

The Elicitation of Audiovisual Steady-State Responses: Multi-Sensory Signal Congruity and Phase Effects

Julian Jenkins III · Ariane E. Rhone ·
William J. Idsardi · Jonathan Z. Simon ·
David Poeppel

Received: 21 June 2010 / Accepted: 23 February 2011 / Published online: 6 March 2011
© Springer Science+Business Media, LLC 2011

Abstract Most ecologically natural sensory inputs are not limited to a single modality. While it is possible to use real ecological materials as experimental stimuli to investigate the neural basis of multi-sensory experience, parametric control of such tokens is limited. By using artificial bimodal stimuli composed of approximations to ecological signals, we aim to observe the interactions between putatively relevant stimulus attributes. Here we use MEG as an electrophysiological tool and employ as a measure the steady-state

response (SSR), an experimental paradigm typically applied to unimodal signals. In this experiment we quantify the responses to a bimodal audio-visual signal with different degrees of temporal (phase) congruity, focusing on stimulus properties critical to audiovisual speech. An amplitude modulated auditory signal ('pseudo-speech') is paired with a radius-modulated ellipse ('pseudo-mouth'), with the envelope of low-frequency modulations occurring in phase or at offset phase values across modalities. We observe (i) that it is possible to elicit an SSR to bimodal signals; (ii) that bimodal signals exhibit greater response power than unimodal signals; and (iii) that the SSR power at specific harmonics and sensors differentially reflects the congruity between signal components. Importantly, we argue that effects found at the modulation frequency and second harmonic reflect differential aspects of neural coding of multisensory signals. The experimental paradigm facilitates a quantitative characterization of properties of multi-sensory speech and other bimodal computations.

Keywords Audio-visual · Cross-modal · Magnetoencephalography · Speech · Multi-sensory

J. Jenkins III (✉) · J. Z. Simon
Department of Biology, University of Maryland,
College Park, 1206 Biology-Psychology Building,
College Park, MD 20742, USA
e-mail: julianj@umd.edu

J. Z. Simon
e-mail: jzsimon@umd.edu

A. E. Rhone · W. J. Idsardi
Department of Linguistics, University of Maryland, College
Park, 1401 Marie Mount Hall, College Park, MD 20742, USA
e-mail: arhone@umd.edu

W. J. Idsardi
e-mail: idsardi@umd.edu

J. Jenkins III · A. E. Rhone · W. J. Idsardi
Cognitive Neuroscience of Language Laboratory,
University of Maryland, College Park, 1401 Marie Mount Hall,
College Park, MD 20742, USA

J. Z. Simon
Department of Electrical and Computer Engineering,
University of Maryland, College Park, 2209 AV Williams,
College Park, MD 20742, USA

D. Poeppel
Department of Psychology, New York University,
6 Washington Place, New York, NY 10003, USA
e-mail: david.poeppel@nyu.edu

Introduction

The majority of naturalistic sensory experiences require the observer not only to segregate information into separate objects or streams but also to integrate related information into a coherent percept across sensory modalities—as well as across space and time (Amedi et al. 2005; Kelly et al. 2008; Lalor et al. 2007; Macaluso and Driver 2005; Miller and D'Esposito 2005; Molholm et al. 2002, 2004, 2007; Murray et al. 2005; Senkowski et al. 2006). Integration of this information not only unifies the perception of events,

but the presence of redundant information also facilitates recognition, increases signal-to-noise ratios, and decreases reaction times to cross-modal events (Driver and Spence 1998; Hershenson 1962; Senkowski et al. 2006; Stein et al. 1989). Studies examining the simultaneous serial and parallel computations and physiological responses underlying the integration of information and the recognition of a unified percept have important implications for advancing the understanding of the binding of cross-modal information for ecologically valid behaviors such as motion perception and speech recognition and comprehension (Baumann and Greenlee 2007; Lakatos et al. 2008; Miller and D'Esposito 2005; Schroeder and Lakatos 2009; Schroeder et al. 2008).

While it has traditionally been thought that processing of cross-modal events occurs primarily in association cortices (Jones and Powell 1970; Mesulam 1998), much recent evidence indicates that information from other sensory modalities can influence cortical areas conventionally assumed to be unimodal (Ghazanfar and Schroeder 2006). For example, electroencephalographic (EEG), functional magnetic resonance imaging (fMRI) and magnetoencephalographic (MEG) studies in humans have provided evidence that visual and somatosensory signals can influence neuronal activity in the auditory cortex (e.g., see Schroeder and Foxe 2005 for a review). Intracranial recordings and anatomical tracings in macaques have affirmed the existence of multisensory inputs to putatively unimodal cortical areas (Kayser et al. 2008). In humans, several functional imaging and intracranial studies have identified cortical networks involved in object recognition, auditory-somatosensory and visual-somatosensory processing and integration of audio-visual speech (Calvert et al. 1999, 2000, 2001; Molholm et al. 2004, 2006; Senkowski et al. 2008). Human imaging studies have identified the superior colliculus, superior temporal sulcus, intraparietal sulcus, insula and several frontal cortical areas as being involved in crossmodal computation (Calvert et al. 2001). With regard to speech, the traditional speech areas (perisylvian) have been implicated, as well as the superior parietal, inferior parietal, inferior frontal, superior temporal sulcus and left claustrum areas (Calvert et al. 2000; Campbell 2008; Fort et al. 2002; Olson et al. 2002). These findings emphasize the importance of rapid synchronization of crossmodal information in heteromodal cortical areas.

A number of event-related potential (ERP) studies have examined the temporal aspects of cross-modal interactions, motivated by the hypothesis that the decrease in reaction time and facilitation of object recognition should be visible in electrophysiological recordings. These studies have found significant activity within several latency windows, with the most surprising results for audio-visual interactions coming at ~50–100 ms post-stimulus onset, suggesting early cortical processing of audiovisual interactions (Molholm et al.

2002). In addition, several ERP studies have also evaluated facilitation of bimodal interactions via an additive model (Besle et al. 2004). These studies typically have shown amplitude and latency facilitation due to bimodal interactions localized to multi-modal cortical areas, as well as suppression of electrophysiological responses with cortical generators in (putatively) unimodal areas.

A slightly different electrophysiological paradigm for investigating the computational advantages of cross-modal interactions is provided by the steady-state response (SSR), which is the result of entrainment to the temporal properties of a modulated stimulus. This response has been documented for both visual and auditory signals and has been used extensively for clinical and diagnostic purposes (Sohmer et al. 1977). Auditory SSRs are generally elicited by amplitude or frequency modulated signals, or both (e.g., Luo et al. 2006), while visual SSRs are typically elicited by transient high-contrast stimuli such as checkerboard reversals or luminance flicker. Though commonly measured with EEG, the same principles of frequency entrainment to periodic stimuli have been evaluated in MEG (Müller et al. 1997; Ross et al. 2000). Ecological stimuli that are temporally extended and have a quasi-steady-state nature, such as speech, can also be modeled via stimuli that approximate the excitation produced by domain-specific information (Grant and Seitz 2000). SSRs have a potential further advantage: they can be used to exploit endogenous cortical oscillations. These oscillations are amplified when preferential stimuli (i.e., stimuli that match the frequency and phase of the endogenous oscillations) constitute the sensory input (Schroeder and Lakatos 2009; Schroeder et al. 2008; Senkowski et al. 2008). Oscillatory activity of particular interest occurs in frequency ranges that are important for relevant behaviors such as speech comprehension, working memory function and selectional attention (Luo et al. 2010; Luo and Poeppel 2007; Senkowski et al. 2008; Talsma et al. 2006).

The motivation for the current study was to model an ecologically typical audio-visual interaction, multi-sensory speech, incorporating some of its critical temporal attributes. The auditory component of speech consists of relatively rapid frequency fluctuations in the spectral domain, along with slower amplitude modulation (i.e., the envelope)—reminiscent of an amplitude-modulated (AM) sinusoidal auditory signal. The speech signal itself shows significant AM activity in the 2–16 Hz range (Steeneken and Houtgast 1980), and it has been shown that cortical decomposition of the speech envelope is particularly sensitive to frequencies in the range of 4–16 Hz. For example, recent MEG evidence supports this generalization: Luo and Poeppel (2007) and Howard and Poeppel (Howard and Poeppel 2010) observed that fluctuations in the speech envelope are associated with intrinsic oscillations in the theta frequency band (~4–8 Hz). Luo et al. (2010)

extended that to the delta band (1–3 Hz) when AV speech was used. Paired with the auditory signal is a visual component in which facial features—and especially mouth movements—aid comprehension, especially in noisy environments (Sumby and Pollack 1954). We thus crafted stimuli consisting of modulated auditory and visual components within the frequency range of the envelope of speech. By building on results investigating SSRs to auditory and visual stimuli presented alone, we assess the SSR to bimodal audio-visual signals.

For the experiment reported in this paper, the visual signal consists of an ellipse to approximate a mouth opening and closing, and the auditory signal consists of amplitude-modulated three-octave pink noise to approximate the envelope and the wideband carrier features of the speech signal. We hypothesize that the SSRs elicited by concurrently modulated (comodal) audio-visual signals should be greater than the responses elicited by unimodally modulated auditory or visual stimuli, as reflected by the amplitude spectrum at the modulation frequency and its harmonics. The increased signal power of the comodal conditions relative to unimodal conditions might reflect increased activity due to synchrony of different neural populations involved in evaluating the multimodal signal. By manipulating the phase congruence of one modality relative to the other, we additionally aimed to elucidate the online cross-talk between modalities. In the experiment presented here, we demonstrate the feasibility of bimodal SSR as an experimental paradigm, the cortical/sensor areas that are predictors for evaluating signal component synchronicity, and the particular SSR component that indexes signal envelope component synchronicity.

