

RESEARCH ARTICLE | *Sensory Processing*

Auditory modulation of spiking activity and local field potentials in area MT does not appear to underlie an audiovisual temporal illusion

Hulusi Kafaligonul,^{1,2} Thomas D. Albright,³ and Gene R. Stoner³

¹National Magnetic Resonance Research Center, Bilkent University, Ankara, Turkey; ²Interdisciplinary Neuroscience Program, Bilkent University, Ankara, Turkey; and ³Vision Center Laboratory, The Salk Institute for Biological Studies, La Jolla, California

Submitted 20 November 2017; accepted in final form 19 June 2018

Kafaligonul H, Albright TD, Stoner GR. Auditory modulation of spiking activity and local field potentials in area MT does not appear to underlie an audiovisual temporal illusion. *J Neurophysiol* 120: 1340–1355, 2018. First published June 20, 2018; doi:10.1152/jn.00835.2017.—The timing of brief stationary sounds has been shown to alter the perceived speed of visual apparent motion (AM), presumably by altering the perceived timing of the individual frames of the AM stimuli and/or the duration of the interstimulus intervals (ISIs) between those frames. To investigate the neural correlates of this “temporal ventriloquism” illusion, we recorded spiking and local field potential (LFP) activity from the middle temporal area (area MT) in awake, fixating macaques. We found that the spiking activity of most MT neurons (but not the LFP) was tuned for the ISI/speed (these parameters covaried) of our AM stimuli but that auditory timing had no effect on that tuning. We next asked whether the predicted changes in perceived timing were reflected in the timing of neuronal responses to the individual frames of the AM stimuli. Although spiking dynamics were significantly, if weakly, affected by auditory timing in a minority of neurons, the timing of spike responses did not systematically mirror the predicted perception of stimuli. Conversely, the duration of LFP responses in β - and γ -frequency bands was qualitatively consistent with human perceptual reports. We discovered, however, that LFP responses to auditory stimuli presented alone were robust and that responses to audiovisual stimuli were predicted by the linear sum of responses to auditory and visual stimuli presented individually. In conclusion, we find evidence of auditory input into area MT but not of the nonlinear audiovisual interactions we had hypothesized to underlie the illusion.

NEW & NOTEWORTHY We utilized a set of audiovisual stimuli that elicit an illusion demonstrating “temporal ventriloquism” in visual motion and that have spatiotemporal intervals for which neurons within the middle temporal area are selective. We found evidence of auditory input into the middle temporal area but not of the nonlinear audiovisual interactions underlying this illusion. Our findings suggest that either the illusion was absent in our nonhuman primate subjects or the neuronal correlates of this illusion lie within other areas.

audiovisual interactions; motion processing; multisensory; temporal ventriloquism; visual area MT

INTRODUCTION

Although it was long thought that information from different sensory modalities only converged in higher-order “association areas,” the emerging view is that some types of cross-modal interactions occur within areas previously thought to be “sensory specific” (Driver and Noesselt 2008; Schroeder and Foxe 2005; Senkowski et al. 2008). Audiovisual interactions have been especially studied well. Of particular interest here is the phenomenon of temporal ventriloquism, in which the timing of brief sounds drives the perceived timing of visual events (Fendrich and Corballis 2001; Morein-Zamir et al. 2003; Recanzone 2003). Temporal ventriloquism has been extensively studied in the context of visual apparent motion (AM) in which spatially separated visual stimuli turned off and on sequentially can give rise to a sense of visual motion (Kolars 1972; Nakayama 1985). In particular, the timing of brief sounds has been shown to alter the perceived direction and/or speed of AM stimuli, presumably by altering the perceived timing of the individual frames of the AM stimuli and/or the duration of the interstimulus intervals (ISIs) between those frames (Freeman and Driver 2008; Getzmann 2007; Kafaligonul and Stoner 2010, 2012; Shi et al. 2010; Staal and Donderi 1983).

By devising AM stimuli that selectively engaged low-level visual motion mechanisms, we found psychophysical evidence that the ability of sound timing to alter the perceived speed of visual motion is subserved, at least in part, within early stages of visual motion processing (Kafaligonul and Stoner 2010, 2012). Specifically, we used AM stimuli designed to engage the cortical middle temporal area (area MT). Area MT is an extrastriate visual cortical area with neurons tuned to direction and speed when tested with smoothly moving visual stimuli (Albright 1984; Dubner and Zeki 1971; Maunsell and Van Essen 1983; Perrone and Thiele 2001) as well as with the type of AM stimuli used in this study (Churchland et al. 2005; Mikami 1991, 1992; Mikami et al. 1986a, 1986b; Newsome et al. 1986). Area MT’s responses to AM stimuli can also be characterized in terms of ISI tuning, with ISI and speed inversely correlated for fixed spatial intervals between visual stimuli (see DISCUSSION for more details). In addition to area MT’s ISI/speed selectivity for these types of stimuli, there are two other reasons for thinking that area MT might be involved in this audiovisual temporal illusion. First, studies using transcranial magnetic stimulation have found evidence that human

Address for reprint requests and other correspondence: G. R. Stoner, Vision Center Laboratory, The Salk Institute for Biological Studies, La Jolla, CA 92037-1099 (e-mail: gene@salk.edu).

area MT+ is also involved in the perceptual timing of visual events (Buetti et al. 2008), and second, several neuroimaging studies on humans have found auditory influences on human area MT+ (Alink et al. 2008; Calvert et al. 1999; Molholm et al. 2002; Scheef et al. 2009). Human area MT+ is believed to be composed of homologs to the macaque motion-responsive visual areas, MT and the medial superior temporal area (MST; Huk et al. 2002).

Motivated by the above evidence, we looked for neuronal correlates of this form of temporal ventriloquism within visual area MT. We examined neuronal responses, single-unit and multiunit activity and LFPs, to three-frame visual AM stimuli accompanied by three clicks having ISIs that were either shorter or longer than the visual ISIs. We first asked whether the ISI/speed tuning of area MT neurons, based on spike rates (LFP response magnitudes were not found to be tuned to ISI/speed for our stimuli), were affected by auditory timing. We found no significant effect of auditory timing on ISI/speed tuning. We next asked whether the predicted changes in perceptual timing were reflected in the timing (rather than magnitude) of neuronal responses to our audiovisual stimuli. Specifically, we tested the hypothesis that AM stimuli with auditory ISIs that were longer than the visual ISIs resulted in responses of longer duration (and corresponding changes in onset and/or offset) than AM stimuli with auditory ISIs that were shorter than the visual ISIs. We found that auditory timing weakly, but significantly, modulated the timing of spiking responses to AM stimuli in a minority of neurons. These effects were not, however, generally consistent with our hypothesis. In contrast, we found that the duration of LFP responses to our audiovisual stimuli was generally consistent with our hypothesis. We discovered, however, that auditory stimuli presented alone elicited LFP responses and, moreover, that responses to audiovisual stimuli were not, on average, statistically distinguishable from the linear sum of the responses to the component visual and auditory stimuli. We thus found no evidence that sounds nonlinearly modulate LFP response durations to visual stimuli. In conclusion, we find evidence of auditory input into area MT but no obvious nonlinear neuronal correlate of the temporal ventriloquism illusion.

MATERIALS AND METHODS

Subjects and Surgical Procedures

Two adult male rhesus monkeys (*Macaca mulatta*; monkeys S and B) were used in these experiments. Experimental protocols were approved by the Salk Institute Animal Care and Use Committee and conform to U.S. Department of Agriculture regulations and to the National Institutes of Health guidelines for the care and use of laboratory animals. Procedures for surgical preparation, behavioral training, and electrophysiological recording were routine and similar to those described previously (Duncan et al. 2000; Krekelberg and Albright 2005). Briefly, each monkey was implanted with a stainless-steel head post and a recording cylinder oriented vertically allowing recording from neurons in area MT. The positioning of the chamber was guided by magnetic resonance imaging (MRI) scans obtained at the University of California, San Diego, Center for Magnetic Resonance Imaging. During neural recording, monkeys were seated in a standard primate chair (Crist Instrument) with the head post rigidly supported by the chair frame to prevent head movement.

Stimuli

Stimulus presentation, behavioral paradigm, and data acquisition were controlled by specialized CORTEX software (Laboratory of Neuropsychology, National Institute of Mental Health, Bethesda, MD; <https://www.nimh.nih.gov/labs-at-nimh/research-areas/clinics-and-labs/ln/shn/index.shtml>). Visual stimuli were presented on a 19-in. cathode ray tube monitor (Trinitron E500; Sony; $1,024 \times 768$ -pixel resolution and 100-Hz refresh rate) at a viewing distance of 57 cm. A PR701S photometer (Photo Research, Syracuse, NY) was used for luminance calibration and gamma correction of the display. Sounds were emitted by two speakers positioned at the top of the visual display. The center-to-center vertical distance between the speakers and monitor was 26 cm corresponding to 26° at a viewing distance of 57 cm. Sound intensities were checked regularly with a sound level meter (33-2055; RadioShack). Timing of visual and auditory stimuli was confirmed with a digital (TDS 1002; Tektronix) oscilloscope connected to the computer sound card and a photodiode. All stimuli were presented while monkeys performed a visual fixation task (see MATERIALS AND METHODS, *Behavioral Paradigm*). Stimuli are described in the relevant sections below.

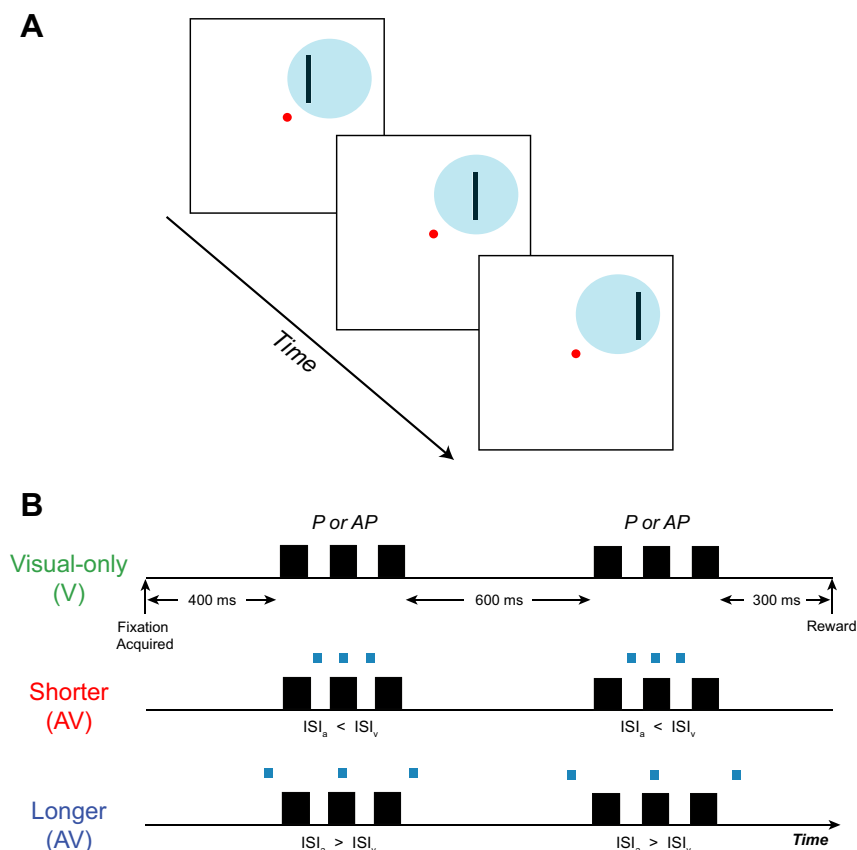
Initial estimate of preferred direction. Before mapping the classical receptive field (CRF) of a given neuronal recording, we estimated the preferred direction using a large random-dot patch ($37 \times 28^\circ$) undergoing circular translation (Schoppmann and Hoffmann 1976). Dot luminance was 20.71 cd/m^2 , and background luminance was 0.67 cd/m^2 . This method allows a continuous and complete mapping of directional responses in a single trial. Ten to twenty trials were typically used to estimate the preferred direction of each neuron. This method has been shown to agree nicely with conventional methods for estimating preferred direction (see below).

CRF mapping. CRFs were mapped by recording responses to square-wave gratings that drifted in the preferred direction as estimated above. Gratings were presented within individual squares of a spatial grid (usually $25 \times 20^\circ$, sometimes $40 \times 30^\circ$) with grid lines separated by 5° . Each square (and hence each grating) in this grid was $5 \times 5^\circ$. Gratings appeared, were static for 50 ms, and then moved for 500 ms. Only one grating was shown per trial. The raw CRF map was interpolated using the MATLAB 7.10 (The MathWorks, Natick, MA) function “interp2” at an interval of 0.5° , using “bicubic” interpolation. The location in the interpolated map that gave rise to the highest firing rate was taken as the CRF center, over which stimuli were then centered. Gratings were presented at least four times at each spatial location until the CRF map stabilized (i.e., the estimated CRF center and size did not change significantly).

Determining preferred and antipreferred directions. After mapping the CRF, we further characterized directional tuning using square-wave gratings drifting in one of eight directions at $8^\circ/\text{s}$. The gratings were viewed through an invisible circular aperture with a diameter of 6° centered on the CRF (see above). The preferred (P) direction was determined using an online analysis script running in MATLAB 7.10 (The MathWorks). With few exceptions, and consistent with previous findings (Huang et al. 2007, 2008), the preferred direction estimated this way matched the initial estimation of the preferred direction using the circular translation stimuli described above. For those exceptions, CRF mapping was repeated using the newly determined preferred direction. The antipreferred (AP) direction was defined as the direction opposite to the preferred direction.

