.

Meyer GF, Wuerger SM (2001) Cross-modal integration of auditory and visual motion signals. Neuroreport 12:2557–2560.

Mitra P, Bokil H (2008) Observed brain dynamics. Oxford University Press.

Movshon JA, Adelson EH, Gizzi MS, Newsome WT (1985) The analysis of moving visual patterns In: Pontificiae Academiae Scientiarum Scripta Varia , pp117–151.

Nahorna O, Berthommier F, Schwartz J-L (2012) Binding and unbinding the auditory and visual streams in the

McGurk effect. J Acoust Soc Am 132:1061–1077.

Nahorna O, Berthommier F, Schwartz J-L (2015) Audio-visual speech scene analysis: Characterization of the dynamics of unbinding and rebinding the McGurk effect. J Acoust Soc Am 137:362–377.

Newsome WT, Britten KH, Movshon JA (1989) Neuronal correlates of a perceptual decision. Nature 341:52–54.

Olcese U, Iurilli G, Medini P (2013) Cellular and synaptic architecture of multisensory integration in the mouse neocortex. Neuron 1:1–15.

Osmanski MS, Wang X (2011) Measurement of absolute auditory thresholds in the common marmoset (Callithrix jacchus). Hear Res 277:127–133.

Palmer SM, Rosa MGP (2006a) A distinct anatomical network of cortical areas for analysis of motion in far peripheral vision. Eur J Neurosci 24:2389–2405.

Palmer SM, Rosa MGP (2006b) Quantitative analysis of the corticocortical projections to the middle temporal area in the marmoset monkey: evolutionary and functional implications. Cereb Cortex 16:1361–75.

Panzeri S, Harvey CD, Piasini E, Latham PE, Fellin T (2017) Cracking the neural code for sensory perception by combining statistics, intervention and behavior. Neuron 93:1–30.

Parise C V., Spence C, Ernst MO (2012) When correlation implies causation in multisensory integration. Curr Biol 22:46–9.

Pavani F, Macaluso E, Warren J, Driver J, Griffiths T (2002) A common cortical substrate activated by horizontal and vertical sound movement in the human brain. Curr Biol 12:1584–1590.

Pei Y-C, Bensmaia SJ (2014) The neural basis of tactile motion perception. J Neurophysiol 112:3023–3032.

Poirier C, Collignon O, DeVolder AG, Renier L, Vanlierde A, Tranduy D, Scheiber C (2005) Specific activation of the V5 brain area by auditory motion processing: An fMRI study. Cogn Brain Res 25:650–658.

Quian Quiroga R, Panzeri S (2009) Extracting information from neuronal populations: information theory and decoding approaches. Nat Rev Neurosci 10:173–185.

Rajan R (2000) Centrifugal pathways protect hearing sensitivity at the cochlea in noisy environments that exacerbate the damage induced by loud sound. J Neurosci 20:6684–6693.

Rajan R, Dubaj V, Reser DH, Rosa MGP (2013) Auditory cortex of the marmoset monkey - complex responses to tones and vocalizations under opiate anaesthesia in core and belt areas. Eur J Neurosci

37:924–941.

Reig R, Silberberg G (2014) Multisensory integration in the mouse striatum. Neuron 83:1200–1212.

Rockland KS, Ojima H (2003) Multisensory convergence in calcarine visual areas in macaque monkey. Int J Psychophysiol 50:19–26.

Rosa MGP, Elston GN (1998) Visuotopic organisation and neuronal response selectivity for direction of motion in visual areas of the caudal temporal lobe of the marmoset monkey (Callithrix jacchus): middle temporal area, middle temporal crescent, and surrounding cortex. J Comp Neurol 393:505–527.

Ruff DA, Cohen MR (2016) Stimulus dependence of correlated variability across cortical areas 36:7546–7556.

Schroeder CE, Foxe JJ (2002) The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Brain Res Cogn Brain Res 14:187–198.

Seitz AR, Kim R, Shams L (2006) Sound facilitates visual learning. Curr Biol 16:1422–1427.

Sekuler R, Sekuler AB, Lau R (1997) Sound alters visual motion perception. Nature 385:308–308.

Slee SJ, Young ED (2010) Sound localization cues in the marmoset monkey. Hear Res 260:96–108.

Solomon SS, Chen SC, Morley JW, Solomon SG (2015) Local and global correlations between neurons in the middle temporal area of primate visual cortex. Cereb Cortex 25:3182–3196.

Solomon SS, Tailby C, Gharaei S, Camp AJA, Bourne JA, Solomon SG (2011) Visual motion integration by neurons in the middle temporal area of a New World monkey, the marmoset. J Physiol 23:5741–5758.

Soto-Faraco S, Spence C, Kingstone A (2005) Assessing automaticity in the audiovisual integration of motion. Acta Psychol (Amst) 118:71–92.

Stein BE, Stanford TR (2008) Multisensory integration: current issues from the perspective of the single neuron. Nat Rev Neurosci 9:406–406.

Stein BE, Stanford TR, Rowland BA (2014) Development of multisensory integration from the perspective of the individual neuron. Nat Rev Neurosci 15:520–535.

Stoner GR (2012) Static sound timing alters sensitivity to low-level visual motion 12:1–9.

Strnad L, Peelen M V, Bedny M, Caramazza A (2013) Multivoxel pattern analysis reveals auditory motion information in MT+ of both congenitally blind and sighted individuals. PLoS One 8:e63198.

Treue S, Martinez-Trujillo JC (1999) Feature-based attention influences motion processing gain in macaque visual cortex. Nature 399:575–579.

Treue S, Maunsell JHR (1996) Attentional modulation of visual motion processing in cortical areas MT and

MST. Nature 382:539–541.

von Saldern S, Noppeney U (2013) Sensory and striatal areas integrate auditory and visual signals into behavioral benefits during motion discrimination. J Neurosci 33:8841–9.

Wallace M, Meredith MA, Stein B (1992) Integration of multiple sensory modalities in cat cortex. Exp Brain Res 91:484–488.

Wallace MT, Ramachandran R, Stein BE (2004) A revised view of sensory cortical parcellation. Proc Natl Acad Sci 101:2167–2172.

Wang Y, Celebrini S, Trotter Y, Barone P (2008) Visuo-auditory interactions in the primary visual cortex of the behaving monkey: electrophysiological evidence. BMC Neurosci 9:79.