Materials and Methods

The experimental design and data presented in this manuscript result from an earlier experimental design where we determined the feasibility of eliciting a bimodal SSR (at two distinct modulation frequencies) as well as the most appropriate methods to analyze the data. The major methodological points from that pilot experiment are as follows: (i) pretest stimuli to determine sensors that responded preferentially to a given modality (see pretests below); (ii) for the SSR analysis, we first averaged trial presentations, multiplied the data within the signal evaluation window in the temporal domain by a Kaiser window ($\beta = 13$) to remove spurious frequency contributions and minimize onset and offset responses and then evaluated the Fourier transform; (iii) evaluated the significance of SSR activity using a combination of F tests and Rayleigh's phase test; (iv) evaluated onset responses using principal component analysis (PCA); (v) performed a cross-modal control analysis to

verify the responses recorded were not sensor-dependent and (vi) evaluated the significance of the results using a combination of parametric and non-parametric statistical tests.

The results of the prior experiment revealed whether or not there was any asymmetry in the response power to each modality, and also allowed us to fine-tune the signal evaluation methods and statistics used to evaluate the responses, which are applied to the data presented below.

Participants

Fourteen participants (thirteen right-handed; one ambidextrous, as tested by the Edinburgh Handedness Inventory (Oldfield 1971); six female) with normal hearing and normal or corrected-to-normal vision underwent MEG scanning. Data from two participants were excluded due to an insufficient signal-to-noise ratio for all conditions. Age range was 18–27 (mean 20.1 years). Participants were compensated for their participation (\$10/hr). Presentation of stimuli and biomagnetic recording was performed with the approval of the institutional committee on human research of the University of Maryland, College Park. Prior to the start of the experiment, written informed consent was obtained from each participant.

Stimuli

In the strictest sense, all signals presented were bimodal; we use the terms “unimodal” and “comodal” in a specific manner here to distinguish signal types. Unimodal refers to conditions where only one modality undergoes envelope modulation, while the other modality is presented with a static signal to control for basic sensory excitation. Comodal conditions refer to simultaneous envelope modulation of both modalities.

As mentioned in the “[Introduction](#)”, the modulation frequency employed was intended to model the naturalistic modulation rates found in speech. Though the majority of SSR studies in both the visual and auditory domains use high modulation rates (e.g., 30–80 Hz), we observed unique challenges most likely due to the shape and contrast of the experimental signals and the extended duration of the experiment (approximately 60 min). During a piloting phase, we tested modulation frequencies of 6–8 Hz. Although we observed SSR activity at these frequencies, modulation at these rates caused participant eye fatigue and discomfort, leading us to lower the modulation frequency to 3.125 Hz. This modulation frequency satisfies several criteria: (i) it is in the range of speech modulation, (ii) it falls within a specific frequency bin to eliminate the need for filtering and windowing the SSR response to make signal extraction easier and reduce information loss and (iii) it was more comfortable for participants to attend to.

The SSR-inducing stimuli consisted of five types of audio-visual signals presented at one modulation frequency, along with three target signals, for a total of eight signals. The five types of SSR-inducing signals were: (i) amplitude-modulated three-octave pink noise presented concurrently with a static white rectangle on black background; (ii) a radius-modulated white ellipse on black background concurrently presented with approximately Gaussian white acoustic noise; (iii–v) a radius-modulated ellipse paired with amplitude modulated three-octave pink noise at one of three envelope phase relationships (in phase, $\pi/2$ radians out of phase, π radians out of phase). The amplitude-modulated three-octave pink noise and radius-modulated ellipses were modulated at 3.125 Hz with a modulation depth of 25% of peak amplitude and radius for audio and visual signals, respectively (Fig. 1). The SSR-inducing signals were 3.520 s in duration. For the comodulated conditions, the auditory and visual signal components had the same onset and offset, with the

auditory component reaching the maximum value of the modulation envelope first for out-of-phase conditions.

Auditory signal components were generated with MATLAB (R2009a, The Mathworks, Natick, MA) and consisted of a cosine wave envelope (3.125 Hz modulation frequency) applied to a three-octave pink noise carrier signal with 6 ms \cos^2 onset and offset ramps presented at approximately 65 dB SPL. Using the cosine function for the envelopes was done to enhance the onset responses; the cosine function starts at the maximum value, which will generate a more robust onset response. The pink noise was generated using the NSL Toolbox (Chi and Shamma, <http://www.isr.umd.edu/Labs/NSL/Software.htm>) for MATLAB. The three-octave pink noise contained a lowest frequency of 125 Hz; these parameters cover the fundamental frequency range of the human voice as well as the frequency region where most of the energy arising from the first formant tends to be concentrated. The signals were sampled at 44.1 kHz with 16-bit resolution. Visual signal

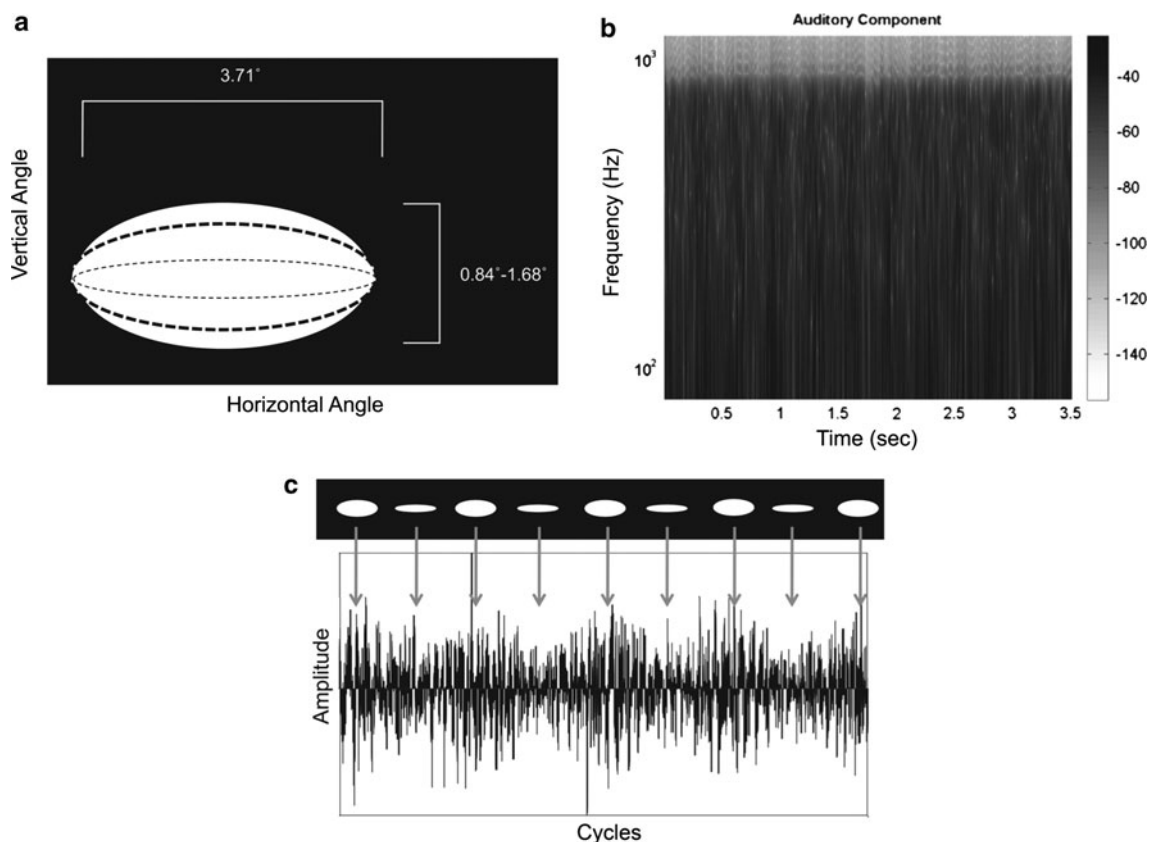


Fig. 1 **a** Schematic of stimuli employed. This panel illustrates the movement of the visual signal component throughout the duration of stimulus (see “Materials and Methods” for details). The stimuli were presented at $F_m = 3.125$ Hz, modulation depth was 25%. The x-axis shows the horizontal visual angle of the signal component and the y-axis the ranges of vertical visual angles. **b** This panel illustrates the auditory signal component spectral structure. This component consisted of three-octave pink noise (lowest frequency: 125 Hz), amplitude modulated at 3.125 Hz, 25% modulation depth. Intensity

values are plotted using a grayscale intensity axis. The x-axis is time in seconds; the y-axis is Frequency (Hz) on a logarithmic scale. **c** Schematized temporal evolution of comodulated signal, signal component envelopes completely in phase, over several cycles. Top portion illustrates the visual component; bottom portion the auditory component. The visual component is seen to ‘open’ and ‘close’ over the duration of the signal; it is easily seen that in this experimental condition, the ellipse modulation is synchronized with the amplitude modulations in the auditory component

components were generated using Gnu Image Manipulation Program (www.gimp.org). The radius-modulated white ellipses were centered on a 640×480 pixel black background, and ranged from 0.84° to 1.68° visual angle for the minor radius and 3.71° visual angle for the major radius. The minor radius was modulated to simulate mouth movements. The individual frames were compiled into Audio–Video Interleave (AVI) format using VirtualDub (www.virtualdub.org) for presentation. Stimulus timing/frequency was verified with an oscilloscope. The visual components were projected on a screen approximately 30 cm from the participant's nasion. Participants were supine in the MEG scanner for the duration of the experiment.