Three-frame AM and auditory stimuli. AM stimuli consisted of three sequentially “flashed” (30 ms) bars ($0.3 \times 3.0^\circ$) centered within the CRF (Fig. 1A). The spatial displacement (i.e., center-to-center separation) of each consecutively flashed bar was 0.4° . The visual ISIs (ISI_v s) between the bars in a given sequence were chosen pseudorandomly from eight values: 20, 40, 60, 80, 100, 120, 160, and 200 ms. On the basis of the formula we used to calculate physical speed [$\text{speed} = 0.4^\circ / (30 \text{ ms} + \text{ISI}_v)$], these eight ISI_v s corresponded to speeds of 8, 5.71, 4.44, 3.64, 3.08, 2.67, 2.11, and $1.74^\circ/\text{s}$. ISI_v s

Fig. 1. *A*: visual stimuli. Apparent motion (AM) stimuli consisted of three flashed bars. AM stimuli were centered within the receptive field. Motion was either in the preferred (P) or antipreferred (AP) direction of the recorded neuron(s). *B*: visual and auditory stimulus timing for each audiovisual configuration. Visual bars and auditory clicks are indicated by black rectangles and blue squares, respectively. The three clicks were temporally centered relative to the three bars. Relative durations of each visual and auditory event are indicated by the thickness of black rectangles and blue squares, respectively (height of these icons distinguishes stimulus modality). Four hundred milliseconds after fixation acquisition, AM stimuli first moved either in the P or in the AP direction and then moved in the opposite direction after an interval of 600 ms. Interstimulus intervals between each bar presentation [visual interstimulus intervals (ISI_v)] varied from 20 to 200 ms. Visual-only (V) stimuli had no sounds (i.e., clicks). For audiovisual (AV) stimuli, the clicks had auditory interstimulus intervals (ISI_a) that were either shorter (AV shorter) or longer (AV longer) than the ISI_v s. Previous studies have found that the shorter condition is perceived as moving faster than the longer condition, consistent with a reduction in the perceived ISI_v . In a subset of recording sessions, we also presented auditory-only stimuli that were identical to AV stimuli but without visual stimuli (see MATERIALS AND METHODS).



and speed covaried (inversely) but we will refer mostly to ISI_v s rather than speeds when discussing our results. The luminance of each bar was 21.71 cd/m², and the background was 0.67 cd/m². Auditory stimuli were three successive 10-ms clicks. Each click comprised a rectangular-windowed 480-Hz sine wave carrier, sampled at 22 kHz with 8-bit quantization. The sequence of three clicks was temporally centered with respect to the three visual frames (Fig. 1*B*).

For each trial, three visual frames were presented in both the P and AP directions with a 600-ms temporal offset between the two AM stimuli (i.e., between the end and the start of the flashed bars in each of the AM stimuli). The order of P and AP directions (i.e., P then AP vs. AP then P) was chosen pseudorandomly. As shown in Fig. 1*B*, each neurophysiological recording included a mixture of two audiovisual (AV) stimulus conditions, shorter (auditory ISIs were 20 ms) and longer (auditory ISIs were 50 ms longer than visual ISIs), and one visual-only (V: no clicks) stimulus condition. The two audiovisual conditions thus had the same number of auditory events, but with different timings. These two audiovisual conditions, together with the visual-only condition, constitute the three main audiovisual configurations in our experiments. In addition, some recording sessions ($n = 38$) included two auditory-only (A) shorter and longer conditions. These conditions were restricted to four visual ISIs (i.e., shorter and longer conditions of 20-, 60-, 100-, and 160-ms visual ISIs) and, excepting the absence of visual stimuli, were identical to corresponding audiovisual shorter and longer conditions. Since there were no P and AP directions for these conditions, the number of trials for each of these conditions was half of those in the corresponding AV and V conditions so that the number of stimulus presentations was the same for all conditions. Conditions were intermixed pseudorandomly and were presented 16 times during each recording session, with the exception of 3 recording sessions, for which each AM stimulus condition was presented 10–12 times.

Behavioral Paradigm

Eye position was sampled at 60 Hz using an infrared video-based tracking system (ISCAN, Burlington, MA). Monkeys were required to maintain fixation on a small (0.2° diameter) centrally located red spot within a 2 × 2° window during each experimental trial. This window allowed for drifts in the centering of eye position on the fixation target: deviations from fixation were typically very much less than half of this window size. In each trial, after the monkey had acquired and maintained fixation for 400 ms (Fig. 1*B*), the visual stimulus appeared. After the offset of the last visual flash, the fixation spot remained for another 300 ms. Upon successful completion of a trial, monkeys were rewarded with juice. Trials were separated by a 1-s intertrial interval in addition to the time it took for the monkey to achieve fixation.

Electrophysiological Recording

We recorded neuronal activity using tungsten microelectrodes (FHC; 3–5 MΩ base impedance), which were driven into cortex using a hydraulic micropositioner (model 650; David Kopf Instruments). Neurophysiological signals were filtered, sorted, and stored using the Plexon system (Plexon, Dallas, TX). The electrode signal was passed through a headstage with unit gain and then split to separately extract spiking activity and LFPs. For LFPs, the signal was filtered by a band-pass hardware filter (3–90-Hz range) before being amplified and digitized at 1 kHz. For spike recordings, the signal was band-pass filtered between 250 and 8,000 Hz, amplified, and digitized at 40 kHz. Single-unit spiking activity was then isolated using a window discriminator.

MRI scans were used to guide electrode placement. We identified area MT by its characteristically large proportion of directionally selective cells, its small CRFs relative to those of neighboring area MST, and its location on the posterior bank of the superior temporal

sulcus. Recording depths of physiologically identified MT neurons agreed well with the expected anatomical location based on structural MRI scans. Action potentials were classified as “single unit” (i.e., as coming from an individual neuron) if those waveforms were, on the basis of the raw waveforms and the principal component analysis of the Plexon spike sorter, clearly clustered and distinct from the baseline noise and other clusters of spikes. Action potentials that crossed a magnitude threshold and had stable waveforms but did not meet the criteria for a single unit were grouped together and classified as “multiunit.”

Data Analysis

We analyzed neuronal responses with software developed by us and written in MATLAB 7.10 (The MathWorks). Details of our response metrics are given below.

ISI Tuning

Spike analyses for ISI tuning. The first hypothesis we tested was that the ISI/speed tuning of neuronal responses in area MT reflected selectivity for perceived ISI/speed as influenced by auditory timing. We thus predicted that ISI tuning would be shifted to reflect the shifts in perceived speed as observed in previous studies using these stimuli. This prediction rests on two assumptions: 1) Monkeys passively fixating while viewing these stimuli are subject to the temporal ventriloquism illusion documented in human subjects, and 2) area MT neurons are tuned to the perceived rather than the physical ISI between AM frames. To test this prediction, we computed ISI tuning curves for responses to AM stimuli moving in these P and AP directions and then for the response differences (P-AP) to those two directions. To ensure that our computation of average firing rate for each stimulus presentation included onset and offset responses (the timing and magnitude of which might be modulated by the sounds), we conservatively averaged over a response window beginning at the onset of the first visual frame and ending 200 ms after the onset of the third (final) frame. This response window necessarily varied with ISI of the visual stimuli (ISI_v) thereby introducing a potential confounder in the interpretation of P and AP ISI tuning curves. If, for example, responses were identical in amplitude and duration for all ISI_v s, then computing the average firing rate over a larger response window for longer ISI_v s would yield tuning curves that appear to favor shorter ISI_v s. Indeed, such apparent low-pass ISI tuning is the dominant trend in the ISI tuning curves computed for the P direction and, to a lesser degree, for the AP direction. Although drawing conclusions about ISI tuning from the shape of P or AP ISI tuning curves is problematic because of this response window confounder, an impact of auditory stimulation on ISI tuning for P and AP stimuli can be inferred from significant interactions between ISI_v and audiovisual configuration (2-way ANOVA: visual ISI and audiovisual configuration as factors). As discussed in MATERIALS AND METHODS, *Stimuli*, there were three main audiovisual configurations: AV shorter, AV longer, and visual-only. Significant interactions would indicate that the presence and/or timing of auditory stimulation affected ISI tuning.

Since P and AP responses were computed over the same time period, ISI tuning curves based on P minus AP responses (yielding a measure of direction selectivity) do not suffer from the response window confounder discussed above: the difference between them thus accurately identifies how directional selectivity varies as a function of ISI_v . Previous studies similarly examined directional selectivity for AM stimuli with different ISI_v s on the basis of P-AP responses (Mikami 1991, 1992). In our study, we presented P and AP directions in the same trial (Fig. 1B) and thus had a trial-by-trial measure of P-AP and hence of directional selectivity. The impact of ISI_v and audiovisual configuration on directional selectivity (P-AP responses) was examined by two-way ANOVA (visual ISI and audiovisual configuration as factors).

LFP analyses for ISI tuning. Raw LFP signals from each trial were first filtered with a second-order notch filter (quality factor 10) with a 60-Hz center frequency to remove 60-Hz noise artifacts. To compute ISI tuning curves from each recording session, single-trial instantaneous LFP amplitudes were extracted by wavelet decomposition (Morlet wavelets) on 92 scales from 1 to 92 Hz. For frequencies lower than 30 Hz, we defined frequency bands using the conventions of human electroencephalography (EEG frequency bands): θ , 4–8 Hz; α , 8–12 Hz; β , 13–30 Hz (Buzsáki 2006). For frequencies higher than 30 Hz (γ -band) and up to 90 Hz, we used successive, nonoverlapping, 30-Hz-wide bands (γ_1 , 30–60 Hz; γ_2 , 60–90 Hz).

Mean LFP amplitude over time for each frequency band was computed by averaging wavelet-transformed amplitudes over each of these five frequency ranges. These frequency-specific average magnitudes were summed over the same temporal interval used to measure neuronal responses and then divided by that interval. For each trial, we computed these magnitudes for P, AP, and P-AP. We analyzed LFP tuning curves using ANOVA as described above for spike data. We also used a bootstrap procedure to test for significant differences between P and AP (see MATERIALS AND METHODS, *Audiovisual Interactions*, below).

Response Timing

We also tested the hypothesis that the timing of the auditory stimuli shifted the timing of neuronal responses to the visual AM stimuli.

Spiking dynamics. To identify auditory timing-induced changes in spiking dynamics, we first determined whether each neuron responded significantly to AM stimuli moving in the P and/or AP directions and whether the response difference between these directions (P-AP) significantly varied across time. For each trial of each condition, spikes were counted in 10-ms time bins beginning at the onset of the first AM frame and ending 200 ms after the last frame. We then performed a two-way ANOVA (time bin and audiovisual configuration as factors) for each ISI_v condition and direction (i.e., P, AP, and P-AP). Audiovisual configurations again consisted of AV shorter, AV longer, and visual-only. In subsequent analyses, we only used recordings in which spike responses varied significantly with time bin, which merely indicated a significant response to the stimulus in question: estimating response duration obviously only makes sense for significant responses. Nonsignificant cases were typically for stimuli with large ISI_v s for which neuronal responses were weak.

We hypothesized that the timing of each auditory event (i.e., click) would shift (or “capture”) the timing of neuronal responses to each visual event (i.e., AM frame) in the direction of the auditory timing. This hypothesis follows from the assumption that changes in perceived visual ISI (ISI_v) result from auditory-induced changes in the timing of the responses to the individual visual events. On the basis of this assumption and our previous demonstration of auditory-induced changes in perceived ISI_v /speed of AM stimuli (Kafaligonul and Stoner 2010, 2012), we specifically hypothesized that the response duration for the longer auditory timing condition would be longer than for the shorter auditory timing condition.

To estimate response duration, we first estimated onset and offset latencies, using methods adapted from previous studies (Huang et al. 2007; Kayser et al. 2008; Maunsell and Gibson 1992). First, for each neuron, the peristimulus time histograms (PSTHs with 10-ms bins) were smoothed with a Savitzky-Golay (SG) filter with a third-degree underlying polynomial and a window size of five bins. We estimated the mean and the standard deviation of the baseline firing rate over the 170-ms period before the onset of the first click in the longer auditory condition. We established the onset latency by first locating the first three successive bins that exceeded the baseline rate by three standard deviations. The onset latency was then taken to be the middle time point of the first bin. To calculate the offset latency, we examined activity beginning 50 ms after physical motion offset (corresponding to 20 ms after the offset of the last click in the longer auditory

condition). We determined the first three successive bins that were within three standard deviations of the baseline rate. The offset latency was taken to be the middle time point of the third bin. Response durations were then computed as the difference between onset and offset latencies.

Additionally, we estimated the response duration of the averaged neuronal responses observed in the population PSTHs. To compute the population PSTHs, the raw PSTHs (10-ms bins) for each neuron were first normalized to the maximum response across all conditions. These normalized PSTHs were then averaged across those neurons having neuronal responses that showed significant dependency on time bin (i.e., were significantly modulated by the stimulus). This population PSTH was then smoothed with an SG filter with a third-degree underlying polynomial and a window size of five bins. The response durations in these averaged PSTHs were then estimated as described above. The SG filter, also known as a digital smoothing filter, has been commonly used to remove high-frequency ripples from averaged activity (e.g., Huang and Lisberger 2013; Huang et al. 2007). In our data analysis, we carefully arranged the width and degree of SG filter such that the filtering procedure preserved the location and shape of the averaged spiking activity. We also confirmed that the filtering procedure did not introduce a significant artifact in the estimates of response duration.