Warren J, Zielinski B, Green G (2002) Perception of sound-source motion by the human brain. Neuron 34:139–148.

Wuerger SM, Hofbauer M, Meyer GF (2003) The integration of auditory and visual motion signals at threshold. Percept Psychophys 65:1188–1196.

Zavitz E, Yu H-H, Row EG, Rosa MGP, Price, Nicholas SC (2016) Rapid adaptation induces persistent biases in population codes. J Neurosci 36:4579–4590.

Zeki SM (1974) Functional organization of a visual area in the posterior bank of the superior temporal sulcus of the rhesus monkey. J Physiol 236:549–573.

Zeki SM, Watson JD, Lueck CJ, Friston KJ, Kennard C, Frackowiak RS (1991) A direct demonstration of functional specialization in human visual cortex. J Neurosci 11:641–649.

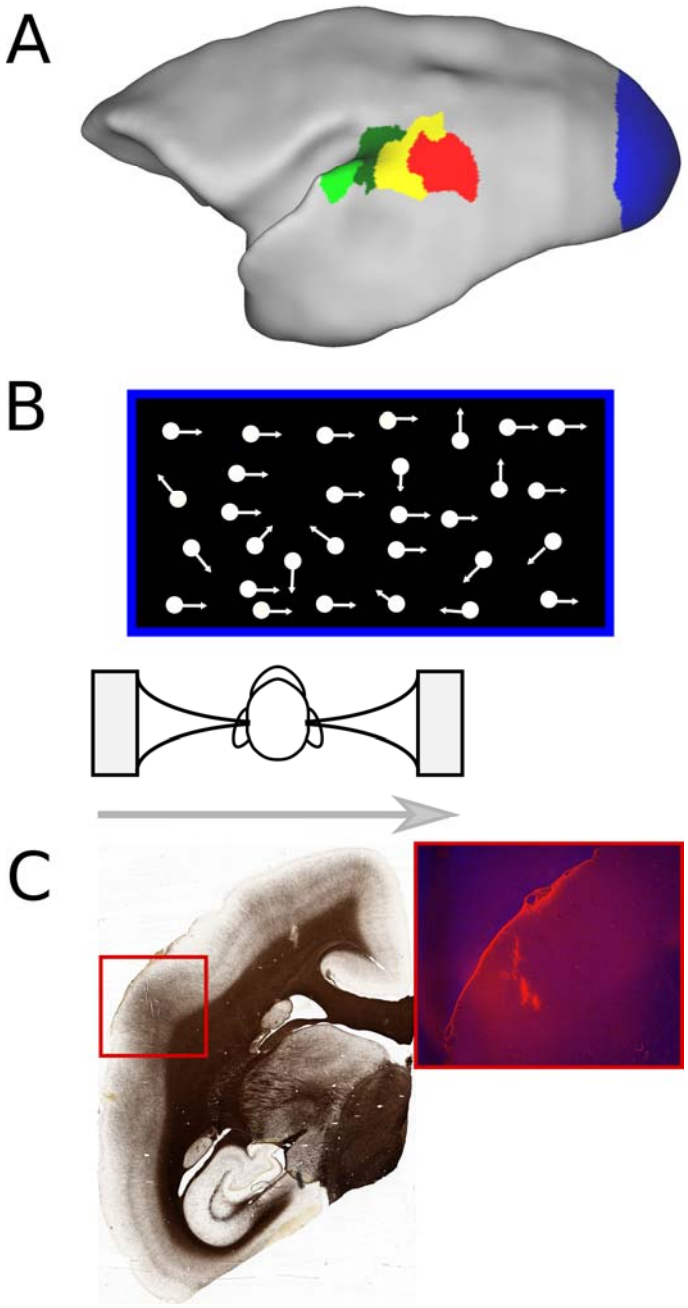


Figure 1: Experimental design and setup. A: Lateral view of a marmoset cerebral cortex model, with the relevant cortical areas - MT (red), MST (yellow), caudal higher-order auditory cortex (dark green), primary auditory cortex (light green) and primary visual cortex (blue, shown for reference). B: Experimental setup. A computer screen was positioned in front of the animal was used to present visual dot motion, either left or right, and congruent auditory motion was simulated through headphones, shown by the grey arrow. C: Histology. The image on the left shows a coronal myelin section from one case. Area MT is identifiable by the densely myelinated region in the red box. The image on the right shows the region in the red box as a fluorescence image, showing the fluorescent tracks made by the DiI-coated electrodes in area MT.