To maintain participant vigilance to both modalities, brief targets were pseudorandomly interleaved throughout the experimental trials. Targets were of three types: (i) an auditory only target consisting of approximately Gaussian white noise; (ii) a visual only target consisting of a white crosshair on a black background; (iii) an audiovisual target consisting of a white crosshair on black background paired with approximately Gaussian white noise. Target duration was 500 ms.

Experimental stimuli were presented in six blocks, with 15 repetitions per signal per block, for a total of ninety trials per condition. Presentation of conditions was randomized within blocks. The SSR-inducing materials were passively attended to; no response to those signals was required. For the target signals (38% of trials), participants were required to press a button indicating their detection of the target.

Delivery

All experimental stimuli were presented using a Dell Optiplex computer with a M-Audio Audiophile 2496 sound card (Avid Technology, Inc., Irwindale, CA) via Presentation stimulus presentation software (Neurobehavioral Systems, Inc., Albany, CA). Stimuli were delivered to the participants binaurally via Eartone ER3A transducers and non-magnetic air-tube delivery (Etymotic, Oak Brook, IL). The inter-stimulus interval varied pseudo-randomly between 980 and 2000 ms.

Recording and Filtering

Data were acquired using a 160-channel whole-head biomagnetometer with axial gradiometer sensors (KIT System, Kanazawa, Japan). Recording bandwidth was DC–200 Hz, with a 60 Hz Notch filter, at 1000 Hz sampling rate. The data were noise reduced using time-shifted PCA (de Cheveigné and Simon 2007) and trials were averaged offline (artifact rejection ± 2.5 pT) and baseline corrected. Data for the SSR analysis were not filtered; however, data for examining the onset responses were filtered. The filter employed was a 4th order low-pass

elliptical filter with a 40 Hz cutoff frequency, 0.5 dB peak-to-peak ripple and at least 60 dB stopband attenuation.

Sensor Selection from Pre-Test

Determination of maximally responsive auditory and visual channels was performed in separate pre-tests. The auditory pre-test consisted of amplitude-modulated sinusoidal signals with an 800 Hz sinusoidal carrier signal, modulation frequency (F_m) 7 Hz, modulation depth 100% and 11.3 s duration. The visual pre-test consisted of a checkerboard flicker pattern ($F_m = 4$ Hz), of 240 s duration. The sensor space was divided into quadrants to characterize the auditory response and sextants to characterize the visual response based on the peak and trough field topography expected for each modality as recorded from axial gradiometers (see Fig. 2). Sensor channel designations were anterior temporal (front of head), posterior temporal (rear quadrants/middle of head) and occipital (back of head overlying occipital lobe). Five channels from source and sink from each sensor division (i.e., ten channels for auditory response and five channels for visual response per hemisphere; 15 channels per hemisphere total) with the maximum measured magnetic field deflection were used for subsequent analyses. The analysis window for the PSD analysis of the visual pretest was 10 s and for the auditory pretest 11 s.

Onset Response Evaluation

The signal evaluation window (averaged and filtered sensor data) ranged from 500 ms pre-trigger to 3519 ms post-trigger. For several participants with exceptionally clean and robust onset responses, examination of the data revealed three distinct evoked peaks: (i) in the range of ~ 70 – 85 ms post-stimulus onset, with an auditory magnetic field topography; (ii) in the range of ~ 120 – 150 ms post-stimulus onset, with a visual field topography and (iii) in the range of ~ 180 – 240 ms post-stimulus onset, which was a combination of auditory and visual topographies. For the majority of participants however, such clear response patterns were not observed. Due to the univariate design, we calculated the RMS of the individual participant RMS vectors, and grand averages of the magnetic field deflections. Permutation tests were performed on time ranges obtained from visible peaks in the grand averaged waveform. Latency values used in permutation tests were taken from individual RMS vectors in the time ranges of the peaks observed. These data were taken from the filtered and baseline corrected individual participant data; baseline correction within participants served as a normalization of the data used in the grand averages. The number of trials averaged was, at a minimum, 80 out of 90 presentations.

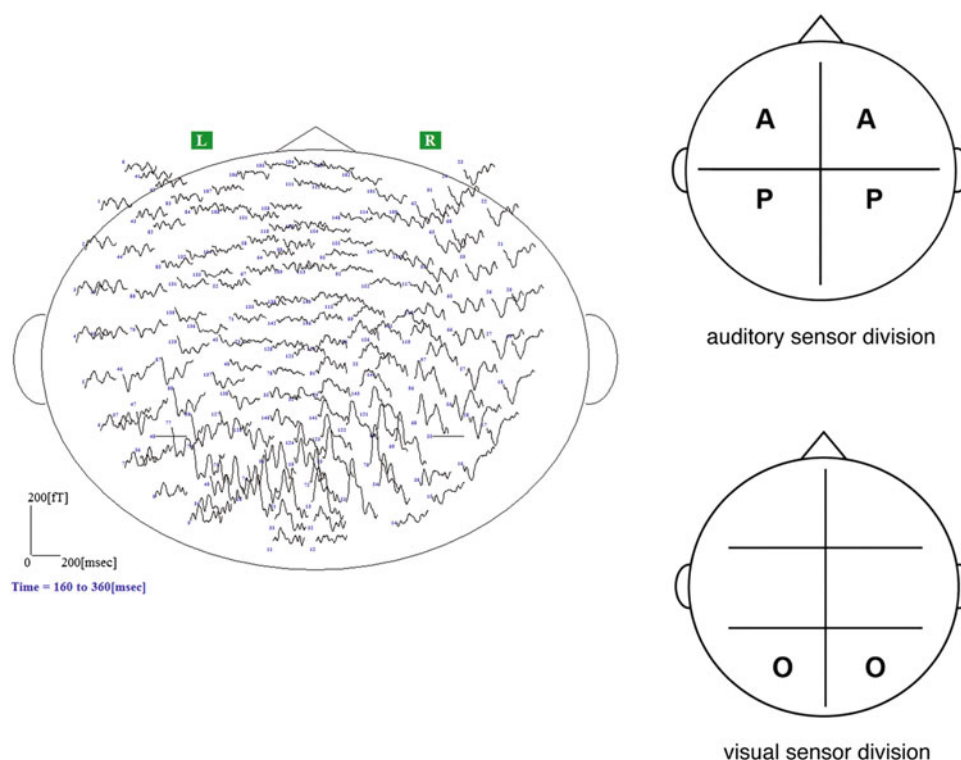


Fig. 2 *Left panel:* Sensor layout of whole-head biomagnetometer with field deflections overlaid. Anterior portion of the head is at the top. This figure gives a sense of the positioning of the sensors in the dewar; robust evoked activity is seen at the channels overlying the occipital and posterior temporal lobes. Data are from a single participant, $\Phi = 0$ comodul condition. *Right panel:* Division of magnetoencephalographic sensors. Top panel shows division of auditory sensors for experimental pre-test; bottom panel shows sensor

division for visual pre-test. Sensor division was based on expected field topography for auditory and visual cortical responses recorded from axial gradiometer sensors (see “[Materials and Methods](#)” for details). Sensor designation is as follows: A = anterior temporal sensors, P = posterior temporal sensors, O = occipital sensors. Placement of letters roughly corresponds to the locations of the sensors selected for the analysis of the experimental data

SSR Analysis

The magnitude and phase spectra of the SSR were determined using the Fast Fourier Transform (FFT) of the baseline corrected channel data. The FFT was calculated from 320 ms post-stimulus onset to the end of the signal evaluation window (3519 ms) for a total of 3200 samples analyzed; this yielded frequency bins commensurate with the modulation frequency and its harmonics. The magnitude of the response was calculated using the RMS of the FFT across channels. The phase response was determined by calculating the mean direction as described by Fisher (1996) based on the phase angle of the Fourier transformed data. The across participant response power was determined by calculating the mean of the individual participant power vectors. To determine the across participant phase response, the mean direction of the individual mean directions was calculated.

Across-Participant Response Averaging

Onset responses were collected and evaluated as described above. A similar procedure was used for the Fourier

transformed data (collection of FFT vectors and grand averages computed). Individual participant vectors for response power (squared magnitude) and phase were collected and the relevant statistics calculated as described below.