LFP dynamics. For each recording location, we computed the normalized LFP amplitudes within each of five frequency bands (see above). To do so, we first averaged all of the trial-by-trial LFPs for each experimental stimulus. We then filtered these average responses with a second-order notch filter (quality factor 10) with a 60-Hz center frequency. Notch filters have been commonly used to remove line noise in LFP signals (e.g., Wang et al. 2011). We adjusted the filter parameters in controlled and initial data analyses so that filtering was mostly restricted to frequencies ~ 60 Hz. Instantaneous amplitude was extracted by wavelet decomposition (Morlet wavelets) on 92 scales from 1 to 92 Hz. These wavelet-transformed data were normalized to the maximum amplitude within that session. The mean LFP amplitudes for each frequency band were then computed by averaging the wavelet-transformed normalized amplitudes within each of the five frequency ranges. For the lowest frequencies analyzed (e.g., 10 Hz), the wavelet procedure substantially temporally blurs the amplitude signal, which can result in low-frequency, i.e., θ (4–8 Hz) and α (8–12 Hz), responses that appear to commence before stimulus onset. Additionally, for the highest frequency range (γ_2 , 60–90 Hz), the effect of blurring over the frequency domain becomes dominant over the baseline power during the first 70 ms of the wavelet-transformed amplitude. To minimize these problems, we used frequency band-specific methods to estimate response durations (see below).

As we did for the spike data, we estimated response durations by subtracting the onset latency from the offset latency. For θ - and α -frequencies, we computed the onset and offset latencies using an approach similar to that used by Sundberg et al. (2012). We first estimated the midpoint of maximum and minimum LFP amplitudes [i.e., $0.5 \times (\text{maximum} - \text{minimum}) + \text{minimum}$]. We established the onset latency by locating the first 20 successive 1-ms bins that exceeded this midpoint estimate. The onset latency was taken to be the first of these bins. We estimated offset latency within the period beginning 50 ms after the visual stimulus offset (corresponding to 20 ms after the offset of the last click in the longer auditory condition). The offset latency was taken to be the first of the 20 successive 1-ms bins with values that fell below the midpoint value. For β -, γ_1 -, and γ_2 -frequency bands, we computed onset and offset latencies by using mean baseline amplitudes plus standard deviations as the reference point. To accurately estimate the amplitude and standard deviation of baseline activity, we used time windows that were outside the blurring effects that arise from wavelet transform (Chandran KS et al. 2016). The starting points of these windows were 200, 200, and 100 ms before visual motion onset for the β -, γ_1 -, and γ_2 -frequency bands,

respectively. The corresponding end points of these baseline time windows were 80, 50, and 30 ms before visual motion onset.

To compute response onset, we looked for the first 20 consecutive bins after the end of the baseline window that had values exceeding the baseline amplitude by at least 3 standard deviations. The onset was taken to be the first of these 20 consecutive points. Similarly, we looked for the response offset beginning 50 ms after visual motion offset with the offset taken to be the first of 20 consecutive points that lay within 3 standard deviations of the baseline amplitude.

Audiovisual Interactions

To determine whether LFPs to audiovisual stimuli reflected non-additive audiovisual interactions, responses to audiovisual (AV) stimuli were compared with the arithmetic sum of responses to auditory-only (A) stimuli and responses to visual-only (V) stimuli. We applied a bootstrap procedure using averaged normalized LFPs (see above) from the audiovisual, auditory-only, and visual-only stimuli from each recording session in which we included auditory-only conditions ($n = 38$). For each AV response having corresponding auditory-only and visual-only responses (i.e., 20-, 60-, 100-, and 160-ms visual ISI conditions), we generated a data set including all A and V pairwise combinations ($n = 38 \times 38$). We then randomly selected a sample from this data set and computed the summed response (A+V). This bootstrap procedure was repeated 10,000 times, yielding 10,000 summed unisensory response samples. From these bootstrap samples, the mean summed response and 99% confidence intervals were estimated for each data point in time. For each time point, we asked whether the actual bimodal AV responses were either significantly larger (superadditive) or smaller (subadditive) than the summed “unisensory” signals (A+V). This criterion was met if the AV responses were outside the range of the confidence intervals of the A+V response. Additionally, to assess superadditive and subadditive effects for each frequency band, we applied the same approach on the wavelet-transformed LFP amplitudes. Because of the nonlinearity of the wavelet transformation (Senkowski et al. 2007), unisensory responses were summed before the wavelet transformation. We also used this same general approach to compare P and AP LFP responses.

Analysis of Eye Movements

To address the possible influences of eye movements (which might be elicited by auditory stimuli) on LFP activity, we measured 1) mean eye position and 2) standard deviation of eye position for each trial of the three main audiovisual configurations (AV shorter, AV longer, and visual-only) for all visual ISIs. We then performed ANOVA and *t*-tests to determine whether the mean and standard deviation of eye position differed as a function of the presence of auditory stimuli, auditory timing, and visual ISI.

RESULTS

ISI Tuning

Spike tuning curves. We analyzed 152 neuronal recordings, including both single units ($n = 97$) and multiunits ($n = 55$). Receptive field eccentricities ranged from 0.50 to 11.51°, with a median eccentricity of 7.16°. As we found no consistent difference in the pattern of results from single-unit and multi-unit recordings, we henceforth refer to both as unit recordings. For each recording, we presented three-frame AM stimuli with varying visual ISIs (i.e., temporal separations between each flash) in the preferred (P) and antipreferred (AP) directions. Figure 2 shows responses from an example unit recording. For this neuron, the difference in response magnitudes for the P and AP directions (i.e., directional selectivity) was greatest for

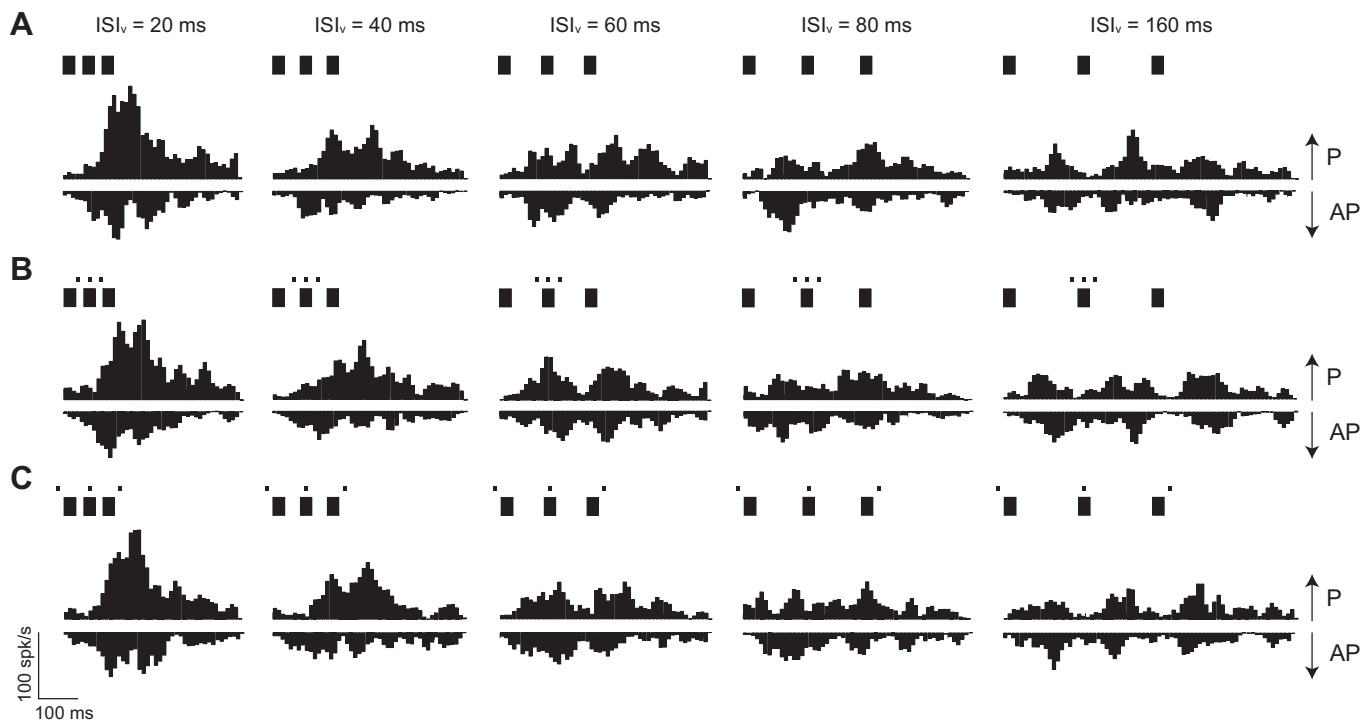


Fig. 2. Responses of a middle temporal area neuron. As in Fig. 1, the bars and clicks are indicated by large rectangles and small squares, respectively. Responses to motion in the preferred (P) and antipreferred (AP) directions are plotted upward and downward, respectively. *A*: responses to visual-only conditions. Peristimulus time histograms (bin width = 10 ms) show firing rate as a function of time for five of the eight visual interstimulus intervals (ISI_v s) tested. *B*: responses to audiovisual (AV) shorter stimulus. *C*: responses to AV longer stimulus. Here, spk, spikes.

the smallest visual ISI (ISI_v). As the ISI_v was increased, the average spiking rate decreased (mostly for the P direction). For $ISI_v > 80$ ms, the response magnitudes for the two directions converged with resultant loss of directional selectivity. This dependency on ISI_v is clearly seen in the ISI_v tuning curves estimated from the responses of the same neuron (Fig. 3). As discussed in MATERIALS AND METHODS, we did not look for significant ISI_v tuning for the P and AP tuning curves because of the response window confounder. The ANOVA test (visual ISI and audiovisual configuration as factors) applied to the P-AP tuning curves (Fig. 3, *right*) revealed that directional selectivity (P-AP) was significantly dependent on ISI_v

($F_{7,360} = 26.32$, $P < 0.0001$). Moreover, this example of P-AP tuning, like those from most recordings showing significant P-AP ISI_v tuning, exhibited low-pass tuning over the ISI_v range we examined (Fig. 3C). Note that because of the inverse relationship between ISI_v and speed, low-pass ISI_v tuning implies high-pass speed tuning and vice versa. The firing rate of this neuron was not, however, significantly modulated by the auditory stimuli: We observed no main effect ($F_{2,360} = 2.46$, $P = 0.0871$) of audiovisual configuration (i.e., neither the presence nor the timing of auditory stimuli had a significant effect on firing rate). And most important, there was no significant interaction between audiovisual configuration and

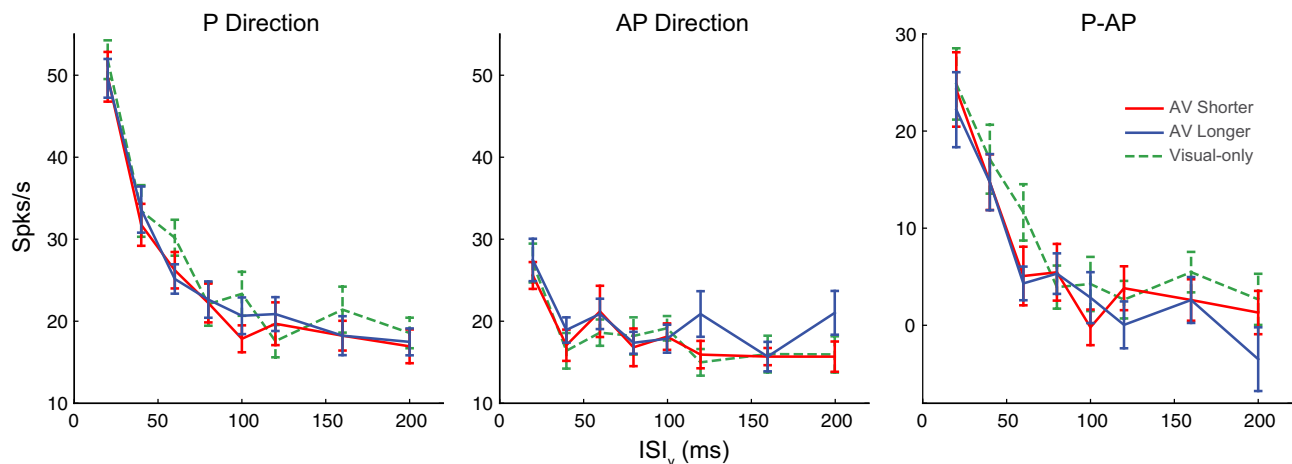


Fig. 3. Visual interstimulus interval (ISI_v) tuning curves for the neuronal responses shown in Fig. 2. Tuning curves for preferred (P) and antipreferred (AP) directions and the response difference between P and AP directions (P-AP) are shown in separate panels [P direction (*left*), AP direction (*middle*), and P-AP (*right*)]. The dashed green, solid red, and solid blue curves correspond to visual-only, audiovisual (AV) shorter, and AV longer conditions, respectively. Each data point is the mean of 16 trials, and error bars correspond to $\pm SE$. Spks, spikes.

Table 1. Two-way ANOVAs on the local field potential P-AP interstimulus interval tuning curves

	ISI _v	Audiovisual Configuration	ISI _v × Audiovisual Configuration
θ (4–8 Hz)	12	4 (2)	4 (1)
α (8–12 Hz)	5	3 (0)	3 (0)
β (13–30 Hz)	5	2 (1)	3 (1)
γ ₁ (30–60 Hz)	10	1 (0)	1 (0)
γ ₂ (60–90 Hz)	4	0 (0)	2 (1)

Visual interstimulus interval (ISI_v) and audiovisual configuration as factors. Each column shows the number of recording sessions having a significant dependency on ISI_v, auditory configuration, and/or a two-way interaction out of our 61 recording sessions. The values in parentheses in the second and third columns correspond to the number of recording sessions (out of just the significant cases in the first column) having significant dependency on auditory configuration and interaction with ISI_v, respectively. The number of significant recording sessions for each frequency band is indicated in separate rows. P-AP, response difference between preferred and antipreferred directions.