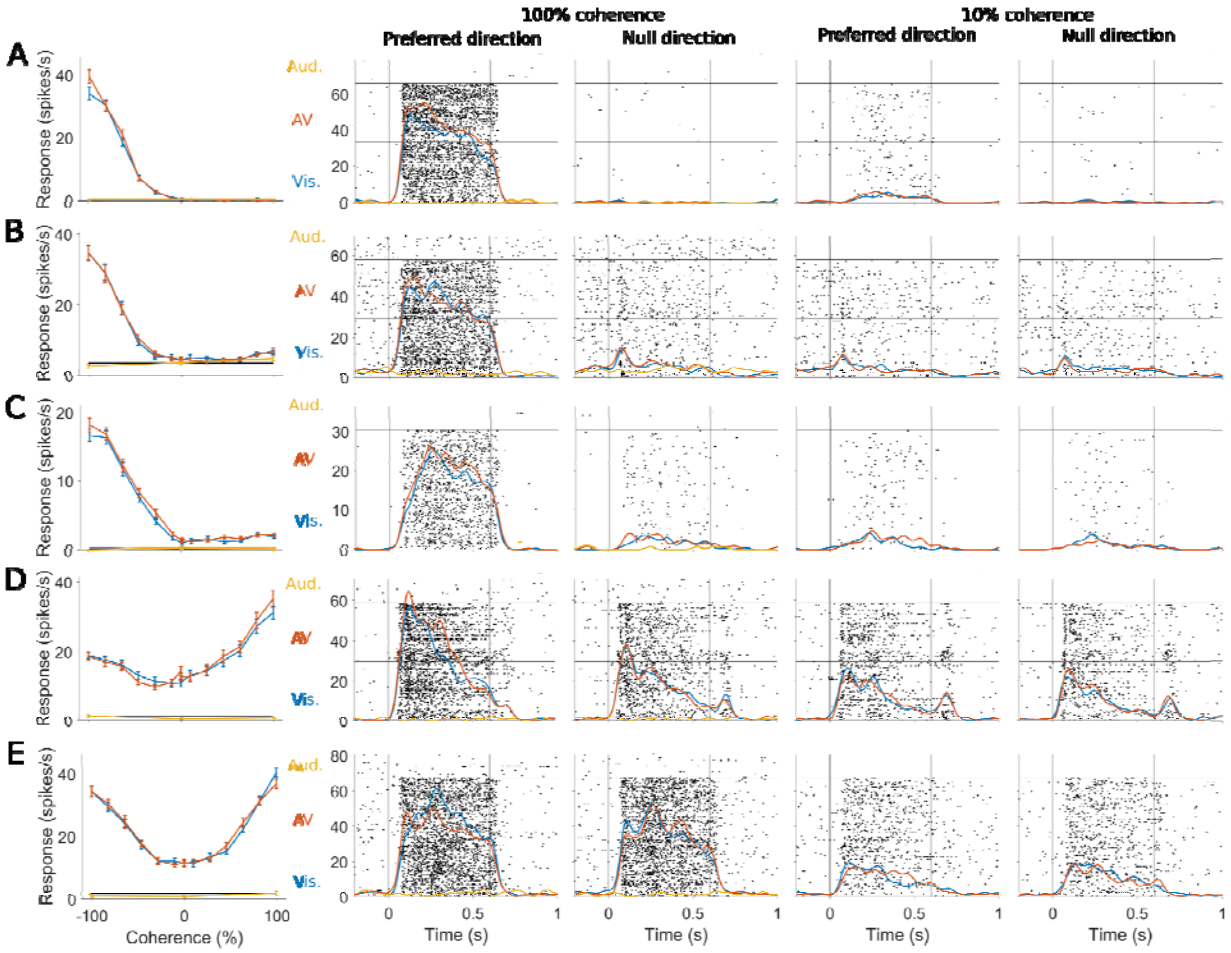


Figure 2: Example units showing clear and robust responses to visual but not auditory stimuli. Each row (A-E) shows a different unit - A: single-unit in MST, B: multi-unit in MST, C: single-unit in MT, D & E: multi-units in MT. In each row, the first panel shows the mean firing rate in response to leftwards or rightwards motion. For visual and audiovisual motion, the level of dot coherence was manipulated, as shown by the x-axis, with leftwards motion represented by negative values, and rightwards motion represented by positive values. For the auditory stimuli, only 3 conditions were used – leftwards, rightwards and stationary (plotted as responses to -100, 100 and 0% coherence respectively). The remaining panels show the spiking responses to motion in the preferred and null directions at 100% and 10% coherence, with the spike rate density functions overlaid on top of raster plots. The raster plots are separated into visual, audiovisual and auditory conditions by the horizontal lines, as shown the by the legend in the first column (note there were fewer trials for auditory only stimuli, and no equivalent of 10% coherence for auditory stimuli). The vertical lines show the onset and offset

of the stimuli.

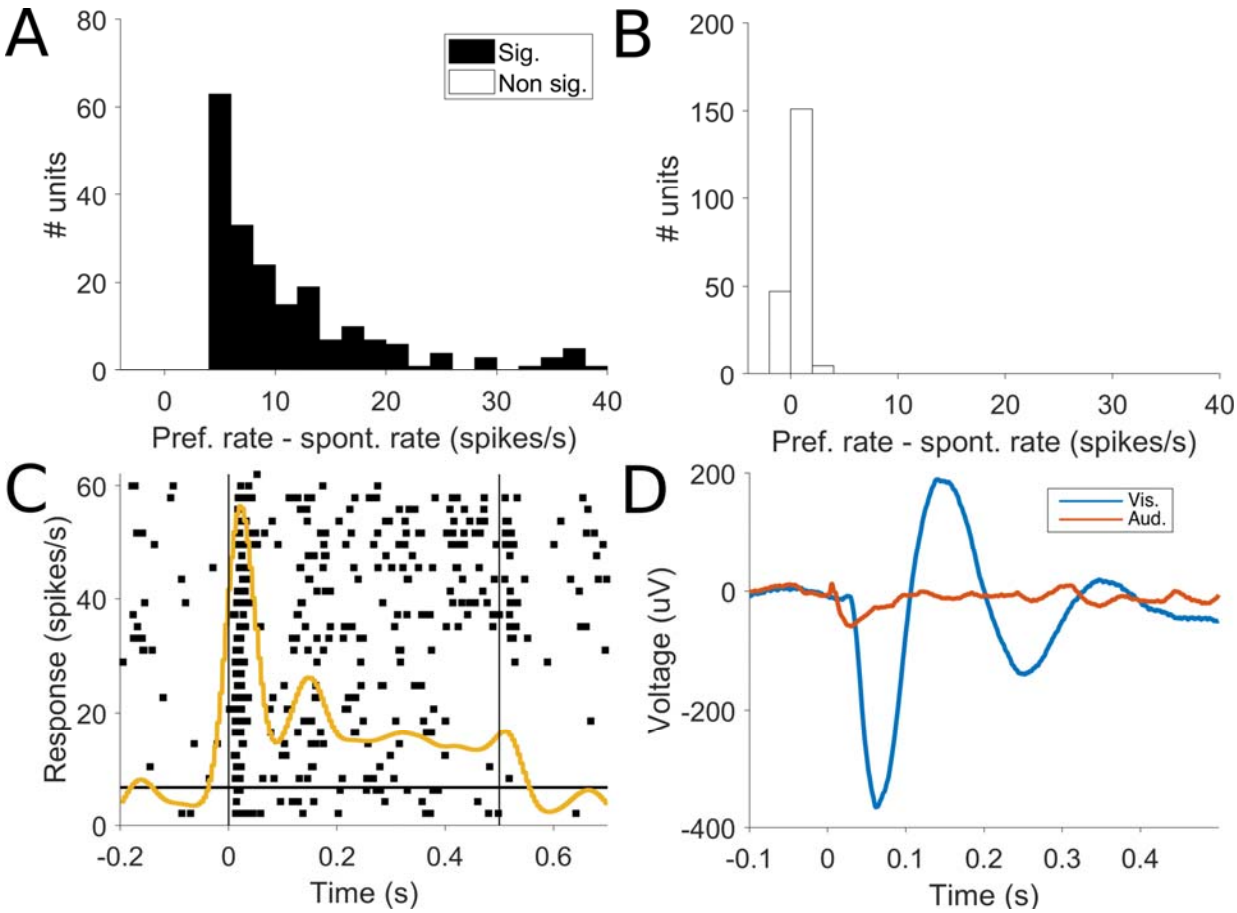


Figure 3: Responses to visual and auditory motion. A: Firing rates of visually responsive units to the best direction of motion (left or right) from 3 animals in which visual and auditory stimuli were interleaved in the same testing block. By definition, all units show statistically significant different evoked rates compared to the spontaneous rate, and are therefore colored black, as indicated in the legend. B: Responses of the same units in A to the best direction of auditory motion (left or right). None of these responses were significantly different from the spontaneous rates, and are colored white to denote this. C: An example raster plot and spike rate density function of multi-unit activity in the auditory cortex in response to auditory motion, demonstrating that this stimulus can evoke spiking activity. D: Visual flash (blue) and auditory click (red) evoked potentials, averaged across all visually responsive channels from all animals.