Statistical Analyses

The significance of the SSR amplitude at a specific frequency was analyzed by performing an F test on the squared RMS (power) of the Fourier transformed data using the MATLAB Statistics Toolbox. The F test takes into account both amplitude and phase (Valdes et al. 1997; Picton et al. 2003). For the across-participant data, F tests were performed on the power of the SSR at the modulation frequency and the second harmonic. The response power in linear values and decibels (dB) was assessed using ANOVAs as well as general linear models (GLMs) using the “languageR” statistical package (R Foundation for Statistical Computing, v. 2.10.1; Baayen 2008). Factors for both sets of statistical tests were Hemisphere, Harmonic, Condition, and Sensor Area, with Participant as a random

effect. To determine the separation of densities, distributions of the responses for each hemisphere, harmonic, condition and area were compared using Kolmogorov–Smirnov tests. Use of ANOVAs is standard when comparing responses across participants in electrophysiological experiments (Jenkins et al. 2010); we used the power afforded by GLMs to determine more robustly the predictors of the response. K–S tests were used to see if the response distributions in each of the sensor areas were statistically different.

Additionally, we compared response additivity using the AV versus ($A + V$) model, but not via RMS of the recorded responses (if the additivity were assessed using RMS, this would assume a single source is generating the response; because the responses examined involve two sensory domains, this would not be parsimonious). As such, additivity evaluations were made using the complex representation from the Fourier transform of the data on the frequency bins containing the frequencies of interest, specifically the modulation frequency and the second harmonic. Responses at the third harmonic were not statistically different from background noise. Statistical differences were assessed using Wilcoxon signed-rank tests in order to decrease the assumptions concerning the distribution of the data recorded between pairs of conditions.

Participant Head Location

Though we did not perform dipole localization, we did take marker measurements and recorded digitized headshapes for the participants. The marker coils were placed by common anatomical markers: by each preauricular point and three frontal coils based on spacing from the nasion. Head position measurements were taken prior to and after experimental completion to determine proper head placement within the dewar and that the sensors were recording from the entire head (occipital, posterior temporal/parietal, anterior temporal/frontal areas). This also aided in ensuring that sensor selection from the pretests was correct.

Results

Figure 3 illustrates the nature of the data recorded using a single participant; magnetic flux is in black, RMS in red. The panels, from top to bottom, (i) data from pre-stimulus onset to the end of the analysis frame, (ii) a zoomed-in view of the onset response and (iii) the steady-state portion of the evoked response. For the onset response, there are two distinct peaks in the RMS, one occurring at ~ 140 ms and the other at ~ 210 ms post-stimulus onset. The oscillatory activity seen in the evoked response indicates entrainment to the periodic properties of the experimental

stimuli (verified via the Fourier transform of the magnetic signals and statistical assessment).

SSR responses were reliably generated. The response pattern observed indicated (as measured using ANOVAs and GLMs as well as data visualization) that there was no difference between hemispheres in the power of the response, and that the posterior temporal and occipital channels captured the response best (Fig. 4).

Examination and analysis of the SSR power indicated that it would be more advantageous to analyze the responses in terms of decibel (dB) power, rather than linear power values, due to the effectively normally distributed nature of dB power measurements (Dobie and Wilson 1996; see Fig. 5). Data visualization of power densities was performed using the “ggplot2” package for R (Wickham 2009). The dB values readily yield to a more robust and easily comprehensible statistical analysis.

Across-Participant Power Analysis

Most of the response power was generated in the sensors overlying the posterior temporal and occipital areas. Response power was concentrated at the modulation frequency and the second harmonic, and the power values at those frequencies were used for the subsequent statistical analyses. Statistical significance was assessed using F tests with 2 and 12 degrees of freedom ($df = 2, 12, \alpha = 0.05$) and was confirmed by comparing the average power of the background noise (surrounding frequency bins) with the bin containing the modulation frequency. On average, the frequency bins containing the frequencies of interest were an order of magnitude (~ 10 dB) greater than the background, with exceptions for certain sensor areas and conditions (i.e., responses measured at anterior temporal sensors).

For the unimodal modulation conditions, statistically significant F ratios were found at the modulation frequency for the occipital sensors in both hemispheres (LH: $F = 37.441, P < 0.01$; RH: $F = 10.539, P < 0.01$), but not for the anterior and posterior temporal sensors; the second harmonic F ratio was significant only in the RH occipital sensors ($F = 7.853, P < 0.01$). For the $\Phi = 0$ comodul condition at the modulation frequency, significant F ratios were found for the posterior temporal and occipital sensors in the LH ($F = 7.822, P < 0.01$ and $F = 60.107, P < 0.01$, respectively); the RH occipital sensors F ratio was marginally significant ($F = 4.113, P < 0.05$); this same pattern held for the second harmonic ($F = 4.839, P < 0.05$; $F = 4.733, P < 0.05$; $F = 4.061, P < 0.05$, respectively). For the $\Phi = \pi/2$ condition, significant F ratios were found for the occipital sensors in both hemispheres at the modulation frequency (LH: $F = 74.436, P < 0.01$; RH: $F = 10.04, P < 0.01$) and the LH occipital sensors for the second

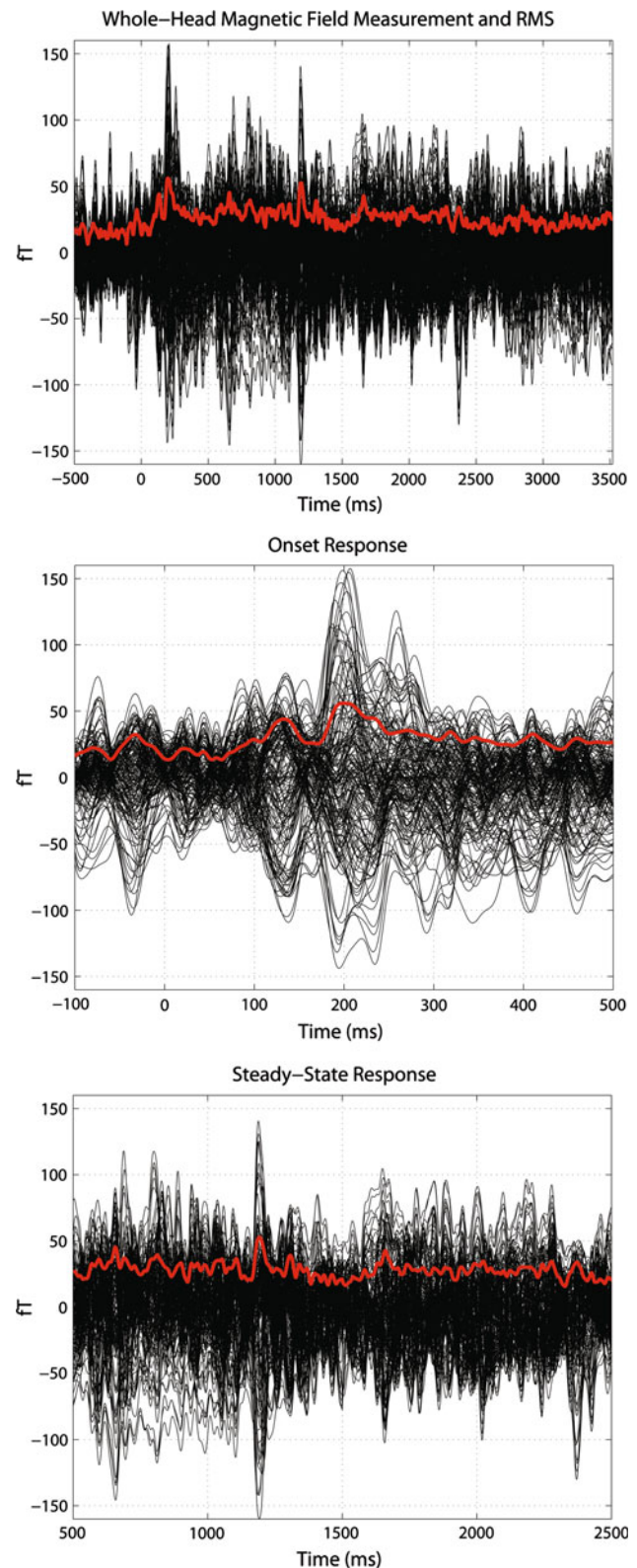
Fig. 3 *Top panel:* MEG waveforms and RMS in the temporal domain from a single participant, $\Phi = 0$ comodul condition. Time is plotted on the x-axis; field deflection in fT on the y-axis. The top panel illustrates the magnetic fields recorded from all 157 data channels. Magnetic fields are in black, RMS is in red. The time duration shown is from 500 ms pre-stimulus onset to the end of the analysis frame. This panel illustrates both the onset (~0–500 ms) and the steady-state response (towards end of onset to end of analysis frame). *Middle panel:* The middle panel provides a more detailed view of the onset response. Conventions are identical to the previous panel. Two clear peaks can be observed in the onset response at ~140 ms and ~210 ms post-stimulus onset (see “Results” for details). *Bottom panel:* Bottom panel illustrates the steady-state portion of the magnetic fields recorded. Conventions are the same as in the previous two panels. Oscillatory activity is clearly observed in the RMS of the signal, indicating entrainment to the physical structure of the comodul signals

harmonic ($F = 37.351$, $P < 0.01$). For the $\Phi = \pi$ condition, significant F ratios were found for the posterior temporal (LH: $F = 16.833$, $P < 0.01$; RH: $F = 7.358$, $P < 0.01$) and occipital sensors (LH: $F = 23.954$, $P < 0.01$; RH: $F = 12.864$, $P < 0.01$) at the modulation frequency; at the second harmonic significant F ratios were found for the occipital sensors (LH: $F = 12.663$, $P < 0.01$; RH: $F = 8.127$, $P < 0.01$) and the RH posterior temporal sensors ($F = 3.901$, $P < 0.05$).