ISI_v for the P-AP ISI tuning curve ($F_{14,360} = 0.46$, $P = 0.9509$) and hence no evidence of an auditory-induced change in ISI_v tuning.

We found that 59 of our 152 unit recordings were significantly ISI_v tuned for P-AP (39%). Of the total population, 13 (~8.5%) units exhibited a significant main effect of audiovisual configuration and 9 (~6%) units exhibited 2-way interactions. We next asked whether the presence of auditory stimulation and/or auditory timing had a significant impact on response rate for the 59 units that exhibited significant P-AP ISI_v tuning. Only four of these units were significantly impacted by audiovisual configuration (~6.7%). Within this subset of neurons, there was only one unit recording (~1.6%) that exhibited a significant interaction between audiovisual configuration and ISI_v. These very low numbers of unit recordings and percentages indicate that auditory stimulation had no meaningful and consistent effect on ISI tuning or, more generally, on neuronal response magnitudes. On the basis of these findings, we reject the hypothesis that ISI_v tuning within MT is sensitive to the timing of accompanying auditory stimulus.

LFP tuning curves. As described in the following sections, all AM stimuli elicited significant LFP responses as observed in both the raw LFP and in the wavelet-transformed LFP amplitudes (over the frequency ranges used in this study). As with the unit recordings, we looked for significant interactions

between ISI_v and audiovisual configuration for the P-AP responses. For the LFP recordings, we did this for each of the five frequency bands examined. Unlike what was found for the unit recordings, LFP response magnitudes for P and AP directions were not statistically different for most cases. Therefore, the number of cases (broken into frequency bands) with significant direction tuning was low (Table 1). Moreover, these few significant cases did not have the typical (low-pass, band-pass, or high-pass) morphology associated with ISI tuning, consistent with the conclusion that these cases were just the spurious deviations expected by chance. We also looked at the number of recordings with significant effects of audiovisual configuration and two-way interaction out of this limited subset of recordings. Again, the number of significant cases was very small (Table 1).

Response Timing

Spiking dynamics. We next asked whether auditory timing modulated neuronal response timing in a manner consistent with the effects on perceived timing. We first tested the general hypothesis that auditory events influenced the dynamics of neuronal response in area MT. The testing of this hypothesis was restricted to recordings with significant responses to AM stimuli (average normalized responses of significant cases for 20-ms ISI_v are shown in Fig. 4). Across all ISI_vs (152 neurons × 8 ISI_v conditions), we found that 83% of the units responded significantly to the AM stimuli (see MATERIALS AND METHODS) for the P direction. The percentages of significant cases for AP and P-AP were 82 and 39%, respectively. To determine whether sounds modulated the dynamics of these significant responses, we looked for significant interactions between time and audiovisual configuration (AV shorter, AV longer, and visual-only). Of these units that responded significantly to the AM stimuli, the percentages with significant modulation ($P < 0.05$) were 16, 14, and 6% for P, AP, and P-AP, respectively. These percentages are greater than the 5% expected by chance and hence suggest that auditory timing influenced the response timing of some area MT neurons.

Given this evidence of an influence on neuronal response timing, we next tested the more specific hypothesis that the duration of neuronal responses for the shorter auditory timing was shorter than for the longer condition (Fig. 5). As shown by

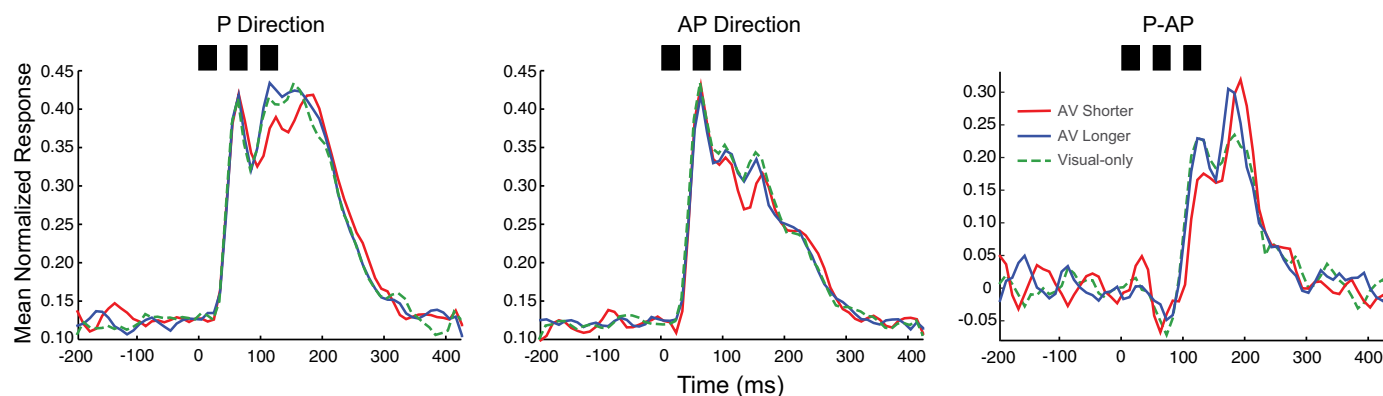


Fig. 4. Peristimulus time histograms (normalized and averaged over all unit recordings with significant responses) reveal robust responses to the 20-ms visual interstimulus interval condition. Plots show responses to preferred (P) and antipreferred (AP) directions and the difference between these responses (P-AP). The unit recordings for each plot (P, AP, and P-AP) were 134, 128, and 83, respectively. In each plot, the dashed green, solid red, and solid blue curves correspond to visual-only (V), audiovisual (AV) shorter, and AV longer conditions, respectively.

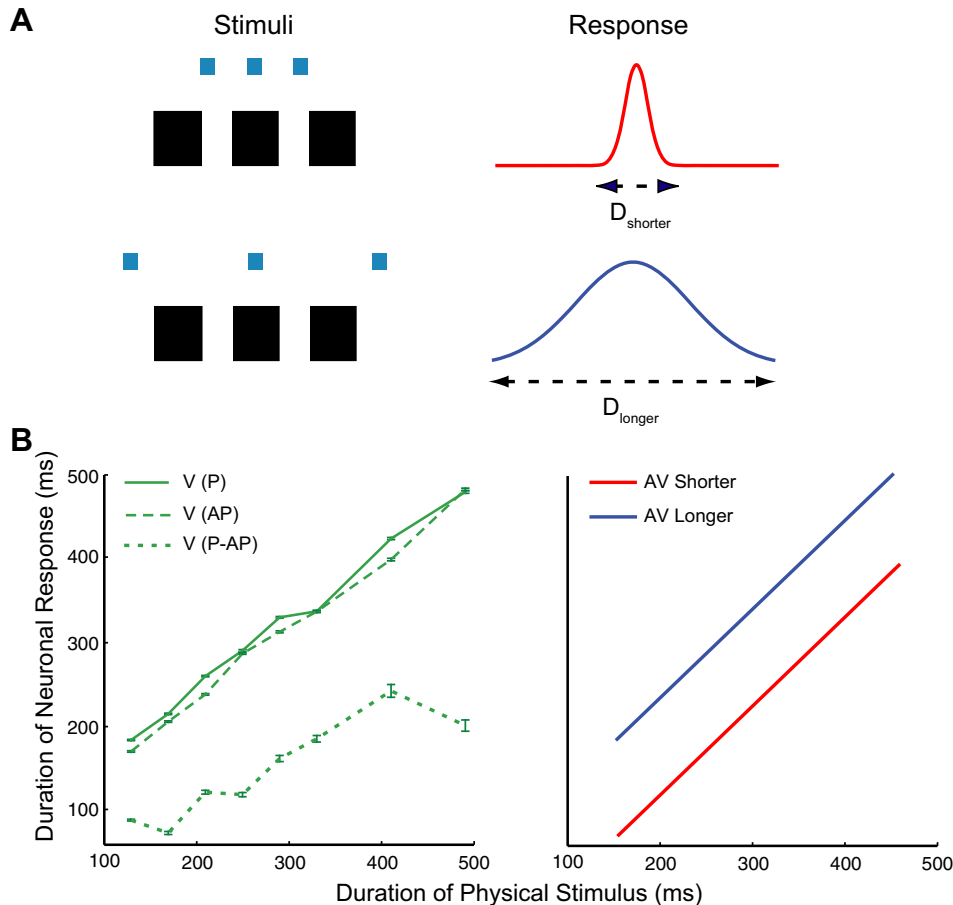


Fig. 5. *A*: schematic predictions of changes in response duration resulting from temporal ventriloquism. The timing of auditory events (blue squares) is predicted to shift the timing of neuronal responses to each visual event (black rectangles). If this is the case, the response durations for the longer auditory timing condition (D_{longer}) should be longer than those for the shorter auditory timing condition (D_{shorter}). *B*: duration of neuronal responses as a function of the physical duration of visual stimulus (apparent motion duration: $2 \times \text{ISI}_v + 90$ ms, where ISI_v is visual interstimulus interval). Green curves in the panel at left correspond to the average of estimated durations from each unit for the visual-only (V) conditions. Error bars correspond to $\pm \text{SE}$. Red and blue curves in the panel at right represent predicted durations for the audiovisual (AV) shorter and longer conditions, respectively. AP, antipreferred direction; P, preferred direction; P-AP, response difference between P and AP directions.

the population (i.e., for all units that responded significantly to the AM stimuli) PSTHs for the 20-ms ISI_v conditions (Fig. 4), the response duration of the shorter condition does not appear to be shorter than that for the longer condition. Our results are thus at odds with the predictions shown in Fig. 5*A*. Response durations for the shorter and longer auditory timings for all ISI_v conditions are shown in Fig. 6. These findings, similarly, do not conform to the prediction shown in Fig. 5*B*. Moreover, two-way ANOVA (visual ISI and auditory timing as factors)

did not reveal any significant effect of auditory timing on the duration of neuronal responses (P: $F_{1,1572} = 1.59$, $P = 0.2078$; AP: $F_{1,1563} = 1.10$, $P = 0.2951$; P-AP: $F_{1,554} = 0.12$, $P = 0.7265$). The interaction between visual ISI and auditory timing was also not significant (P: $F_{7,1572} = 1.16$, $P = 0.3205$; AP: $F_{7,1563} = 0.33$, $P = 0.9388$; P-AP: $F_{7,554} = 0.61$, $P = 0.7452$). We also performed the same statistical analyses on the response durations estimated from just those unit recordings that showed significant interactions between time and audio-

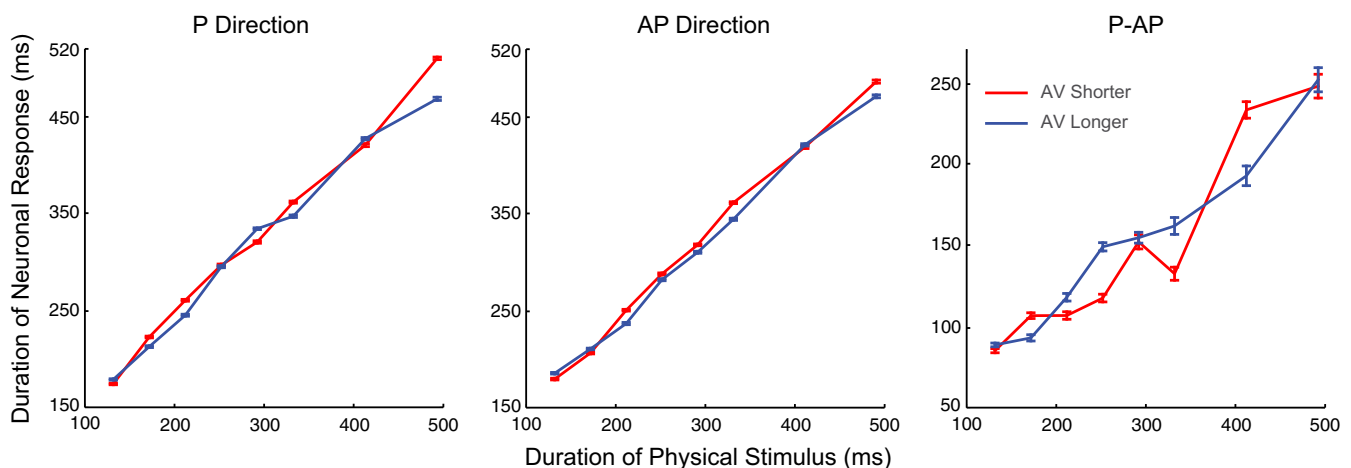


Fig. 6. Duration of neuronal responses as a function of apparent motion duration. The duration values are from the unit recordings with significant responses. Plots show estimated durations for responses to preferred (P) and antipreferred (AP) directions and the difference between these responses (P-AP) with red and blue curves corresponding to audiovisual (AV) shorter and longer conditions, respectively. Each data point is the average of the estimated durations from each single unit and multiunit. Error bars correspond to $\pm \text{SE}$.

visual configuration (P, 16%; AP, 14%; AP, 6%). The average durations of neuronal responses from this subset were similar to the ones in Fig. 6. The main effect of ISI_v was significant ($P < 0.05$). However, the ANOVA test did not reveal any significant effect of auditory timing (P: $F_{1,276} = 0.32$, $P = 0.5706$; AP: $F_{1,227} = 0.01$, $P = 0.9218$; P-AP: $F_{1,25} = 2.46$, $P = 0.1295$) or any two-way interactions (P: $F_{7,276} = 0.8$, $P = 0.5895$; AP: $F_{7,227} = 0.23$, $P = 0.978$; P-AP: $F_{7,25} = 0.72$, $P = 0.6598$).