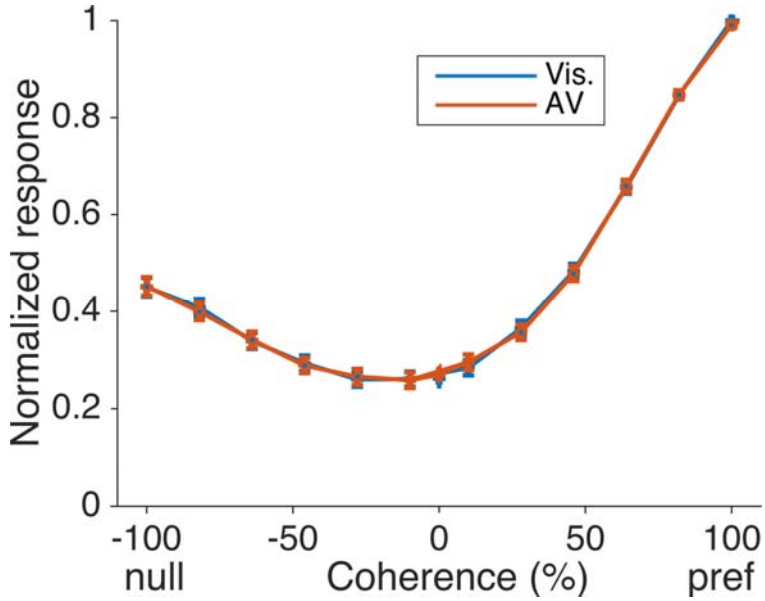


Figure 4: Normalized population coherence response function, the blue line shows the visual response, the red lines shows the audio-visual response and error bars show the standard error of the mean. Motion in the preferred direction is represented by positive coherence values and motion in the null direction is represented by negative coherence values.

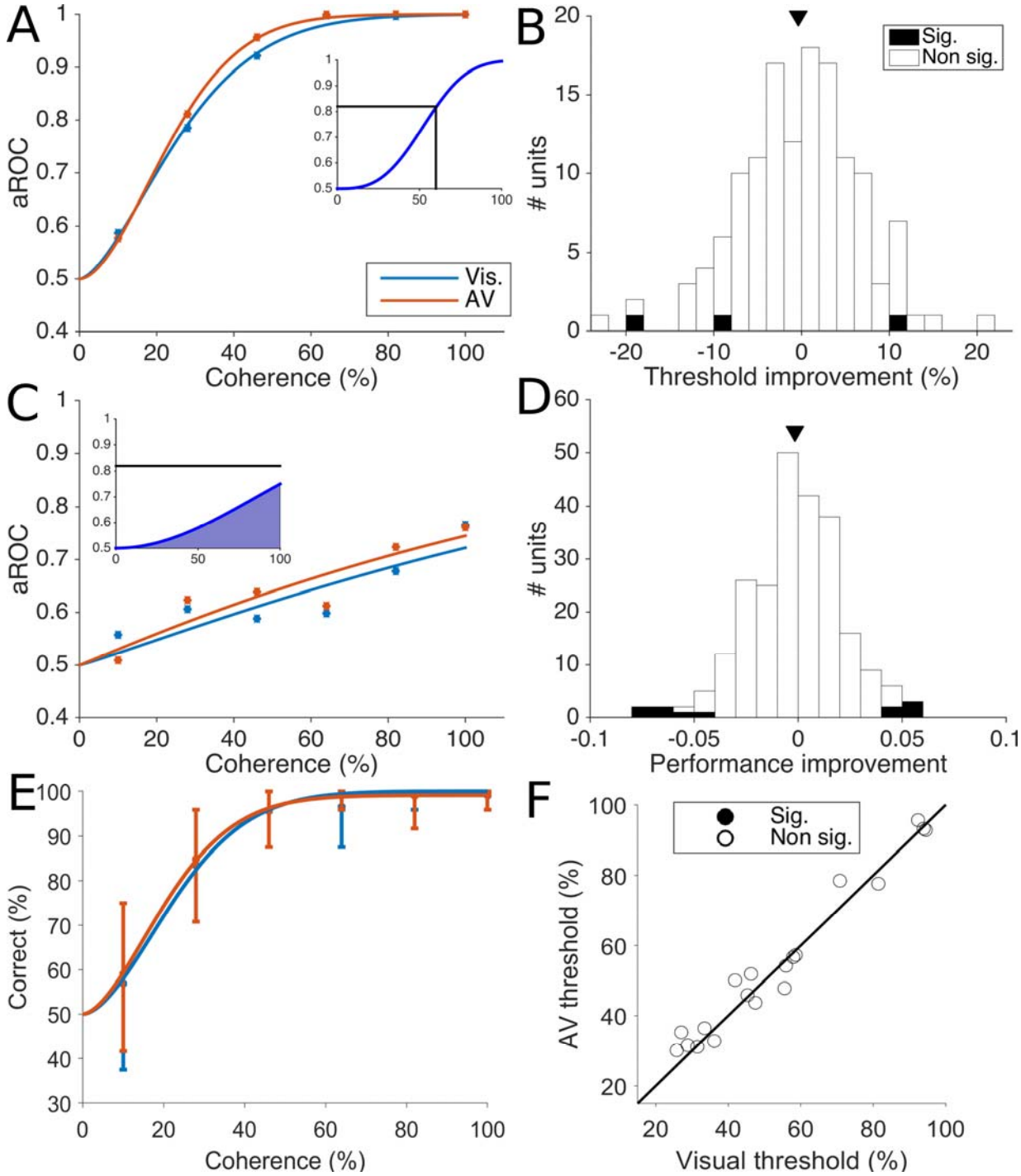


Figure 5: Neurometric thresholds and performance. A: Example neurometric curves for the unit from Figure 3A. The blue line is the visual neurometric curve and the red line is the audio-visual neurometric curve. The inset shows how the neurometric threshold is calculated (the coherence level that results in an aROC of 0.82). B: Distribution of the improvements of the audio-visual versus visual neurometric threshold, with the mean shown by the triangular marker at the top. Significance for individual units is shown as per the legend. C: Example neurometric functions for the unit from Figure 3B, which does not have a defined neurometric