Statistical Summary

Separate ANOVAs were calculated with the following interactions: (i) Hemisphere (two levels) \times Harmonic (two levels) \times Condition (four levels) \times Sensor Area (three levels), (ii) Harmonic \times Condition \times Sensor Area and (iii) Condition \times Sensor Area. For the first ANOVA, significant interactions were found for Harmonic ($F(1,13) = 148.053$, $P < 0.001$), Sensor Area ($F(2,13) = 134.441$, $P < 0.001$), and Condition \times Sensor Area ($F(6,13) = 4.208$, $P < 0.001$); the interaction Hemisphere \times Sensor Area was marginally significant ($F(2,13) = 3.013$, $P = 0.049$). For the second ANOVA, significant interactions were found for Harmonic ($F(1,13) = 150.546$, $P < 0.001$), Sensor Area ($F(2,13) = 136.705$, $P < 0.001$) and Condition \times Sensor Area ($F(6,13) = 4.279$, $P < 0.001$). For the third ANOVA, significant interactions were found for Sensor Area ($F(2,13) = 111.093$, $P < 0.001$) and Condition \times Sensor Area ($F(6,13) = 3.477$, $P < 0.05$).

GLMs were then implemented to statistically determine the predictors of the responses (e.g., hemisphere, harmonic, condition, sensor area). GLMs used the same factors as the ANOVAs to evaluate the response power. For the first and second set of factors the second harmonic ($P < 0.05$), occipital sensors ($P < 0.01$), and the π initial offset condition coupled with the posterior temporal sensors ($P < 0.05$) were predictors of the response power. For the third set of factors, the predictors were the posterior



temporal sensors by themselves ($P < 0.05$), the occipital sensors ($P < 0.01$) and the posterior temporal sensors coupled with the three comodul conditions ($P < 0.05$).

Fig. 4 Grand averaged squared RMS power (linear) for all participants, completely synchronous comodul condition. Left column shows left hemisphere response, right column shows right hemisphere. Rows (top to bottom) show anterior temporal, posterior temporal, and occipital sensors. Hash marks on the x -axis indicate modulation frequency and second harmonic. It is clear that there is significant activity in the posterior temporal and occipital sensors at the frequencies of interest

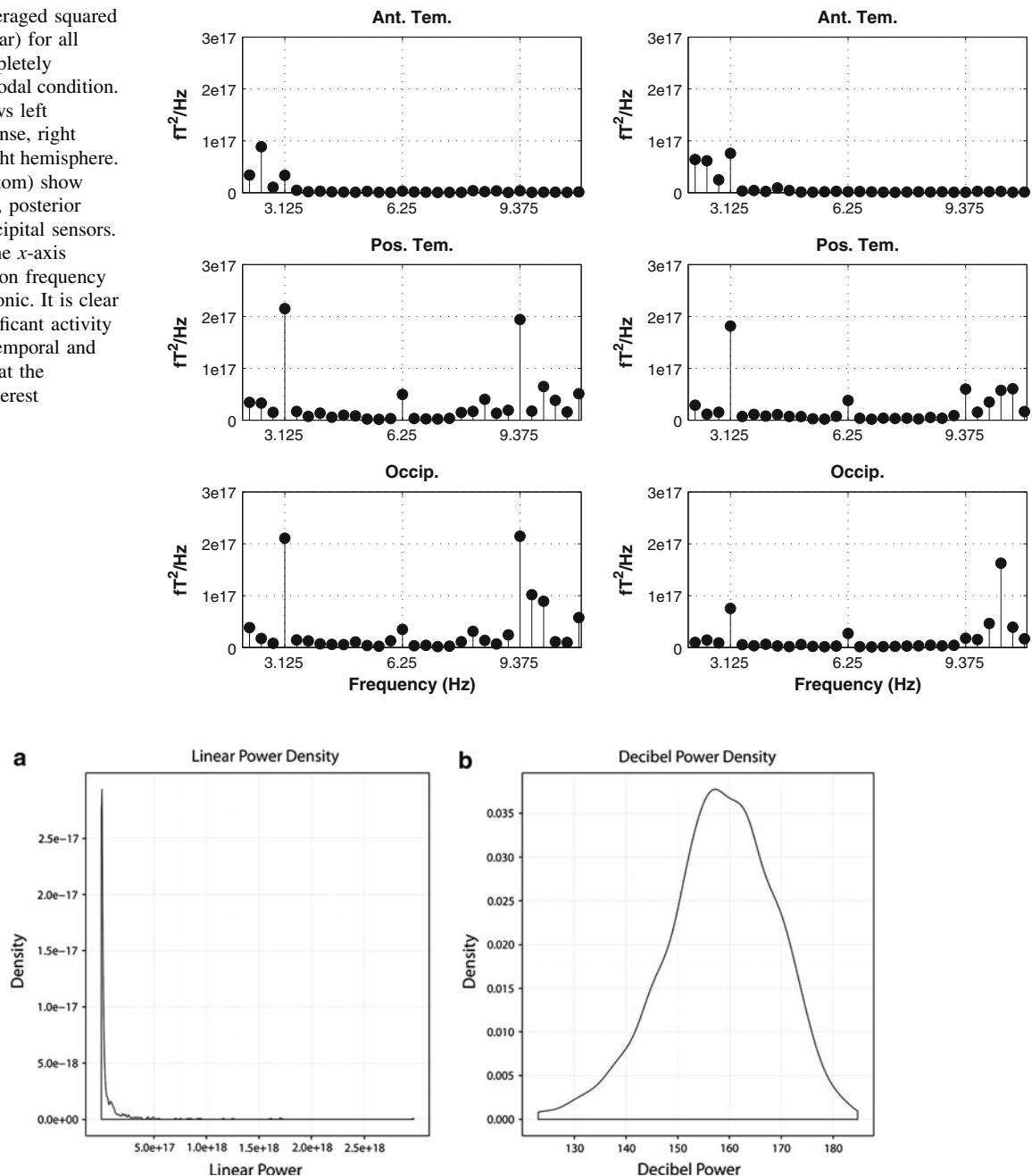


Fig. 5 Density plots for linear (a) and decibel (b) power values. Linear power values are heavily skewed to the right and the combination of large numeric values and the skewedness of the distribution make these data somewhat hard to interpret visually and statistically. Subsequent analyses focus on decibel power values. Though still somewhat skewed, the decibel power values are more

normally distributed than the linear power values, which yields to more easily interpretable visualization and statistical analysis. Additionally, representation and analysis of the data in this manner has been previously performed in the literature (Dobie and Wilson 1996) and may be more biologically plausible

Two-sample Kolmogorov–Smirnov tests indicated that the power distributions for the harmonics ($D = 0.324$, $P < 0.001$), anterior and posterior temporal sensors ($D = 0.455$, $P < 0.001$), anterior temporal and occipital sensors ($D = 0.4821$, $P < 0.001$) and posterior temporal and occipital sensors ($D = 0.134$, $P < 0.05$) differed significantly.

Post hoc analyses on the posterior temporal channels found significant interactions of Harmonic ($F(1,13) = 49.199$, $P < 0.001$; $F(1,13) = 50.157$, $P < 0.001$) and Condition ($F(3,13) = 10.103$, $P < 0.001$; $F(3,13) = 10.300$, $P < 0.001$) for the triple- and double-factor ANOVAs and Condition ($F(3,13) = 8.348$, $P < 0.001$) for

the single-factor ANOVA. GLMs indicated statistically different response power predicted by the second harmonic (triple-factor: $P < 0.05$; double-factor: $P < 0.001$) and the comodal conditions (triple-factor: $P < 0.05$; double- and single-factor: $P < 0.001$).

SSR Power Comparisons

Figure 6 illustrates the differences in overall power between harmonics for each condition for the entire data set for all sensor divisions (collapsed across hemispheres since there was no statistical difference in power between the hemispheres). Plots of the mean dB power show there is no statistical difference in power between the different conditions, but there is a difference in the power between harmonics, with the modulation frequency exhibiting greater power for each condition than the second harmonic. Additionally, though there is no statistical difference between conditions, the relational pattern of topographies observed seems commensurate with the hypotheses regarding representation of the comodal signal either as complete or separate entities (see “Discussion”).