We also considered the possibility that other aspects of the temporal dynamics of neuronal responses (i.e., other than duration) might be influenced by auditory timing and that such changes might be seen more clearly in the frequency domain rather than in the time domain. To investigate this possibility, we performed a power spectra analysis (Bair et al. 1994) on the neuronal responses and then looked for significant interactions between frequency and audiovisual configuration (AV shorter, AV longer, and visual-only). These tests are analogous to those we did in the time domain. Across all ISI_v s, the effects of frequency were found to be significant for 99.3% of the units in the P direction. The percentages of significant cases for AP and P-AP were 99.3 and 88.2%, respectively. These results simply confirmed that statistically significant neuronal responses to our visual stimuli can be observed in the frequency domain. Of these units that responded significantly to the AM stimuli on the basis of our power spectra analysis, we observed 7.3, 8.6, and 3.7% significant ($P < 0.05$) examples of auditory modulation (frequency $\times$ audiovisual configuration) for P, AP, and P-AP, respectively. These percentages are smaller than those observed in the time domain and provide little evidence

of modulation in the frequency domain. We thus have no evidence of any consistent auditory-induced change in the temporal dynamics of neuronal responses in area MT.

LFP dynamics. Figure 7 shows the mean normalized LFP activity (averaged across all recording sessions $n = 61$) for the 20- and 160-ms ISI_v conditions. Visual inspection of these results (together with the other ISI_v conditions) reveals that responses to the P and AP directions are not substantially different. Similarly, when we compare spectrograms for P and AP directions, we do not observe a salient difference between wavelet-transformed (i.e., spectral) LFP amplitudes for these directions across different frequencies (Fig. 8). These impressions are confirmed by a bootstrap analysis comparing P and AP directions for all visual $ISIs$ and audiovisual configurations. In these analyses, we used the same parameters and significance criteria used to test for audiovisual interactions, as described in MATERIALS AND METHODS. These analyses did not reveal a consistent difference between P and AP directions in either the raw LFP or the wavelet-transformed spectrograms. AM stimuli presented in both P and AP directions elicited a clear power increase for frequencies lower than 30 Hz. Although power changes for frequencies higher than 30 Hz are not as obvious in the spectrograms (Fig. 8), another bootstrap analysis (looking for significant deviation from average baseline power) revealed that power in γ low- and high-frequency ranges (γ_1 and γ_2 , respectively) was significantly increased after the presentation of AM stimuli moving in both directions (Fig. 9). We additionally tested whether the response duration (averaged and normalized) within each frequency band significantly depended on the duration of the visual motion. As

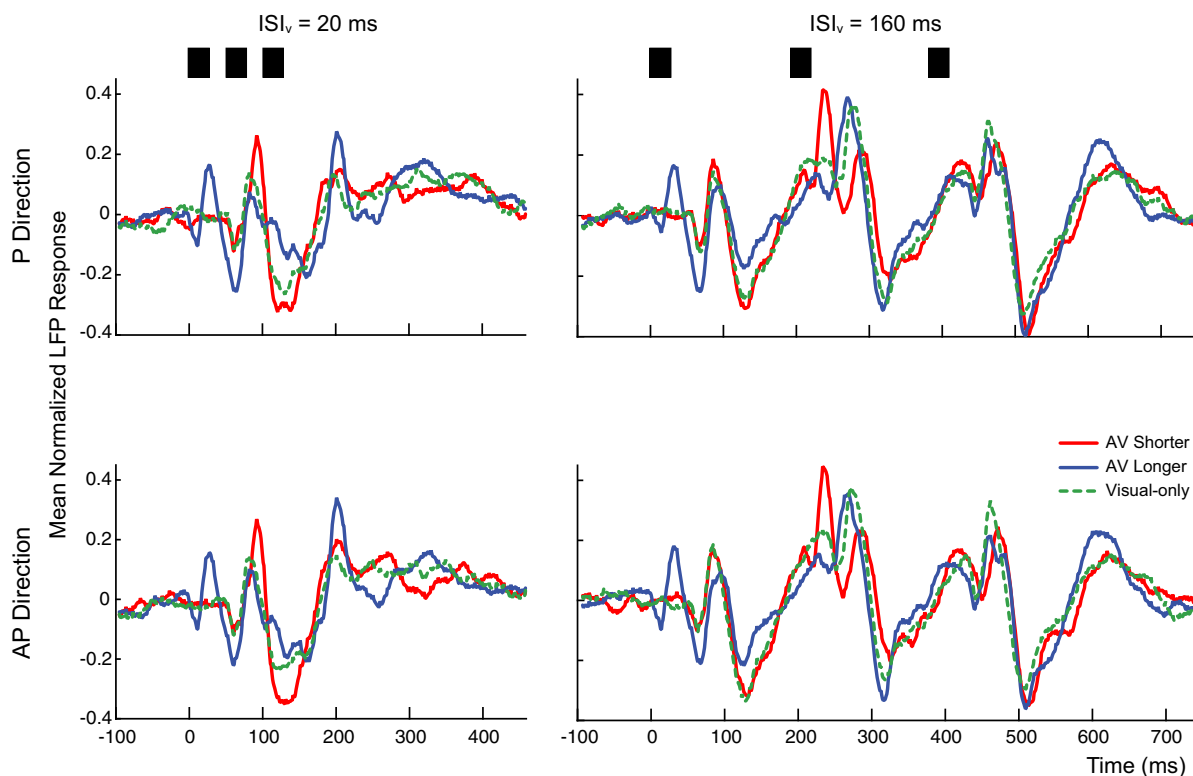


Fig. 7. Normalized and averaged raw local field potential (LFP) responses ($n = 61$) for two of the eight visual interstimulus intervals (ISI_v s) tested. The dashed green, solid red, and solid blue curves correspond to visual-only, audiovisual (AV) shorter, and AV longer conditions, respectively. The onset of the apparent motion is at 0 ms. The timing and duration of each visual bar are indicated by the position and thickness, respectively, of each black rectangle shown above the panels at top. Top: preferred (P) direction. Bottom: antipreferred (AP) direction.

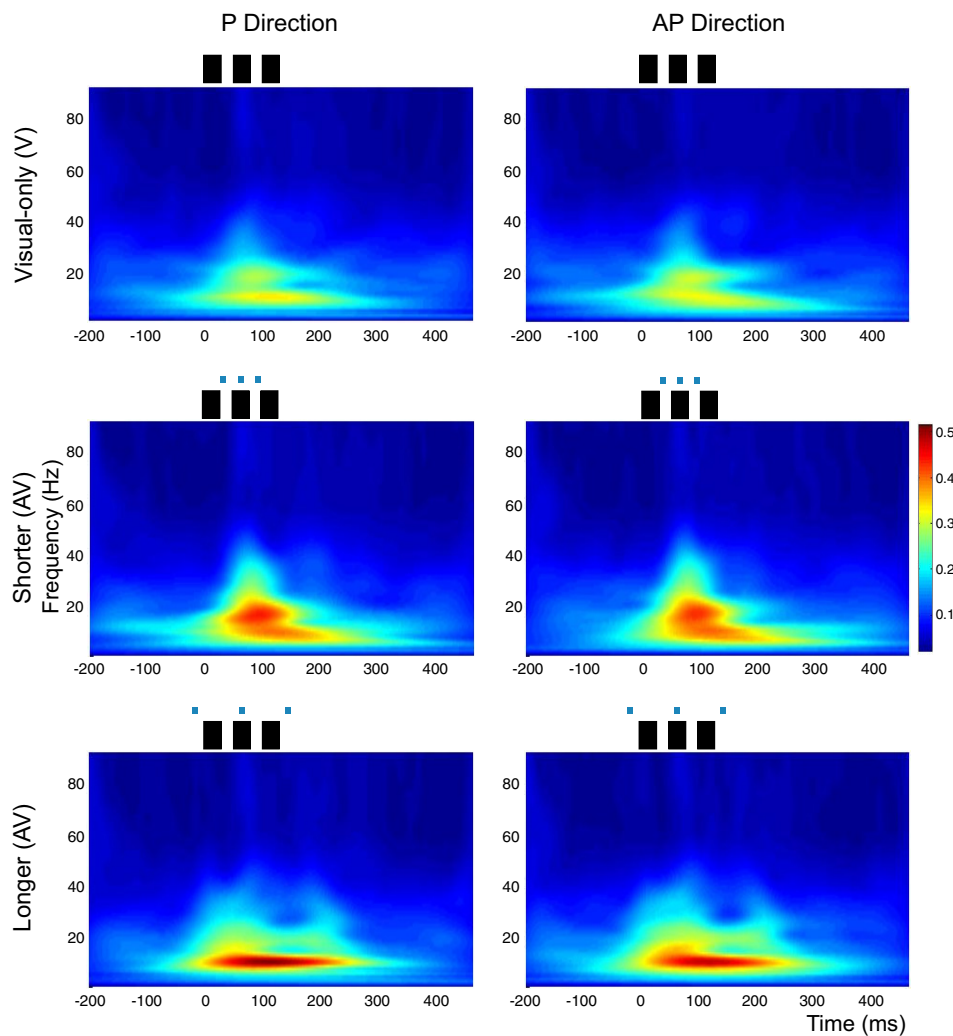


Fig. 8. Normalized and averaged spectrograms ($n = 61$) for 20-ms visual interstimulus intervals. Each time-frequency plot corresponds to wavelet-transformed local field potential amplitudes for a specific motion direction [preferred (P) direction (*left*) and antipreferred (AP) direction (*right*)] and auditory condition [visual-only (V, *top*), audiovisual (AV) shorter (*middle*), and AV longer (*bottom*)]. The black rectangles and blue squares at the top of each time-frequency plot represent visual bars and clicks, respectively. The timing and duration of these stimuli are indicated by the position and thickness of these icons, respectively.

revealed by one-way ANOVA (ISI_v as factor) applied to the visual-only condition, the duration of LFP spectral amplitudes significantly varied with the duration of the physical stimulus for all frequency ranges examined (P and AP directions $P < 0.0001$).

The spectrograms for the two auditory timing conditions of 20-ms ISI_v are also shown in Fig. 8. The differences between the two auditory timing conditions can be clearly seen for frequencies lower than 30 Hz. We observed a reduction in the response duration for the shorter audiovisual condition relative to the longer condition: the response to the longer condition is relatively longer with a bias toward lower frequencies (Fig. 9). These changes are in agreement with the hypothesis that auditory timing can capture the timing of visual responses (Fig. 5). To quantify these changes in LFP dynamics, we estimated the response durations for all ISI_v s and auditory timing conditions used in our recording sessions (Fig. 10). For the β - and γ -frequency bands, we consistently found that LFP response durations for the longer auditory timing were longer than for the shorter auditory timing. This effect of auditory timing was found to be significant (Table 2). Auditory timing had a similar influence on the response durations in the α -frequency band when the ISI_v was 20 ms (Fig. 9), but we did not observe a consistent effect for α -band LFP response durations computed in other ISI_v conditions.

Audiovisual Interactions

Unlike what was found for spiking activity, the durations of the LFP responses (i.e., the duration of spectral amplitudes in β - and γ -frequency bands) to the short versus long audiovisual conditions were significantly different. Although these differences were seemingly in agreement with the auditory capture of visual response timing, we also observed clear LFP responses (unlike the case for spiking responses, Fig. 11A) to auditory clicks presented in the absence of the visual stimuli (Fig. 11B); hence this variation in response duration might correspond to variation in the duration of the auditory response rather than variation in the duration of the visual responses. To determine whether the differences between the LFP responses to the short and long audiovisual conditions were simply due to the addition of the auditory-evoked responses or whether these differences reflected nonlinear audiovisual interactions, we compared the raw LFP responses to combined stimulation (AV) with the summation of the unisensory responses (i.e., AV vs. A+V). A significant difference between AV and A+V indicates a nonlinear interaction (superadditive or subadditive) between the unisensory processes (Kayser et al. 2008; Mercier et al. 2013; Molholm et al. 2002; Stanford et al. 2005). Our bootstrap analyses did not reveal any significant deviation of the audiovisual response from the summed unimodal responses

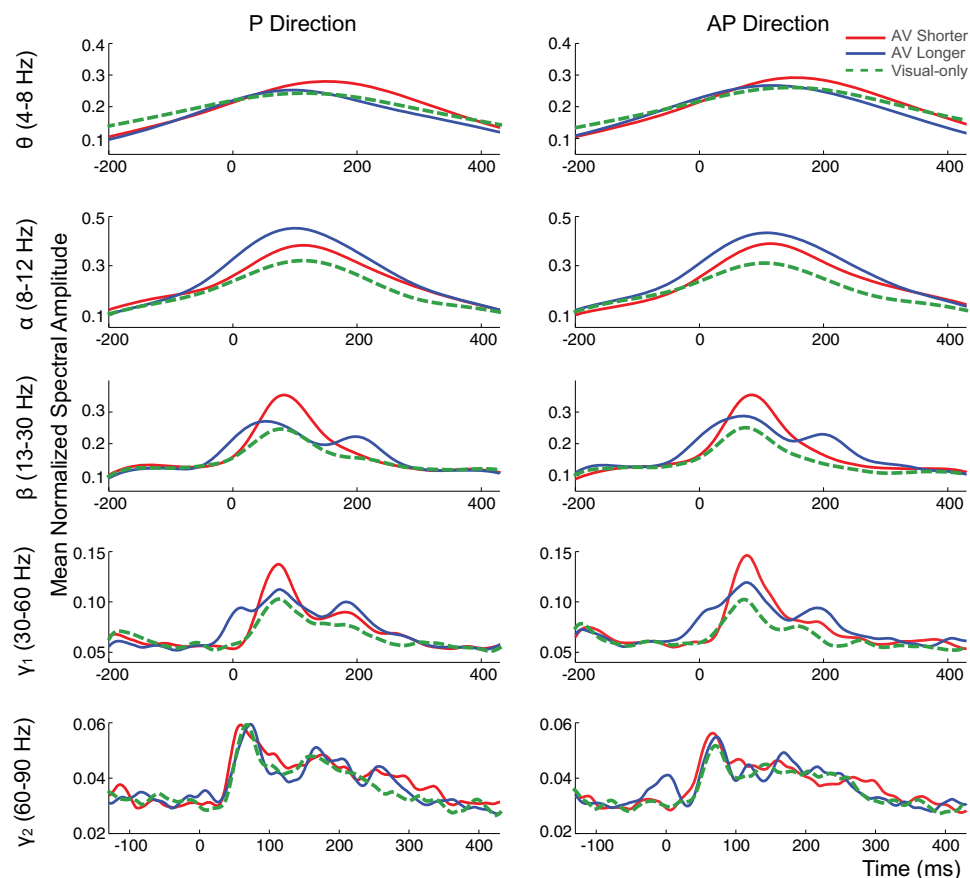


Fig. 9. Mean wavelet transformed local field potential amplitudes ($n = 61$) in different frequency bands (top to bottom) for 20-ms visual interstimulus interval conditions. The mean spectral amplitudes for each motion direction are shown in separate columns [preferred (P) direction (left) and anti-preferred (AP) direction (right)]. The dashed green, and solid red, and solid blue curves in each plot represent mean amplitudes for the visual-only, audiovisual (AV) shorter, and AV longer conditions, respectively.

for any specific time window or for any of the different visual ISI conditions examined. Moreover, neither did the bootstrap analysis applied to the wavelet-transformed LFP amplitudes for each frequency band reveal superadditive or subadditive effects consistent across all ISI_v conditions. On the basis of these analyses, we cannot reject the conclusion that the observed differences in LFP dynamics (and the corresponding changes in the durations of β and γ low-frequency amplitudes) simply reflect the additive superimposition of auditory and visual responses.