threshold (i.e. does not reach an aROC of 0.82). The inset shows how the neurometric performance was calculated (the area under the curve). D: Distribution of the improvement of the audio-visual versus visual neurometric performances; the mean is shown by the triangular marker, and individual units showing significant differences are indicated by bar color. E: Example population decoding from one penetration. This penetration contained 25 individual units. The x-axis shows the coherence level, and the y-axis shows the performance of the classifier, with error bars representing the 95% interval of cross-validation iterations. F: Comparison of visual (x-axis) and audio-visual (y-axis) population neurometric thresholds. Individual penetrations showing significant differences are colored as indicated by the legend, none were significantly different.

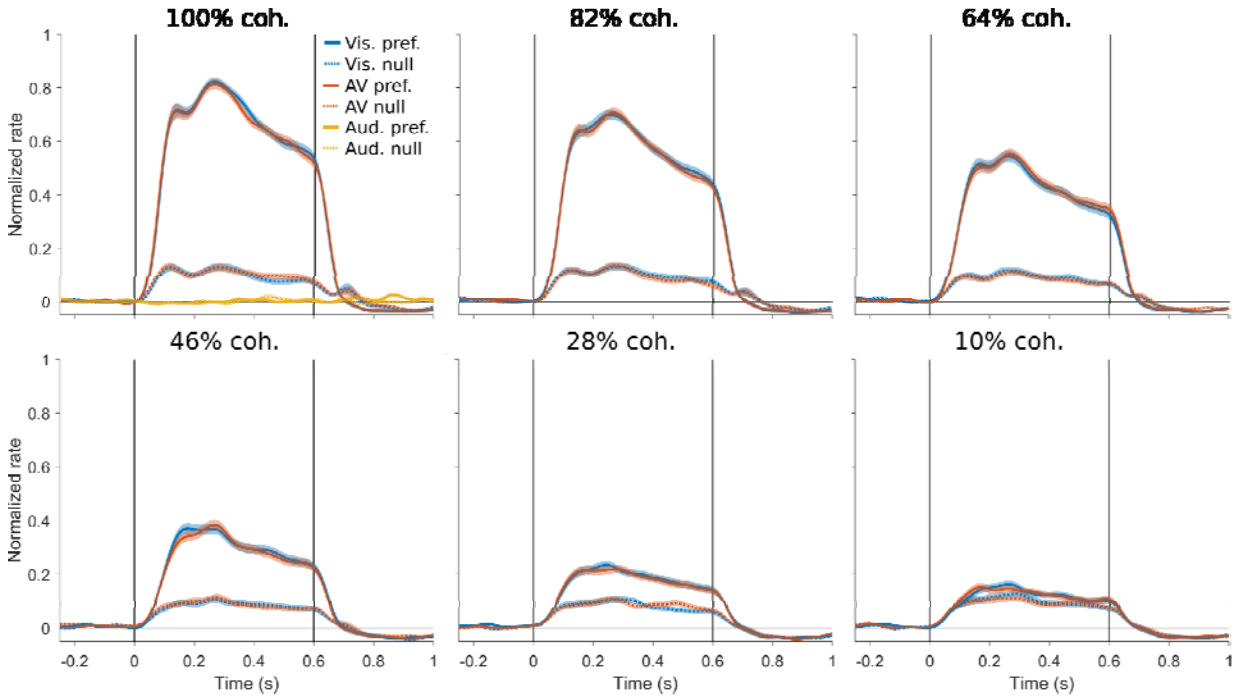
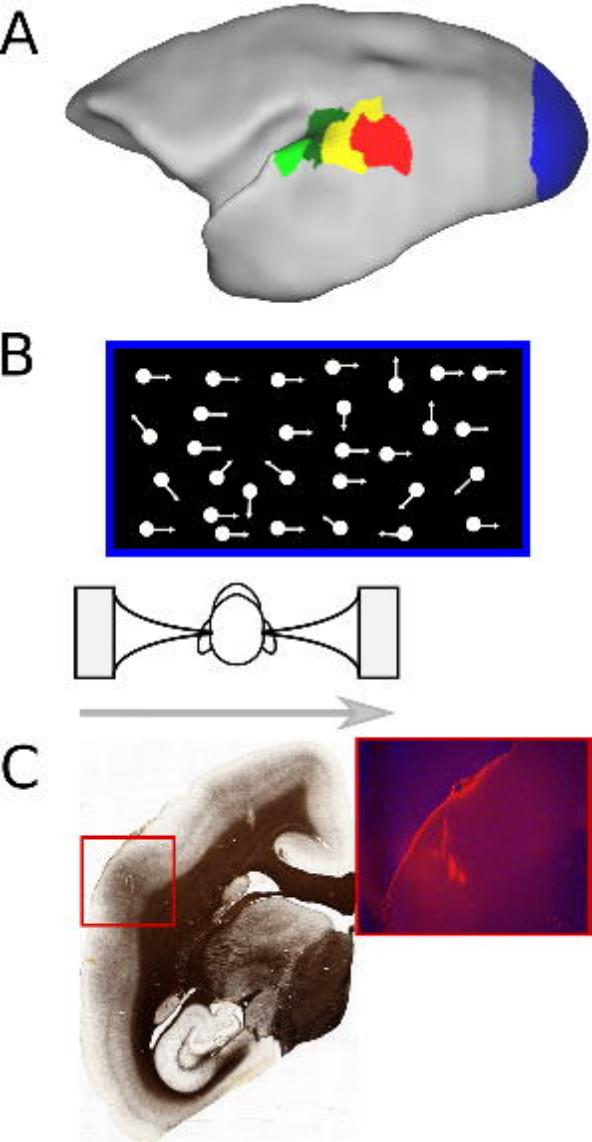
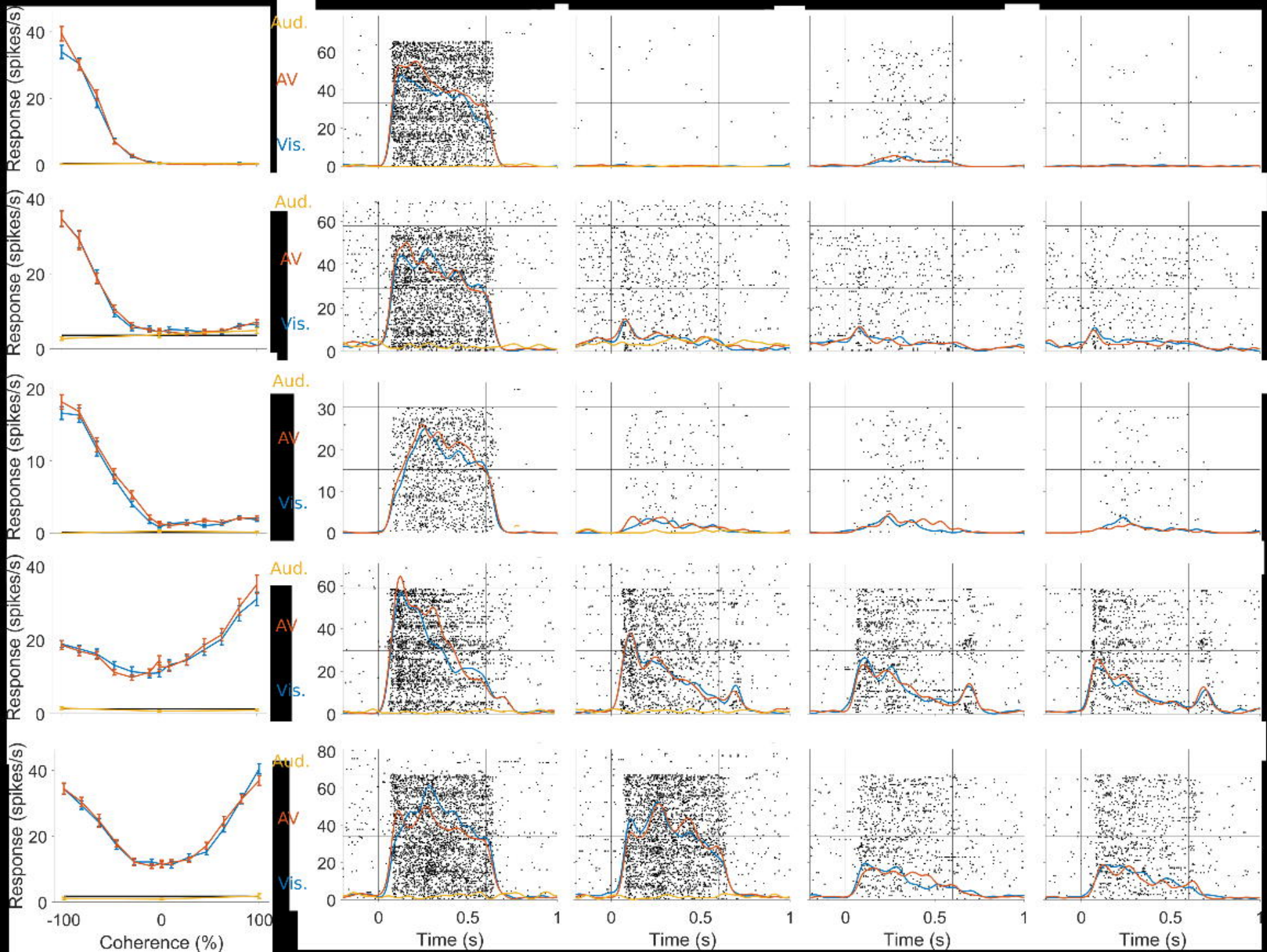
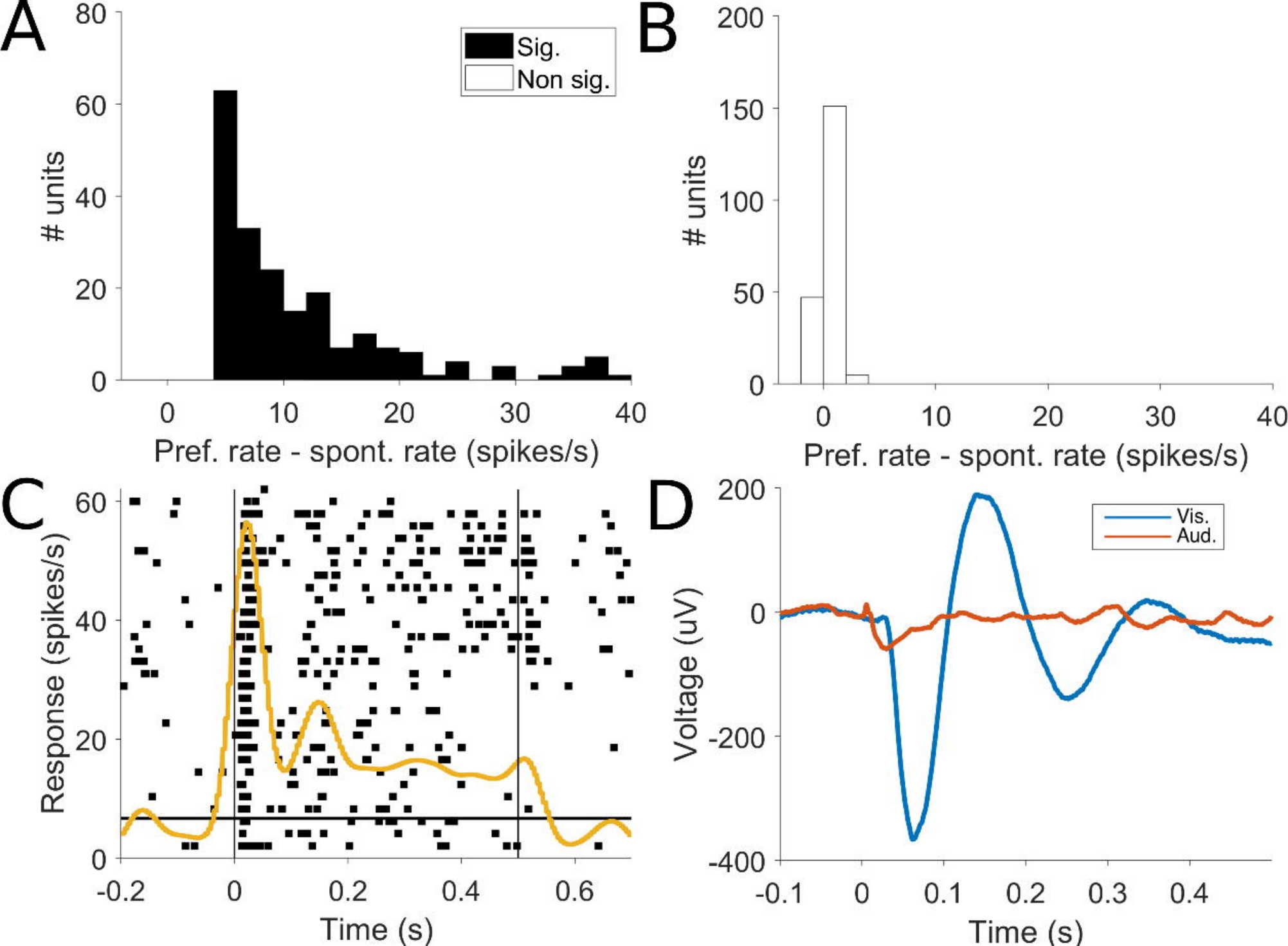
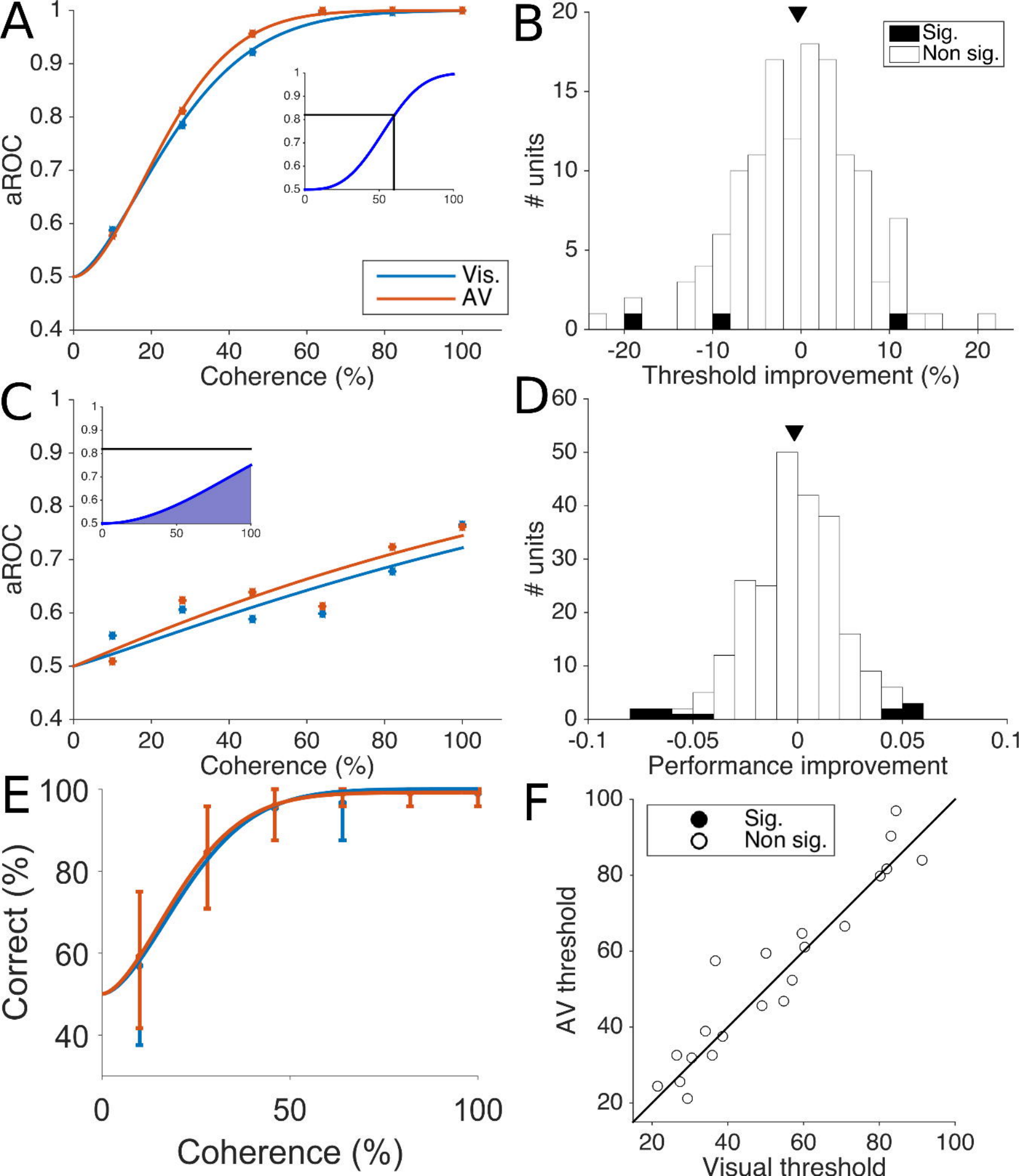