Figure 7 illustrates the changes in response power for the posterior temporal (left panel) and occipital (right panel) sensors. Several trends can be observed. First, there is greater power at the modulation frequency than at the second

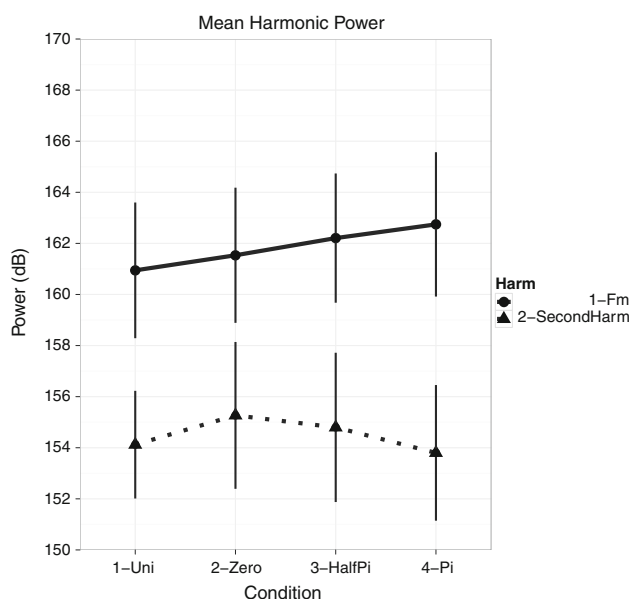


Fig. 6 Mean harmonic power at the modulation frequency (3.125 Hz) and second harmonic (6.250 Hz); experimental condition is on the abscissa and power (dB) is on the ordinate. Power is collapsed across hemispheres and sensor areas. Solid line with circles denotes the modulation frequency and dotted line with triangles denotes the second harmonic. While there is no statistical difference in power between conditions, there is a clear separation in power between harmonics, with the second harmonic exhibiting lower power values than the modulation frequency, a result typical of SSRs

harmonic. Second, the comodal conditions exhibit greater power than the unimodal conditions. Third, and most importantly, the difference in power between unimodal and comodal conditions seems to be directly attributable to the sensors overlying the posterior temporal areas (and possibly parietal lobes). No difference in power for either harmonic across conditions is observed in the occipital sensors.

Results of the Wilcoxon signed-rank tests for additivity indicated that the medians for the unimodal and comodal conditions did not differ except for $\Phi = \pi$ and unimodal modulation pairwise comparison (LH: signed-rank = 3, $Z = -3.107$; RH: signed-rank = 4, $Z = -3.045$). This difference may be due to the nature of the representation used, as the Wilcoxon signed-rank tests for additivity used the complex numbers derived from the Fourier transform and not the power values as were used in the ANOVAs and GLMs.

Figures 8 and 9 illustrate the grand average topography at the modulation frequency and the second harmonic, respectively, in the form of phasor plots, which show the sink-source distribution and the phase of the response (Simon and Wang 2005). Two clear source-sink patterns can be observed for each frequency, while for each comodal condition more complex patterns are observed, especially for the second harmonic. The sink-source distribution (and phase distribution) at the modulation frequency (Fig. 8) for all conditions resembles that of a visual response recorded from axial gradiometer sensors; this is in line with the results from the power analyses, namely that the occipital sensors generated larger responses than the anterior and posterior temporal sensors.

For the response at the second harmonic (Fig. 9), the topographies seen are more complex, as they seem to reflect the degree of AV integration. For the unimodal auditory condition, the sink-source distribution reflects responses typically recorded from auditory cortex. For the unimodal visual condition, the sink-source distribution appears to be somewhat mixed. The sink-source distribution for the comodal conditions indicates (i) the degree of synchronicity and integration between the signal components and (ii) the contribution of the posterior temporal sensors (and perhaps the auditory cortex and/or parietal lobes). For the $\Phi = 0$ condition, a clear auditory sink-source distribution is observed. For the $\Phi = \pi/2$ and $\Phi = \pi$ conditions, especially for the sensors overlying the posterior of the participants' heads, the sink-source distribution reflects the posterior auditory field topography, while for the remaining sensors the magnetic field distribution is not easily interpretable. Taken with the results of the statistical analyses, it is compelling that the changes in the response topographies and response power are due to the second harmonic and information from the posterior temporal lobes and/or auditory cortex, and possibly parietal lobes (Howard and Poeppel 2010).

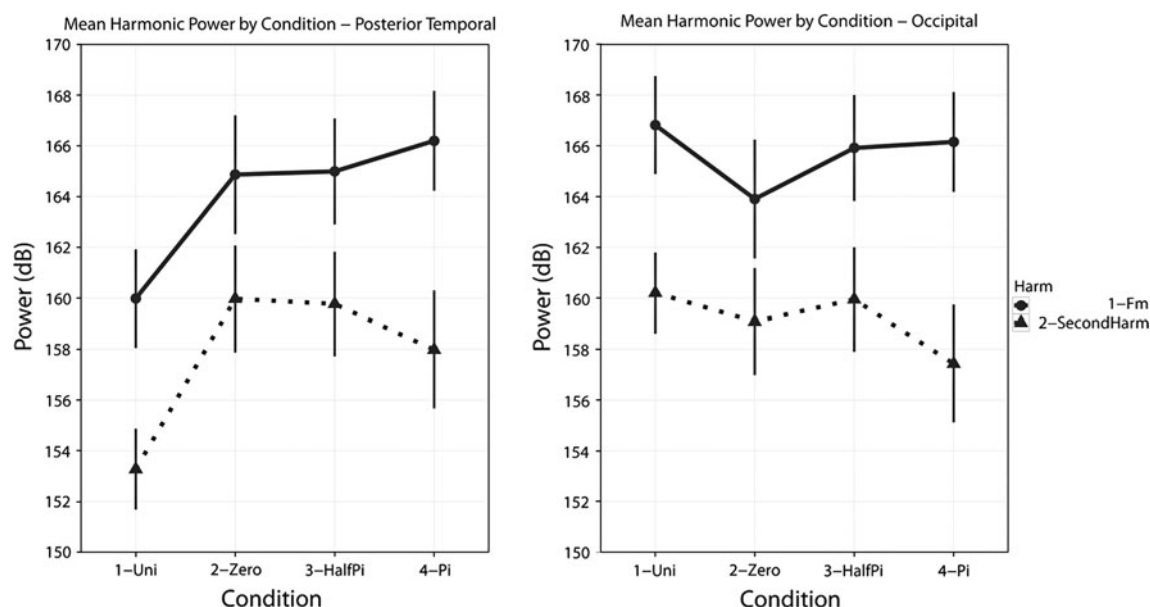
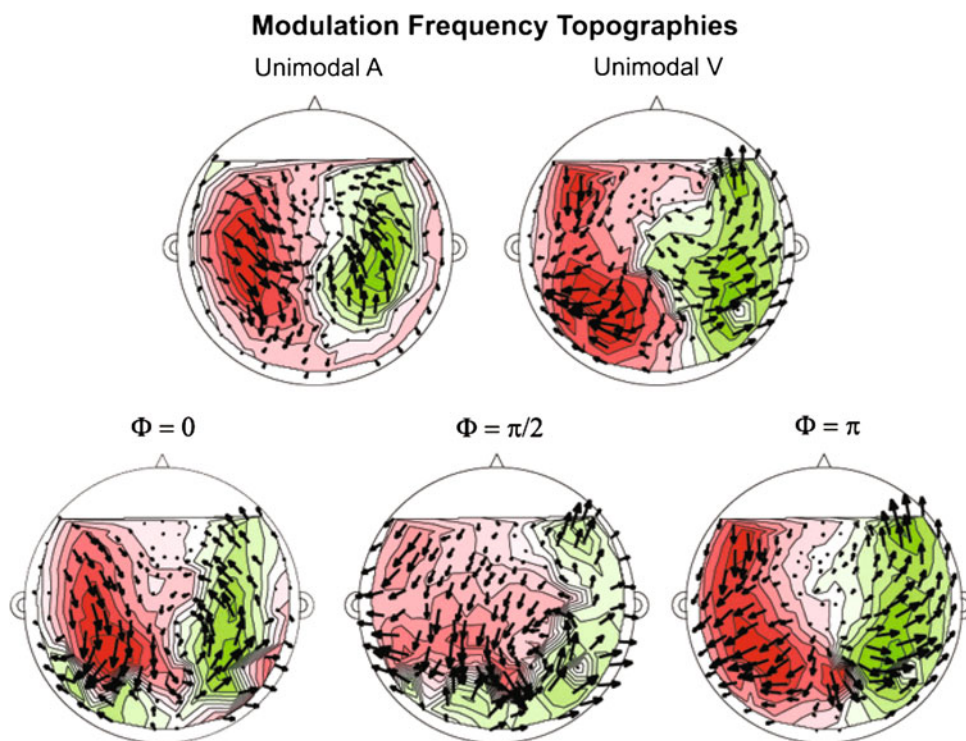


Fig. 7 *Left panel:* Mean harmonic power for the modulation frequency and second harmonic by experimental condition for the posterior temporal sensors. Conventions used are the same as in Fig. 6. Several trends can be observed: (i) response power at the second harmonic is lower than at the modulation frequency; (ii) the power for all three comodal conditions is greater than the unimodal conditions and (iii) there is no overall power difference between the comodal conditions. *Right panel:* Mean harmonic power for the

modulation frequency and second harmonic by experimental condition for the occipital sensors. For the occipital sensors, the power at the modulation frequency is greater than that at the second harmonic and the power for all conditions in the occipital sensors is greater than that of the posterior temporal sensors. For the occipital sensors, there is no statistical difference between either the unimodal or comodal conditions or the comodal conditions themselves

Fig. 8 Phasor plot of grand-averaged complex-valued topography at the modulation frequency (3.125 Hz) for each experimental condition. Top row shows unimodal conditions, bottom row comodal conditions. Magnetic source is indicated by green and magnetic sink by red. Phasors (arrows) indicate overall phase coherence and direction. For all experimental conditions, the source-sink distribution observed resembles that of a visual response as recorded by axial gradiometers. The topographies observed are in accordance with the finding that the overall visual response is greater than the auditory response, even for unimodal auditory modulation



Grand-averaged data yielded two peaks in the RMS for both unimodal modulation conditions and the comodal conditions. For the unimodal auditory condition these

peaks occurred at ~ 140 and ~ 215 ms post-stimulus onset, for the unimodal visual condition the peaks occurred at ~ 109 and ~ 204 ms. For the comodal conditions, these