Control Recordings on Auditory LFPs

We also considered the possibility that auditory-evoked LFP responses might be due to physiological sources distant to MT. One candidate for a non-MT cellular source for the LFP is the “postauricular muscle response” (PAMR), which is electrical activity evoked in the muscle located just behind the ear (Benning 2011; O’Beirne and Patuzzi 1999). The PAMR can be seen in EEG recordings (McDonald et al. 2013) but is not seen in intracranial recordings (Mercier et al. 2013). Nevertheless, it is conceivable that the PAMR might have contaminated our recordings. To test this possibility, we performed a set of control recordings that more generally tested for sources distant from area MT, including physiological sources such as the PAMR, as well as nonphysiological electrical sources (see DISCUSSION). In these control experiments, we recorded electrical activity 1 mm below the dural surface (Fig. 12). Unlike electrical activity generated within (or near) MT (typically ~10 mm below the dural surface), electrical artifacts and the PAMR (as well as other physiological sources with widespread detect-

ability) should be observed outside of MT. We found no hint of significant modulation outside of area MT.

DISCUSSION

Consistent with earlier studies using similar visual AM stimuli, we found that the average firing rate of area MT neurons was tuned for the visual interstimulus interval (ISI_v) between the individual frames of AM stimuli. Since ISI_v and speed covaried in these stimuli, this result is also consistent with speed tuning. We tested the hypothesis that this rate-based tuning reflected the perceived ISI_v /speed as modulated by accompanying brief auditory events (“clicks”), which have been shown to perceptually capture the timing of visual stimuli in human psychophysical experiments. Contrary to that hypothesis, we did not find that the addition of clicks or their timing had a significant impact on ISI_v tuning. Assuming that the monkeys in our experiments perceived the temporal ventriloquism illusion (see below), our findings suggest that (within the context of the audiovisual interactions examined here) ISI_v tuning in area MT reflects not the perceived ISI_v but rather the physical ISI between AM frames.

Additionally, we tested the hypothesis that the duration (i.e., from neuronal onset to offset) of spiking responses reflected the perceived rather than the physical timing of visual stimuli. Specifically, we predicted that the shorter and longer audiovisual conditions would result in shorter and longer response durations, respectively. Whereas we found that sounds had a significant (i.e., at an above-chance level), if weak, influence on the dynamics of some neurons, we did not find a significant

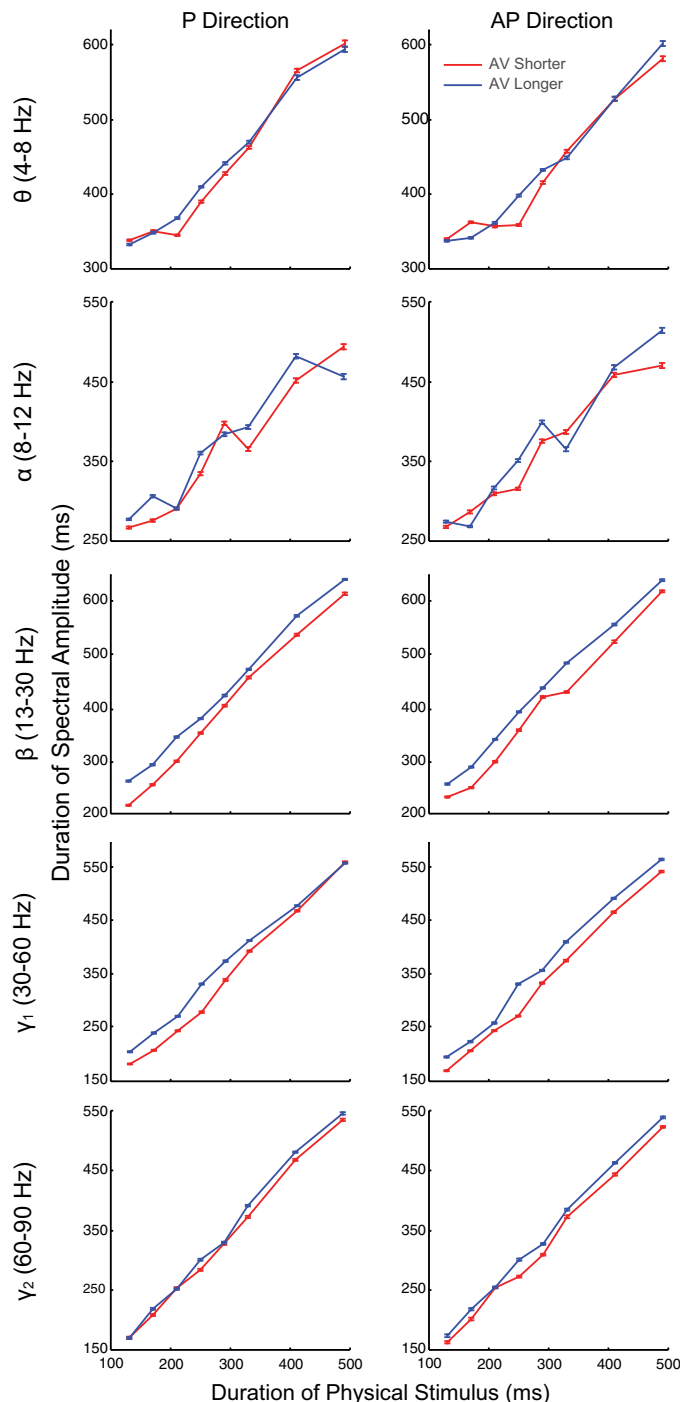


Fig. 10. Durations of frequency band-specific local field potential amplitudes as a function of apparent motion duration. Each plot corresponds to estimated durations for the preferred (P) or antipreferred (AP) motion direction [P direction (*left*) and AP direction (*right*)] within each frequency band (*top* to *bottom*). The red and blue curves in each plot represent mean amplitudes for the audiovisual (AV) shorter and AV longer conditions, respectively. Each data point is the average of the estimated durations from each recording session. Error bars correspond to $\pm$ SE.

difference in the durations of neuronal responses to the shorter versus longer audiovisual conditions.

Unlike the case for spiking activity, we found robust duration differences between the LFP responses to the shorter and longer audiovisual stimuli. Moreover, these duration differ-

ences matched the differences in perceptual duration implied by human psychophysical experiments. We also observed, however, the same pattern of LFP response duration differences to the auditory stimuli presented alone. Moreover, on average, LFP responses were consistent with the linear summation of auditory and visual responses. Our findings thus argue against the hypothesis that auditory timing altered the duration of visual responses.

Area MT and Speed Perception

There is abundant evidence that cortical area MT is involved in speed perception in both human and nonhuman primates. In humans, MT+ (the presumed human homolog of area MT and MST) is preferentially activated during speed discrimination tasks (Huk and Heeger 2000). In nonhuman primates, area MT neurons have been found to be speed tuned (Maunsell and Van Essen 1983; Perrone and Thiele 2001). Moreover, speed discrimination is impaired in nonhuman primates in which area MT has been lesioned (Newsome and Paré 1988; Orban et al. 1995; Rudolph and Pasternak 1999), and microstimulation of area MT alters speed perception (Liu and Newsome 2005). Of relevance to the present study, visual stimulus manipulations (e.g., of contrast, stimulus size, and spatial frequency) affecting the perception of speed in human subjects have been found to also modulate the responses of neurons in area MT of passively fixating macaques (Boyraz and Treue 2011; Priebe and Lisberger 2004). On the basis of these types of findings, it is a matter of continuing debate as to whether simple decoding schemes of area MT responses (such as labeled lines) are sufficient to account for the perception of speed (Brooks et al. 2011; Krekelberg et al. 2006a, 2006b; Priebe and Lisberger 2004; Thompson 1982).

Whereas early characterizations of direction and speed selectivity in area MT used visual stimuli that moved continuously (Albright 1984; Dubner and Zeki 1971; Maunsell and Van Essen 1983), more recent studies typically used visual monitors in which real motion is mimicked by successive changes in the luminance of stationary pixels. The motion produced this way is a type of AM whereby successively activated spatially separated stimuli yield the illusion of motion (Kolars 1972; Korte 1915). If the spatial and temporal intervals are small enough, however, AM cannot be readily distinguished from continuous motion (Watson et al. 1986). Consistent with this perceptual equivalence, the results of neurophysiological studies of directional tuning in area MT that have used monitors to elicit an illusion of continuous motion agree (e.g., Stoner and Albright 1992) with the earlier studies that used “real” or continuous motion (e.g., Albright 1984). AM stimuli with larger spatial and temporal intervals, such as used in our present study, although eliciting a clear sense of visual motion, can, however, be readily distinguished from continuous motion. Studies before ours have found that area MT neurons are direction and speed tuned for these types of AM stimuli (Mikami 1991; Mikami et al. 1986a). For stimuli with a fixed spatial jump (such as in our study), speed is determined by the ISI, with smaller ISIs corresponding to higher speeds.

As mentioned above, it has previously been found that visual stimulus manipulations that impact the perception of speed in humans also impact the speed tuning of area MT neurons in passively fixating monkeys. Our present study can be viewed

Table 2. Two-way ANOVAs on the duration of response windows shown in Fig. 10

	P Direction		AP Direction	
	Auditory Timing	ISI _v × Auditory Timing	Auditory Timing	ISI _v × Auditory Timing
θ (4–8 Hz)	1.73, 0.1885	0.41, 0.8943	1.49, 0.2233	0.95, 0.4676
α (8–12 Hz)	0.73, 0.3939	1.09, 0.3669	1.10, 0.2945	0.96, 0.4618
β (13–30 Hz)	43.50, < 0.0001*	0.71, 0.6599	51.35, < 0.0001*	0.89, 0.5103
γ ₁ (30–60 Hz)	44.10, < 0.0001*	2.36, 0.0218*	51.95, < 0.0001*	1.89, 0.0697
γ ₂ (60–90 Hz)	43.52, < 0.0001*	0.71, 0.6599	11.12, 0.0009*	0.44, 0.8774

Visual interstimulus intervals (ISI_v) and auditory timing [audiovisual (AV) shorter vs. AV longer] as factors. The numbers in each row and column correspond to *F* and *P* values for a specific frequency range and direction. The main effect of ISI_v was found to be significant for all frequency ranges and directions examined (all *P* < 0.0001). AP, antipreferred; P, preferred. *Significant (*P* < 0.05).

as an extension of those studies, except that we asked whether auditory manipulations that affect the perception of speed in humans similarly impact MT responses. Moreover, unlike in those previous studies, we used AM stimuli in which changes in the perception of speed may be determined by changes in perceived ISI between visual frames.

Do Monkeys Perceive the Illusion?