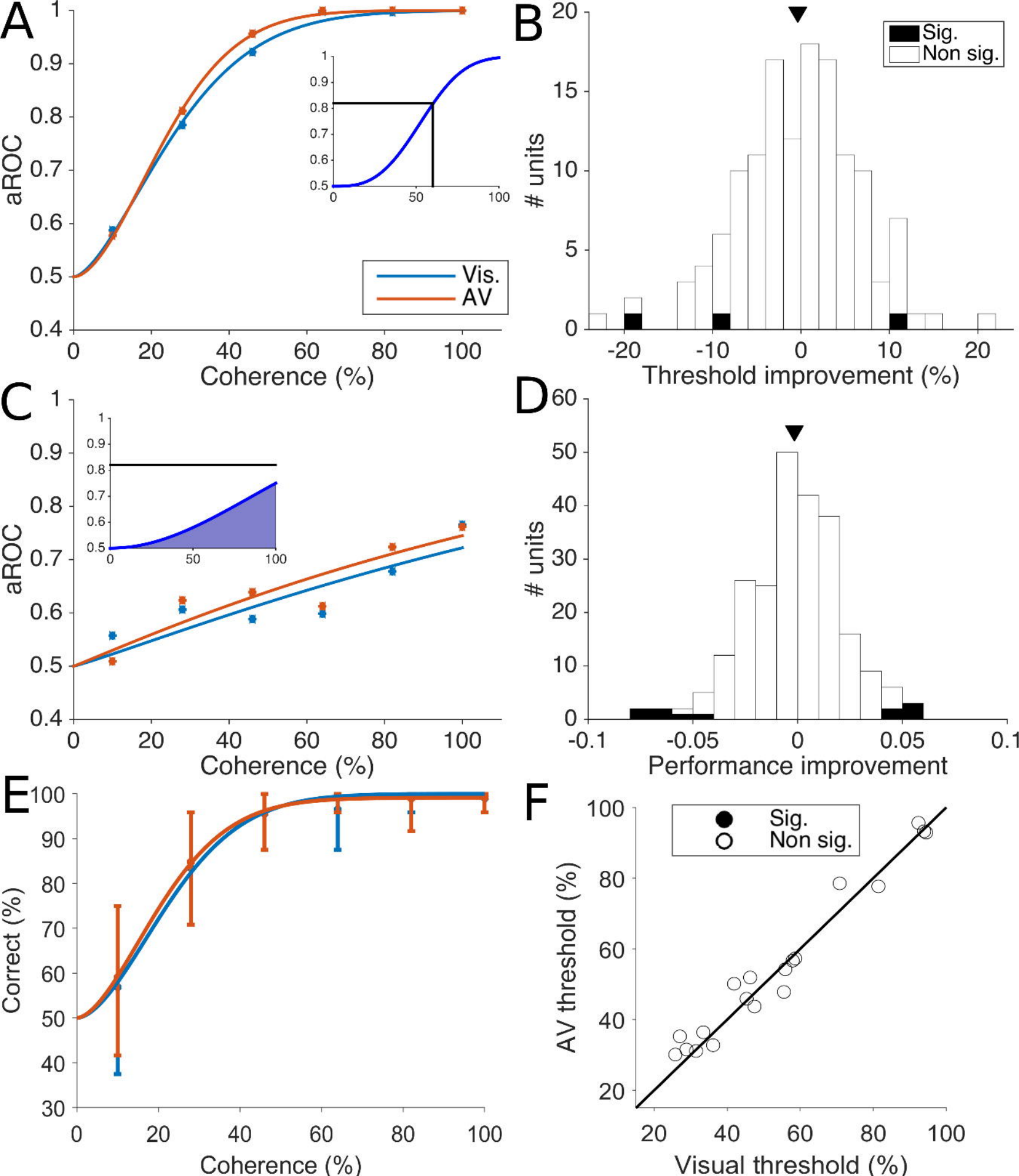
Figure 6: Temporal response profile of spiking activity in response to visual, auditory and audio-visual stimuli. The normalized population-averaged spike rate density function for each direction and coherence is shown in a separate panel, with the standard error of the mean shown by the shading. The response to visual, audio-visual and auditory stimuli in the preferred and null direction of motion are colored as indicated by the legend in the first panel. Only the 100% coherence panel shows auditory responses since there was no equivalent of coherence (stimulus noise) for the auditory stimuli. The spiking response is not modulated by the auditory stimuli at any time point (yellow lines) and the difference between the visual and audiovisual stimuli (blue and red lines respectively) is not different at any time point.