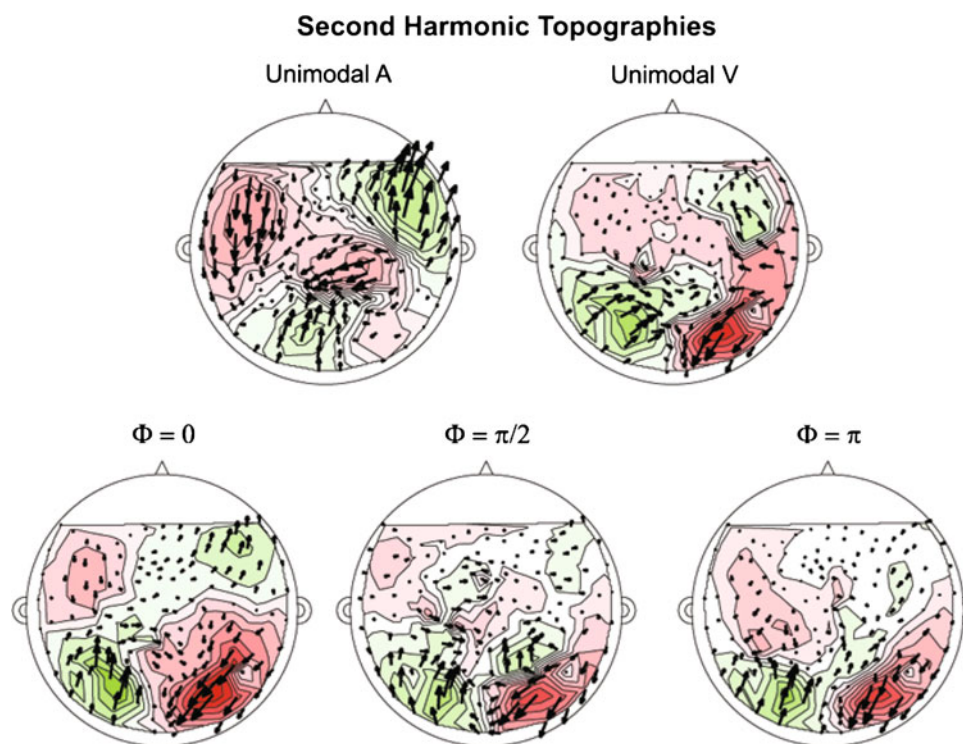


Fig. 9 Phasor plot of grand-averaged complex-valued topography at the second harmonic (6.250 Hz) for each experimental condition. Conventions used are the same as in Fig. 8. The source-sink distributions observed at the second harmonic more closely resemble that of an auditory response as recorded by axial gradiometers. For unimodal auditory modulation, the pattern observed (sink-source distribution and phasor direction and distribution) are rather clear, while for unimodal visual modulation, the topography is muted somewhat, but still observable. The source-sink distribution changes most significantly

for the comodulated conditions. For the $\Phi = 0$ condition, a clear auditory pattern is observed, while for the $\Phi = \pi/2$ and $\Phi = \pi$ conditions the topography seems to be a mix of auditory and visual activation. The constant between the three comodulated conditions is that the sink-source distributions towards the posterior end of the sensor distribution (posterior temporal and occipital sensors) resembles an auditory response topography. These plots agree well with the statistical results that the changes between conditions are predicted by the second harmonic, specifically in the posterior temporal sensors

peak latencies were ~ 140 ms for the first peak and ~ 210 – 217 ms for the second peak. These values suggested that synchronicity was also reflected in the temporal domain, because the peak latencies to comodulated conditions were very close to those observed for unimodal auditory modulation and because the statistics on the SSR power indicated significant auditory contribution to the bimodal responses. However, the permutation tests did not show significant differences in peak latencies between conditions.

Discussion

The audiovisual MEG experiment presented in this paper has (i) extended a paradigm previously used to evaluate unimodal responses to investigate bimodal responses, (ii) elicited the bimodal SSR using novel stimulus types and (iii) elucidated some of the factors affecting the neural signal recorded. In the larger context of AV experiments,

we have replicated several findings: (i) that visual contribution is greater than auditory (in the sense that the response power in visual areas is greater than in auditory areas) and (ii) when change is induced in bimodal signal components, the response in sensors overlying auditory areas changes the most, suggesting that auditory information contributes greatly when comodulated AV signals are presented, in particular when stimuli are temporally aligned across modalities.

Although this experiment contained a large number of trials and hence a high SNR, we did not find any differences between the three comodulated conditions as we initially hypothesized; however, a potential pattern of integration is borne out in the phasor plots. The data we present show effects of condition reflected in the power of the second harmonic for particular sensor areas, suggesting long-term dynamics are reflected in the first two harmonics of the SSR. While we found no statistically significant increase in signal power overall (see Fig. 6), there was a significant increase in power at the second harmonic for the comodulated

signals relative to the unimodal signals. As illustrated in Figs. 8 and 9, which show the magnetic field topographies for each harmonic in phasor plots, there is a clear difference in source-sink distribution for each harmonic. At the modulation frequency, the source-sink distribution mirrors that of a visual response, while at the second harmonic, the distribution observed mirrors that of an auditory response, depending on condition.

These topographic phasor plots (and the statistical results) suggest that the harmonics may be representing differential processing within and across modalities. The activity at the modulation frequency may reflect the modality where attention is directed or which is more salient to the observer (Talsma et al. 2006; Saupe et al. 2009; Gander et al. 2010). Second harmonic activity may reflect envelope congruency changes between modalities, which, based on the observed field patterns, may be related to the degree of statistical regularity and synchronicity in the overall signal. This response, most likely originating from auditory or parietal areas, contributes most to the neurocomputational analysis of comodul AV signals.

As mentioned previously, we did not have access to structural MRs for our participants. However, we are fairly certain of the cortical areas that are most likely generating these signals. First, the lack of MRs did not prevent getting an estimate of participant headshape (see “Materials and Methods”). As such, we had a reliable estimate of the shape and outline of each participant’s head. Second, we were able to place the location of each participant’s head within the scanner by using head marker coils; this allowed us to make sure all cortical areas of interest (based on headshape measurements) were being recorded. Lastly, special care was taken to select the sensors from domain-specific pretests. Combined, these procedures, though without anatomical constraints, assisted us in narrowing down the most likely generators of the responses observed.

Prior to executing the experiment, we had hypothesized that the signals would be represented cortically in three ways. When the signal envelopes are completely congruent, the signals may be observed and ‘computed’ as a single object. When the initial envelope phase offset is $\pi/2$ radians, then over the time course of the comodul signal, the signal components would be alternately perceived as one or two objects, as the synchronicity changes between being out-of-phase and in-phase. Lastly, when the offset is π radians between component envelopes, then each component would be perceived as a single object. As the signal component envelopes are desynchronized, the correlations and redundancies in the bimodal signal decrease, modifying the processing and representation of the percept. To verify these hypotheses, a psychophysical task would have to be incorporated along with characterization of the electrophysiological responses. Although we have

anecdotal data from experimental piloting and participant debriefing, the current data do not support these hypotheses.

The congruency potentially indexed by the phase separation in this paradigm may have practical limits. There is evidence that integration of bimodal signals, with the auditory signal leading, takes place within a 40–60 ms duration window (van Wassenhove et al. 2007). For the modulation frequencies employed here, the incongruity between signal components did not fall within this integration window. It is entirely possible that the response patterns we observe are dependent on the modulation frequency. Higher envelope modulation rates (e.g., 7–11 Hz) with phase separations falling within the temporal window of AV integration could test the SSR response to perceptually simultaneous but physically asynchronous signals.

A related issue is to sample more phase separation values around the entire unit circle. One possible hypothesis is that the representation of the phase separation will be symmetric (except when both signal envelopes are completely synchronized), i.e. the response power for a phase separation of $\pi/2$ radians and $3\pi/2$ radians will be represented equally. The indexing of signal component congruity might also be dependent on which component reaches the maximum of the envelope first. It has been shown that when visual information precedes auditory information, signal detection and comprehension increases (Senkowski et al. 2007; van Wassenhove et al. 2007). In the current study, for the asynchronous bimodal conditions, the auditory component of the signal reached the maximum of the modulation envelope first. It would be useful to examine the interactions that occur when the visual component modulation envelope reaches the maximum value before the auditory envelope.

Adding ‘jitter’ or noise to the signal component envelopes may also yield a more ecologically valid set of stimuli for further experimentation. This would add the variability inherent in speech, while retaining the modulation information of the signal component envelopes. Finally, the modulation depth of the auditory signal component might be made more variable to correspond with the conditions occurring in natural human speech, where the mouth opens and closes fully (modulation depth ranging from 0 to 100%).

Much in the same way as traditional unimodal steady state responses are used to probe auditory and visual function, it may be possible to use the paradigm we introduce to assess audiovisual integration in humans. Deviations from the 40 or 80 Hz aSSR response have been suggested to correlate with impairments in CN VIII, the brainstem, or possibly cortical processing (Valdes et al. 1997). Application of this paradigm could be used as a clinical assessment of audiovisual integration.