Previous studies demonstrated that humans systematically misperceive the speed and/or ISIs of audiovisual stimuli like those used in the present study (Freeman and Driver 2008; Kafaligonul and Stoner 2010). Moreover, using different paradigms (e.g., spatial ventriloquism), previous studies have also shown that monkeys experience audiovisual illusions similar to those of humans and suggest that cross-modal interactions are a basic aspect of the central nervous system and perceptual processing (Bremen et al. 2017; Kopčo et al. 2009; Woods and Recanzone 2004). It is possible of course that the monkeys in our experiments were insensitive to the temporal ventriloquism illusion used in our study. One candidate reason for such insensitivity is the difference in attentional requirements between the human psychophysical experiments and the nonhu-

man neurophysiological experiments: human subjects were required to attend to the audiovisual stimuli (to make speed or temporal judgments), whereas monkeys were only required to fixate. Although we cannot rule out that possibility, there are several reasons for believing that attention may not be required for this illusion. First, robust neuronal audiovisual interactions are found in the absence of attention (Meredith et al. 1987). Second, attention has been shown to have little effect on multisensory integration of audiovisual stimuli that are temporally discrete, such as those in the present study (Donohue et al. 2015). Third, and of direct relevance to the present study, Freeman and Driver (2008) explicitly tested and rejected the hypothesis that this type of audiovisual motion illusion works by attracting attention to specific visual intervals. Fourth, Kafaligonul and Stoner (2012) have found evidence that auditory timing impacts the processing of AM stimuli at the early “motion-energy” stage rather than via modification of attentional tracking. Nevertheless, although attention may not be required for audiovisual interactions to occur (Donohue et al. 2015), there is evidence that attention can have a modulatory influence on certain types of audiovisual interactions in space and time (Chen and Vroomen 2013). Verification of the per-

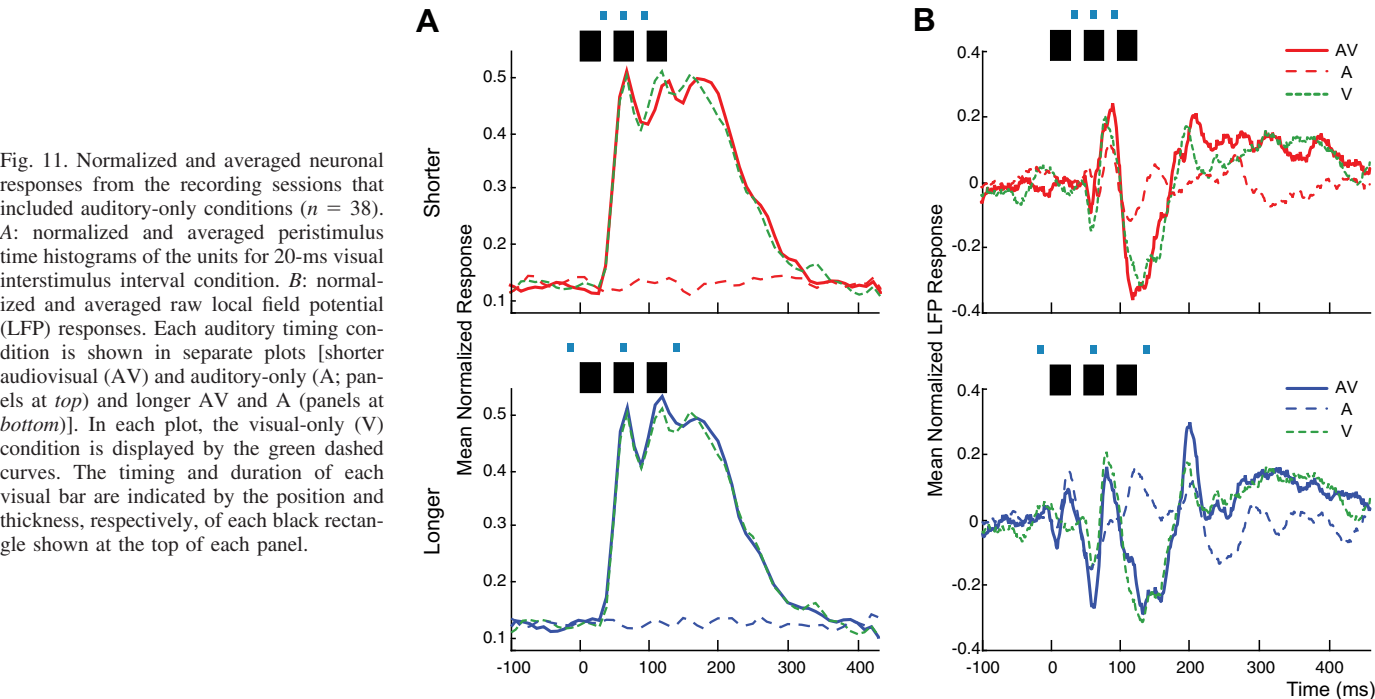


Fig. 11. Normalized and averaged neuronal responses from the recording sessions that included auditory-only conditions (*n* = 38). *A*: normalized and averaged peristimulus time histograms of the units for 20-ms visual interstimulus interval condition. *B*: normalized and averaged raw local field potential (LFP) responses. Each auditory timing condition is shown in separate plots [shorter audiovisual (AV) and auditory-only (A; panels at top) and longer AV and A (panels at bottom)]. In each plot, the visual-only (V) condition is displayed by the green dashed curves. The timing and duration of each visual bar are indicated by the position and thickness, respectively, of each black rectangle shown at the top of each panel.

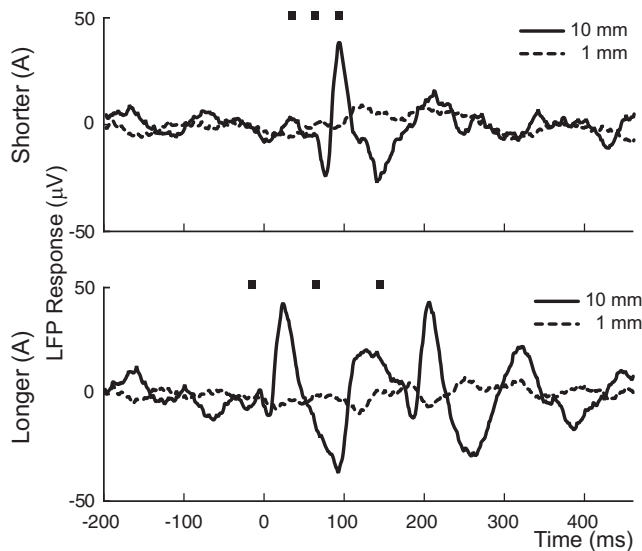


Fig. 12. Local field potential (LFP) responses from a control recording session. In this control session, only auditory events [auditory-only (A) shorter and longer conditions] were used. The timing and duration of each click are indicated by the position and thickness, respectively, of each square shown at the top of each panel. Raw LFP responses are the average of 32 presentations. The dashed and solid curves correspond to 1- and 10-mm recording depths, respectively.

ceptual effects in macaques during simultaneous neuronal recording is required to settle this issue with certainty.

Possible Artifactual Origin of Auditory LFPs

Given the dissociation between spiking activity and the LFP in response to sounds, we considered the possibility that the LFPs elicited by the sounds in our experiments might have a source other than area MT. We considered three possible alternative sources: 1) electrical sources (from the speakers producing the click), 2) physiological sources distant to MT, and 3) changes in visual input secondary to eye movements elicited by the clicks.

The possibility that the auditory LFP might be an electrical artifact arising from the speakers producing the clicks is strongly argued against by the timing of the auditory LFP, which had a latency of 40–60 ms relative to the onset of clicks but would have been nearly instantaneous if it were an electrical artifact. Furthermore, our control experiments ruled out electrical artifacts and other physiological sources with widespread detectability such as PAMR. On the other hand, we cannot definitively rule out the possibility that the LFP originated, in part, from adjacent cortical areas or at the very least from different cortical layers than the spiking activity (Kajikawa and Schroeder 2011). For instance, these auditory LFPs may have conceivably arisen from adjacent areas such as area MST, or even more distant auditory areas within the superior temporal gyrus.

We also considered a possible role for eye movements. Specifically, it is conceivable that the clicks resulted in small (i.e., within our fixation window) eye movements that resulted in small changes in the visual input. There are several arguments against this interpretation. First, the latencies of the auditory LFPs were too short to be a result of a change in visual input due to eye movements. We would expect a minimum saccade latency of 70 ms [for “express” saccades, see Fischer

and Boch (1983)] plus the latency of the visual responses to these stimuli (typically over 50 ms to these stimuli, Fig. 1), the sum of which far exceeds the LFP latencies of 40–60 ms. Second, we would have expected spiking responses to be similarly modulated, whereas significant modulation of spiking activity was only seen in a small number of recordings and this modulation did not parallel that seen in the LFP. Third, we found no consistent and meaningful difference in the mean or standard deviation of eye position between the different audiovisual configuration conditions.

Dissociation Between Spiking and LFP Responses

Although we found only very weak modulation of area MT spiking activity by sounds, we found robust auditory LFPs. Moreover, the weak modulation that we did observe in spiking activity did not parallel that seen in the LFP. As discussed above, our control experiments ruled out artifactual electrical sources as well as physiological sources distant from area MT. Dissociations between LFP and spiking activity are now well established (Denker et al. 2011; Lee et al. 2005; Maier et al. 2008; Mendoza-Halliday et al. 2014). Within the primary visual cortex, the relationship between spiking activity and the LFP has, for example, been found to be context dependent (Maier et al. 2008). Within area MT, Mendoza-Halliday et al. (2014) found that the LFP, but not spiking activity, was linked with short-term memory. One way to reconcile these discrepancies follows from the evidence that the LFP reflects not changes in the average firing rate but changes in neuronal synchrony (Denker et al. 2011). The auditory LFPs we observed might thus reflect auditory-induced changes in synchrony, which might, in turn, underlie changes in the perceived timing of the visual stimuli.

In the present study, another interesting dissociation between spiking and LFP activities was in directional selectivity. For AM with visual ISIs <100 ms, most MT recordings exhibited a significant difference in response magnitude between P and AP directions. However, this was not the case for the LFP. The LFP responses to our AM stimuli were robust, but there was no difference between P and AP directions across the different frequency ranges examined here. Conversely, Liu and Newsome (2006), using more salient and longer-duration motion stimuli (random dots with fully coherent motion), found that LFP activity can be tuned to speed especially for frequencies at and above the γ -band. The absence of LFP directional selectivity in our study is plausibly due to the use of weaker, and much briefer, stimulation (3-frame AM). Another potential factor in the differences in our two sets of findings, is that we, unlike Liu and Newsome, used a hardware band-pass filter (3–90 Hz), which may have resulted in a relative reduction in the γ -band power, within which directional selectivity was most apparent in Liu and Newsome’s study.

Although plausible rationales, as well as precedent, for dissociation of spiking and LFP activity can be advanced, we cannot rule out the possibility (as suggested above) that the observed dissociations arose because of the LFP having, in part, a nonlocal origin (Kajikawa and Schroeder 2011).

Auditory Evoked Potentials over Visual Cortex

Our findings indicate that auditory-alone stimulation evoked LFPs within (or nearby, if we allow for spreading of the signal

from nearby areas) visual area MT. These findings are in general agreement with neuroimaging studies that have found auditory influences in human area MT+. Although most such studies of audiovisual interactions have emphasized the modulation of visual responses in visual cortex by auditory stimuli, there is recent evidence of direct activation of visual cortex by auditory stimuli. Both EEG (Naue et al. 2011) and subdural (Mercier et al. 2013) recordings from human visual cortex revealed auditory evoked potentials over visual cortex. In agreement with the findings reported here, these studies found that brief auditory stimulation resulted in a power increase in mostly the β - and early γ -frequency ranges. Together with recent evidence, our study supports the notion that auditory stimulation may evoke activity within cortical areas that have been characterized as exclusively visual.

ACKNOWLEDGMENTS

We thank Aaron Cortez, Dinh D. Diep, and Jorge Aldana for superb technical assistance.

GRANTS

This research was supported by National Eye Institute (NEI) Grant R01-EY-012872 with core support provided by NEI Grant for Vision Research P30-EY-019005. H. Kafaligonul was also supported by Scientific and Technological Research Council of Turkey (TÜBİTAK) Co-Funded Brain Circulation Fellowship 112C010.

DISCLOSURES

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

AUTHOR CONTRIBUTIONS

H.K., T.D.A., and G.R.S. conceived and designed research; H.K. performed experiments; H.K. and G.R.S. analyzed data; H.K., T.D.A., and G.R.S. interpreted results of experiments; H.K. and G.R.S. prepared figures; H.K. and G.R.S. drafted manuscript; H.K. and G.R.S. edited and revised manuscript; H.K., T.D.A., and G.R.S. approved final version of manuscript.