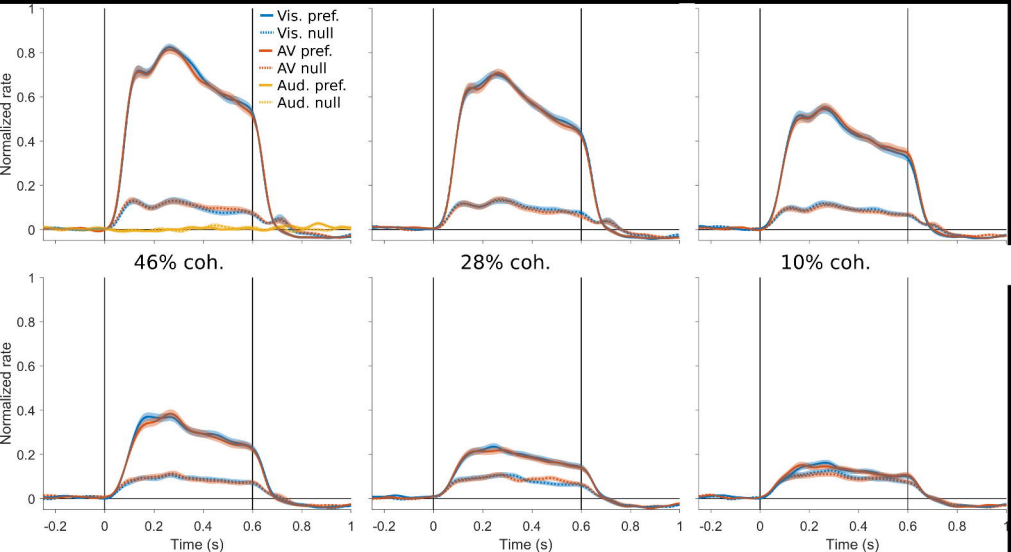

A
B
C


Response (spikes/s)
Aud.
AV
Vis.
Coherence (%)
Time (s)


A
Sig.
Non sig.
units
Pref. rate - spont. rate (spikes/s)
B
units
Pref. rate - spont. rate (spikes/s)
C
Response (spikes/s)
Time (s)
D
Vis.
Aud.
Voltage (uV)
Time (s)


A
aROC
Coherence (%)
Vis.
AV
B
units
Threshold improvement (%)
Sig.
Non sig.
C
aROC
Coherence (%)
D
units
Performance improvement
E
Correct (%)
Coherence (%)
F
AV threshold
Visual threshold
Sig.
Non sig.


A
aROC
Coherence (%)
Vis.
AV
B
units
Threshold improvement (%)
Sig.
Non sig.
C
aROC
Coherence (%)
D
units
Performance improvement
E
Correct (%)
Coherence (%)
F
AV threshold (%)
Visual threshold (%)
Sig.
Non sig.


Vis. pref.
Vis. null
AV pref.
AV null
Aud. pref.
Aud. null
Normalized rate
46% coh.
28% coh.
10% coh.
Time (s)
-0.2
0
0.2
0.4
0.6
0.8
1