In summary, we demonstrate that an experimental technique commonly applied to unimodal signals, the SSR, can be applied to signals of a bimodal nature approximating the spectro-temporal properties of speech. We observed that the presence of bimodal information increased response strength in auditory areas.

Our findings are in line with several studies regarding AV integration, especially with regard to the specific contributions of auditory information (Cappe et al. 2010). In a real-world stimulus analogous to our ‘noislipses’, Chandrasekaran et al. (2009) characterized bimodal speech stimuli (with no phase incongruities) and observed (i) a temporal correspondence between mouth opening and the auditory signal component envelope and (ii) mouth openings and vocal envelopes are modulated in the 2–7 Hz frequency range. That modulation frequencies in this range play a key neurophysiological role in the parsing of neurophysiological signals has now been amply demonstrated. For example, Luo et al. (2010) show that audio-visual movies incorporating conversational speech bear a unique signature in the delta and theta neural response bands, values congruent with the Chandrasekaran behavioral data. Cumulatively, it is now rather clear that low-modulation frequency information lies at the basis of analyzing uni- and multimodal signals that have extended temporal structure. The results of this MEG study offer additional support for this claim, and future iterations of this paradigm could further elucidate the neural computations underlying multisensory perception of ecologically relevant stimuli.

Acknowledgments This project originated with a series of important discussions with Ken W. Grant (Auditory-Visual Speech Recognition Laboratory, Army Audiology and Speech Center, Walter Reed Army Medical Center). The authors would like to thank him for his extensive contributions to the conception of this work. The authors would like to thank Mary F. Howard and Philip J. Monahan for critical reviews of this manuscript. We would also like to thank Jeff Walker for technical assistance in data collection and Pedro Alcocer and Diogo Almeida for assistance with various R packages. This work was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health: 2R01DC05660 to DP, JZS, and WJI and Training Grant DC-00046 support to JJIII and AER. Parts of this manuscript comprise portions of the first two authors’ doctoral theses.

References

- Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ (2005) Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res* 166: 559–571
- Baayen RH (2008) languageR: data sets and functions with “Analyzing Linguistic Data: a practical introduction to statistics”. R package version 0.953
- Baumann O, Greenlee MW (2007) Neural correlates of coherent audiovisual motion perception. *Cereb Cortex* 17:1433–1443
- Besle J, Fort A, Delpuech C, Giard MH (2004) Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur J Neurosci* 20:2225–2234
- Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS (1999) Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport* 10:2619–2623
- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10:649–657
- Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001) Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage* 14:427–438
- Campbell R (2008) The processing of audio-visual speech: empirical and neural bases. *Philos Trans R Soc Lond B Biol Sci* 363:1001–1010
- Cappe C, Thut G, Romei V, Murray MM (2010) Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. *J Neurosci* 30:12572–12580
- Chandrasekaran C, Trubanova A, Stillitano S, Caplier A, Ghazanfar AA (2009) The natural statistics of audiovisual speech. *PLoS Comput Biol* 5:e1000436
- de Cheveigné A, Simon JZ (2007) Denoising based on time-shift PCA. *J Neurosci Methods* 165:297–305
- Dobie RA, Wilson MJ (1996) A comparison of *t*-test, *F*-test, and coherence methods of detecting steady-state auditory-evoked potentials, distortion product otoacoustic emissions, or other sinusoids. *J Acoust Soc Am* 100:2236–2246
- Driver J, Spence C (1998) Crossmodal attention. *Curr Opin Neurobiol* 8:245–253
- Fisher NI (1996) Statistical analysis of circular data. Cambridge University Press, Cambridge
- Fort A, Delpuech C, Pernier J, Giard M-H (2002) Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cereb Cortex* 12:1031–1039
- Gander PE, Bosnyak DJ, Roberts LE (2010) Evidence for modality-specific but not frequency specific modulation of human primary auditory cortex by attention. *Hear Res* 268:213–226
- Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? *Trends Cogn Sci* 10:278–285
- Grant KW, Seitz PF (2000) The use of visible speech cues for improving auditory detection of spoken sentences. *J Acoust Soc Am* 108:1197–1208
- Hershenson M (1962) Reaction time as a measure of intersensory facilitation. *J Exp Psychol* 63:289
- Howard MF, Poeppel D (2010) Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *J Neurophysiol* 105(5):2500–2511
- Jenkins J III, Idsardi WJ, Poeppel D (2010) The analysis of simple and complex auditory signals in human auditory cortex: magnetoencephalographic evidence from M100 modulation. *Ear Hear* 31:515–526
- Jones EG, Powell TP (1970) An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain* 93:793–820
- Kayser C, Petkov CI, Logothetis NK (2008) Visual modulation of neurons in auditory cortex. *Cereb Cortex* 18:1560–1574
- Kelly SP, Gomez-Ramirez M, Foxe JJ (2008) Spatial attention modulates initial afferent activity in human primary visual cortex. *Cereb Cortex* 18:2629–2636
- Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320:110

- Lalor EC, Kelly SP, Pearlmutter BA, Reilly RB, Foxe JJ (2007) Isolating endogenous visuo-spatial attentional effects using the novel visual-evoked spread spectrum analysis (VESPA) technique. *Eur J Neurosci* 26:3536–3542
- Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54:1001–1010
- Luo H, Wang Y, Poeppel D, Simon JZ (2006) Concurrent encoding of frequency and amplitude modulation in human auditory cortex: MEG evidence. *J Neurophysiol* 96:2712–2723
- Luo H, Liu Z, Poeppel D (2010) Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol* 8:e1000445
- Macaluso E, Driver J (2005) Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends Neurosci* 28:264–271
- MATLAB (2009) Version R2009a. The Mathworks, Natick, MA
- Mesulam MM (1998) From sensation to cognition. *Brain* 121:1013–1052
- Miller BT, D'Esposito M (2005) Searching for “the Top” in top-down control. *Neuron* 48:535–538
- Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cognitive Brain Res* 14:115–128
- Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. *Cereb Cortex* 14:452–465
- Molholm S, Sehatpour P, Mehta AD, Shpaner M, Gomez-Ramirez M, Ortigue S, Dyke JP, Schwartz TH, Foxe JJ (2006) Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. *J Neurophysiol* 96:721–729
- Molholm S, Martinez A, Shpaner M, Foxe JJ (2007) Object-based attention is multisensory: co-activation of an object's representations in ignored sensory modalities. *Eur J Neurosci* 26:499–509
- Müller MM, Teder W, Hillyard SA (1997) Magnetoencephalographic recording of steady-state visual evoked cortical activity. *Brain Topogr* 9:163–168
- Murray MM, Foxe JJ, Wylie GR (2005) The brain uses single-trial multisensory memories to discriminate without awareness. *NeuroImage* 27:473–478
- Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9:97–113
- Olson IR, Gatenby JC, Gore JC (2002) A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Cognitive Brain Res* 14:129–138
- Picton T, John M, Dimitrijevic A, Purcell D (2003) Human auditory steady-state responses. *Int J Audiol* 42:177–219
- R computer program, Version 2.10.1. R Foundation for Statistical Computing, Vienna, Austria (2009)
- Ross B, Borgmann C, Draganova R, Roberts LE, Pantev C (2000) A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude modulated tones. *J Acoust Soc Am* 108:679–691
- Saupe K, Widmann A, Bendixen A, Müller MM, Schröger E (2009) Effects of intermodal attention on the auditory steady-state response and the event related potential. *Psychophysiology* 46:321–327
- Schroeder CE, Foxe J (2005) Multisensory contributions to low-level, ‘unisensory’ processing. *Curr Opin Neurobiol* 15:454–458
- Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci* 32:9–18
- Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillations and visual amplification of speech. *Trends Cogn Sci* 12:106–113
- Senkowski D, Molholm S, Gomez-Ramirez M, Foxe JJ (2006) Oscillatory beta activity predicts response speed during a multisensory audiovisual reaction time task: a high-density electrical mapping study. *Cereb Cortex* 16:1556–1565
- Senkowski D, Talsma D, Grigutsch M, Herrmann CS, Woldorff MG (2007) Good times for multisensory integration: effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia* 45:561–571
- Senkowski D, Schneider TR, Foxe JJ, Engel AK (2008) Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci* 31:401–409
- Simon JZ, Wang Y (2005) Fully complex magnetoencephalography. *J Neurosci Methods* 149:64–73
- Sohmer H, Pratt H, Kinarti R (1977) Sources of frequency following response (FFR) in man. *Electroencephalogr Clin Neurophysiol* 42:656–664
- Steeneken HJM, Houtgast T (1980) A physical method for measuring speech-transmission quality. *J Acoust Soc Am* 67:318–326
- Stein BE, Meredith MA, Huneycutt WS, McDade L (1989) Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. *J Cogn Neurosci* 1:12–24
- Sumby WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26:212–215
- Talsma D, Doty TJ, Stowd R, Woldorff MG (2006) Attentional capacity for processing concurrent stimuli is larger across sensory modalities than within a modality. *Psychophysiology* 43:541–549
- Valdes JL, Perez-Abalo MC, Martin V, Savio G, Sierra C, Rodriguez E, Lins O (1997) Comparison of statistical indicators for the automatic detection of 80 Hz auditory steady state responses. *Ear Hear* 18:420–429
- van Wassenhove V, Grant KW, Poeppel D (2007) Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* 45:598–607
- Wickham H (2009) ggplot2: elegant graphics for data analysis: Springer, New York