REFERENCES

- Albright TD. Direction and orientation selectivity of neurons in visual area MT of the macaque. *J Neurophysiol* 52: 1106–1130, 1984. doi:10.1152/jn.1984.52.6.1106.
- Alink A, Singer W, Muckli L. Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. *J Neurosci* 28: 2690–2697, 2008. doi:10.1523/JNEUROSCI.2980-07.2008.
- Bair W, Koch C, Newsome W, Britten K. Power spectrum analysis of bursting cells in area MT in the behaving monkey. *J Neurosci* 14: 2870–2892, 1994. doi:10.1523/JNEUROSCI.14-05-02870.1994.
- Benning SD. Postauricular and superior auricular reflex modulation during emotional pictures and sounds. *Psychophysiology* 48: 410–414, 2011. doi:10.1111/j.1469-8986.2010.01071.x.
- Boyraz P, Treue S. Misperceptions of speed are accounted for by the responses of neurons in macaque cortical area MT. *J Neurophysiol* 105: 1199–1211, 2011. doi:10.1152/jn.00213.2010.
- Bremen P, Massoudi R, Van Wanrooij MM, Van Opstal AJ. Audio-visual integration in a redundant target paradigm: a comparison between rhesus macaque and man. *Front Syst Neurosci* 11: 89, 2017. doi:10.3389/fnsys.2017.00089.
- Brooks KR, Morris T, Thompson P. Contrast and stimulus complexity moderate the relationship between spatial frequency and perceived speed: implications for MT models of speed perception. *J Vis* 11: 19, 2011. doi:10.1167/11.14.19.
- Bueti D, Bahrami B, Walsh V. Sensory and association cortex in time perception. *J Cogn Neurosci* 20: 1054–1062, 2008. doi:10.1162/jocn.2008.20060.
- Buzsáki G. *Rhythms of the Brain*. New York: Oxford University Press, 2006. doi:10.1093/acprof:oso/9780195301069.001.0001.
- Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS. Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport* 10: 2619–2623, 1999. doi:10.1097/00001756-199908200-00033.
- Chandran KS S, Mishra A, Shihhatti V, Ray S. Comparison of matching pursuit algorithm with other signal processing techniques for computation of the time-frequency power spectrum of brain signals. *J Neurosci* 36: 3399–3408, 2016. doi:10.1523/JNEUROSCI.3633-15.2016.
- Chen L, Vroomen J. Intersensory binding across space and time: a tutorial review. *Atten Percept Psychophys* 75: 790–811, 2013. doi:10.3758/s13414-013-0475-4.
- Churchland MM, Priebe NJ, Lisberger SG. Comparison of the spatial limits on direction selectivity in visual areas MT and V1. *J Neurophysiol* 93: 1235–1245, 2005. doi:10.1152/jn.00767.2004.
- Denker M, Roux S, Lindén H, Diesmann M, Riehle A, Grün S. The local field potential reflects surplus spike synchrony. *Cereb Cortex* 21: 2681–2695, 2011. doi:10.1093/cercor/bhr040.
- Donohue SE, Green JJ, Woldorff MG. The effects of attention on the temporal integration of multisensory stimuli. *Front Integr Neurosci* 9: 32, 2015. doi:10.3389/fnint.2015.00032.
- Driver J, Noesselt T. Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron* 57: 11–23, 2008. doi:10.1016/j.neuron.2007.12.013.
- Dubner R, Zeki SM. Response properties and receptive fields of cells in an anatomically defined region of the superior temporal sulcus in the monkey. *Brain Res* 35: 528–532, 1971. doi:10.1016/0006-8993(71)90494-X.
- Duncan RO, Albright TD, Stoner GR. Occlusion and the interpretation of visual motion: perceptual and neuronal effects of context. *J Neurosci* 20: 5885–5897, 2000. doi:10.1523/JNEUROSCI.20-15-05885.2000.
- Fendrich R, Corballis PM. The temporal cross-capture of audition and vision. *Percept Psychophys* 63: 719–725, 2001. doi:10.3758/BF03194432.
- Fischer B, Boch R. Saccadic eye movements after extremely short reaction times in the monkey. *Brain Res* 260: 21–26, 1983. doi:10.1016/0006-8993(83)90760-6.
- Freeman E, Driver J. Direction of visual apparent motion driven solely by timing of a static sound. *Curr Biol* 18: 1262–1266, 2008. doi:10.1016/j.cub.2008.07.066.
- Getzmann S. The effect of brief auditory stimuli on visual apparent motion. *Perception* 36: 1089–1103, 2007. doi:10.1068/p5741.
- Huang X, Albright TD, Stoner GR. Adaptive surround modulation in cortical area MT. *Neuron* 53: 761–770, 2007. doi:10.1016/j.neuron.2007.01.032.
- Huang X, Albright TD, Stoner GR. Stimulus dependency and mechanisms of surround modulation in cortical area MT. *J Neurosci* 28: 13889–13906, 2008. doi:10.1523/JNEUROSCI.1946-08.2008.
- Huang X, Lisberger SG. Circuit mechanisms revealed by spike-timing correlations in macaque area MT. *J Neurophysiol* 109: 851–866, 2013. doi:10.1152/jn.00775.2012.
- Huk AC, Dougherty RF, Heeger DJ. Retinotopy and functional subdivision of human areas MT and MST. *J Neurosci* 22: 7195–7205, 2002. doi:10.1523/JNEUROSCI.22-16-07195.2002.
- Huk AC, Heeger DJ. Task-related modulation of visual cortex. *J Neurophysiol* 83: 3525–3536, 2000. doi:10.1152/jn.2000.83.6.3525.
- Kafaligonul H, Stoner GR. Auditory modulation of visual apparent motion with short spatial and temporal intervals. *J Vis* 10: 31, 2010. doi:10.1167/10.12.31.
- Kafaligonul H, Stoner GR. Static sound timing alters sensitivity to low-level visual motion. *J Vis* 12: 2, 2012. doi:10.1167/12.11.2.
- Kajikawa Y, Schroeder CE. How local is the local field potential? *Neuron* 72: 847–858, 2011. doi:10.1016/j.neuron.2011.09.029.
- Kayser C, Petkov CI, Logothetis NK. Visual modulation of neurons in auditory cortex. *Cereb Cortex* 18: 1560–1574, 2008. doi:10.1093/cercor/bhm187.
- Kolers PA. *Aspects of Motion Perception*. Oxford, UK: Pergamon, 1972.
- Kopčo N, Lin IF, Shinn-Cunningham BG, Groh JM. Reference frame of the ventriloquist aftereffect. *J Neurosci* 29: 13809–13814, 2009. doi:10.1523/JNEUROSCI.2783-09.2009.
- Korte A. Kinematoskopische Untersuchungen [Cinematographic investigations]. *Z Psychol* 72: 194–296, 1915.
- Krekelberg B, Albright TD. Motion mechanisms in macaque MT. *J Neurophysiol* 93: 2908–2921, 2005. doi:10.1152/jn.00473.2004.
- Krekelberg B, van Wezel RJA, Albright TD. Interactions between speed and contrast tuning in the middle temporal area: implications for the neural code

- for speed. *J Neurosci* 26: 8988–8998, 2006a. doi:10.1523/JNEUROSCI.1983-06.2006.
- Krekelberg B, van Wezel RJA, Albright TD. Adaptation in macaque MT reduces perceived speed and improves speed discrimination. *J Neurophysiol* 95: 255–270, 2006b. doi:10.1152/jn.00750.2005.
- Lee H, Simpson GV, Logothetis NK, Rainer G. Phase locking of single neuron activity to theta oscillations during working memory in monkey extrastriate visual cortex. *Neuron* 45: 147–156, 2005. doi:10.1016/j.neuron.2004.12.025.
- Liu J, Newsome WT. Correlation between speed perception and neural activity in the middle temporal visual area. *J Neurosci* 25: 711–722, 2005. doi:10.1523/JNEUROSCI.4034-04.2005.
- Liu J, Newsome WT. Local field potential in cortical area MT: stimulus tuning and behavioral correlations. *J Neurosci* 26: 7779–7790, 2006. doi:10.1523/JNEUROSCI.5052-05.2006.
- Maier A, Wilke M, Aura C, Zhu C, Ye FQ, Leopold DA. Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nat Neurosci* 11: 1193–1200, 2008. doi:10.1038/nn.2173.
- Maunsell JH, Gibson JR. Visual response latencies in striate cortex of the macaque monkey. *J Neurophysiol* 68: 1332–1344, 1992. doi:10.1152/jn.1992.68.4.1332.
- Maunsell JH, Van Essen DC. 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, 1983. doi:10.1152/jn.1983.49.5.1127.
- McDonald JJ, Störmer VS, Martinez A, Feng W, Hillyard SA. Salient sounds activate human visual cortex automatically. *J Neurosci* 33: 9194–9201, 2013. doi:10.1523/JNEUROSCI.5902-12.2013.
- Mendoza-Halliday D, Torres S, Martinez-Trujillo JC. Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nat Neurosci* 17: 1255–1262, 2014. doi:10.1038/nn.3785.
- Mercier MR, Foxe JJ, Fiebelkorn IC, Butler JS, Schwartz TH, Molholm S. Auditory-driven phase reset in visual cortex: human electrocorticography reveals mechanisms of early multisensory integration. *Neuroimage* 79: 19–29, 2013. doi:10.1016/j.neuroimage.2013.04.060.
- Meredith MA, Nemitz JW, Stein BE. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci* 7: 3215–3229, 1987. doi:10.1523/JNEUROSCI.07-10-03215.1987.
- Mikami A. Direction selective neurons respond to short-range and long-range apparent motion stimuli in macaque visual area MT. *Int J Neurosci* 61: 101–112, 1991. doi:10.3109/00207459108986278.
- Mikami A. Spatiotemporal characteristics of direction-selective neurons in the middle temporal visual area of the macaque monkeys. *Exp Brain Res* 90: 40–46, 1992. doi:10.1007/BF00229254.
- Mikami A, Newsome WT, Wurtz RH. Motion selectivity in macaque visual cortex. I. Mechanisms of direction and speed selectivity in extrastriate area MT. *J Neurophysiol* 55: 1308–1327, 1986a. doi:10.1152/jn.1986.55.6.1308.
- Mikami A, Newsome WT, Wurtz RH. Motion selectivity in macaque visual cortex. II. Spatiotemporal range of directional interactions in MT and V1. *J Neurophysiol* 55: 1328–1339, 1986b. doi:10.1152/jn.1986.55.6.1328.
- Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res* 14: 115–128, 2002. doi:10.1016/S0926-6410(02)00066-6.
- Morein-Zamir S, Soto-Faraco S, Kingstone A. Auditory capture of vision: examining temporal ventriloquism. *Brain Res Cogn Brain Res* 17: 154–163, 2003. doi:10.1016/S0926-6410(03)00089-2.
- Nakayama K. Biological image motion processing: a review. *Vision Res* 25: 625–660, 1985. doi:10.1016/0042-6989(85)90171-3.
- Naue N, Rach S, Strüder D, Huster RJ, Zaehle T, Körner U, Herrmann CS. Auditory event-related response in visual cortex modulates subsequent visual responses in humans. *J Neurosci* 31: 7729–7736, 2011. doi:10.1523/JNEUROSCI.1076-11.2011.
- Newsome WT, Mikami A, Wurtz RH. Motion selectivity in macaque visual cortex. III. Psychophysics and physiology of apparent motion. *J Neurophysiol* 55: 1340–1351, 1986. doi:10.1152/jn.1986.55.6.1340.
- Newsome WT, Paré EB. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J Neurosci* 8: 2201–2211, 1988. doi:10.1523/JNEUROSCI.08-06-02201.1988.
- O’Beirne GA, Patuzzi RB. Basic properties of the sound-evoked post-auricular muscle response (PAMR). *Hear Res* 138: 115–132, 1999. doi:10.1016/S0378-5955(99)00159-8.
- Orban GA, Saunders RC, Vandenbussche E. Lesions of the superior temporal cortical motion areas impair speed discrimination in the macaque monkey. *Eur J Neurosci* 7: 2261–2276, 1995. doi:10.1111/j.1460-9568.1995.tb00647.x.
- Perrone JA, Thiele A. Speed skills: measuring the visual speed analyzing properties of primate MT neurons. *Nat Neurosci* 4: 526–532, 2001. doi:10.1038/87480.
- Priebe NJ, Lisberger SG. Estimating target speed from the population response in visual area MT. *J Neurosci* 24: 1907–1916, 2004. doi:10.1523/JNEUROSCI.4233-03.2004.
- Recanzone GH. Auditory influences on visual temporal rate perception. *J Neurophysiol* 89: 1078–1093, 2003. doi:10.1152/jn.00706.2002.
- Rudolph K, Pasternak T. Transient and permanent deficits in motion perception after lesions of cortical areas MT and MST in the macaque monkey. *Cereb Cortex* 9: 90–100, 1999. doi:10.1093/cercor/9.1.90.
- Scheef L, Boecker H, Daamen M, Fehse U, Landsberg MW, Granath DO, Mechling H, Effenberg AO. Multimodal motion processing in area V5/MT: evidence from an artificial class of audio-visual events. *Brain Res* 1252: 94–104, 2009. doi:10.1016/j.brainres.2008.10.067.
- Schoppmann A, Hoffmann KP. Continuous mapping of direction selectivity in the cat’s visual cortex. *Neurosci Lett* 2: 177–181, 1976. doi:10.1016/0304-3940(76)90011-2.
- Schroeder CE, Foxe J. Multisensory contributions to low-level, ‘unisensory’ processing. *Curr Opin Neurobiol* 15: 454–458, 2005. doi:10.1016/j.conb.2005.06.008.
- Senkowski D, Gomez-Ramirez M, Lakatos P, Wylie GR, Molholm S, Schroeder CE, Foxe JJ. Multisensory processing and oscillatory activity: analyzing non-linear electrophysiological measures in humans and simians. *Exp Brain Res* 177: 184–195, 2007. doi:10.1007/s00221-006-0664-7.
- Senkowski D, Schneider TR, Foxe JJ, Engel AK. Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci* 31: 401–409, 2008. doi:10.1016/j.tins.2008.05.002.
- Shi Z, Chen L, Müller HJ. Auditory temporal modulation of the visual Ternus effect: the influence of time interval. *Exp Brain Res* 203: 723–735, 2010. doi:10.1007/s00221-010-2286-3.
- Staal HE, Donderi DC. The effect of sound on visual apparent movement. *Am J Psychol* 96: 95–105, 1983. doi:10.2307/1422212.
- Stanford TR, Quessy S, Stein BE. Evaluating the operations underlying multisensory integration in the cat superior colliculus. *J Neurosci* 25: 6499–6508, 2005. doi:10.1523/JNEUROSCI.5095-04.2005.
- Stoner GR, Albright TD. Neural correlates of perceptual motion coherence. *Nature* 358: 412–414, 1992. doi:10.1038/358412a0.
- Sundberg KA, Mitchell JF, Gawne TJ, Reynolds JH. Attention influences single unit and local field potential response latencies in visual cortical area V4. *J Neurosci* 32: 16040–16050, 2012. doi:10.1523/JNEUROSCI.0489-12.2012.
- Thompson P. Perceived rate of movement depends on contrast. *Vision Res* 22: 377–380, 1982. doi:10.1016/0042-6989(82)90153-5.
- Wang Y, Iliescu BF, Ma J, Josić K, Dragoi V. Adaptive changes in neuronal synchronization in macaque V4. *J Neurosci* 31: 13204–13213, 2011. doi:10.1523/JNEUROSCI.6227-10.2011.
- Watson AB, Ahumada AJ, Farrell JE. Window of visibility: a psychophysical theory of fidelity in time-sampled visual motion displays. *J Opt Soc Am A Opt Image Sci Vis* 3: 300–307, 1986. doi:10.1364/JOSAA.3.000300.
- Woods TM, Recanzone GH. Visually induced plasticity of auditory spatial perception in macaques. *Curr Biol* 14: 1559–1564, 2004. doi:10.1016/j.cub.2004.08.059.