The elicitation of audiovisual steady-state responses: multi-sensory signal congruity and phase effects

Julian Jenkins III[1,3], Ariane E. Rhone[2,3], William J. Idsardi[2,3], and David Poeppel[4]

1. Department of Biology, University of Maryland, College Park
2. Department of Linguistics, University of Maryland, College Park
3. Cognitive Neuroscience of Language Laboratory, University of Maryland, College Park
4. Department of Psychology, New York University

Address for correspondence:

Julian Jenkins III,M.S.
Department of Biology
University of Maryland, College Park
College Park, MD, USA 20742
julianj@umd.edu

Running head: Bimodal SSR in MEG

Abstract

Most ecologically natural sensory experiences are not limited to a single modality. While it is possible to use real ecological materials as experimental stimuli, parametric control of such tokens is limited. By using artificial bimodal stimuli composed of approximations to ecological signals, it can be possible 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. We quantify the responses to a bimodal audio-visual signal with different degrees of temporal (phase) congruity, focusing on properties critical to audiovisual speech. An amplitude modulated auditory signal ('pseudo-envelope') is paired with a radius-modulated disc ('pseudo-mouth'), with the low-frequency modulations occurring in phase or at offset phase values. 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 differentially reflects the congruity between signal components. 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

Introduction

The majority of sensory experiences are not limited to a single modality and thus require the observer to not only 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 A et al., 2005; Kelly SP et al., 2008; Lalor EC et al., 2007; Macaluso E and J Driver, 2005; Miller BT and M D'Esposito, 2005; Molholm S et al., 2007; Molholm S et al., 2004; Molholm S et al., 2002; Murray MM et al., 2005; Senkowski D et al., 2006). The ability to integrate information not only unifies the perception of events, but the presence of redundant information also facilitates recognition, increases signal-to-noise ratio and decreases reaction times to cross-modal events (Driver J and C Spence, 1998; Hershenson M, 1962; Senkowski D *et al.*, 2006; Stein BE et al., 1989). Studies examining the simultaneous serial and parallel computations and physiological responses underlying the integration of information and the cognition 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 O and MW Greenlee, 2007; Lakatos P et al., 2008; Miller BT and M D'Esposito, 2005; Schroeder CE and P Lakatos, 2009; Schroeder CE et al., 2008)

While it has traditionally been thought that processing of cross-modal events occurs primarily in association cortices (Jones EG and TP Powell, 1970; Mesulam MM, 1998), recent evidence indicates that information from other sensory modalities can influence

cortical areas conventionally assumed to be unimodal. Electroencephalographic (EEG), functional magnetic resonance (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 & Foxe (Schroeder CE and J Foxe, 2005) for a review). Intracranial recordings and anatomical tracings in macaques have affirmed the existence of multisensory inputs to unimodal cortical areas (Kayser C 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 GA et al., 1999; Calvert GA et al., 2000; Calvert GA et al., 2001; Molholm S *et al.*, 2004; Molholm S et al., 2006; Senkowski D 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 GA *et al.*, 2001). With regard to speech, the traditional speech areas (perisylvian) have been implicated as well as superior parietal, inferior parietal, inferior frontal, superior temporal sulcus and left claustrum (Calvert GA *et al.*, 2000; Campbell R, 2008; Fort A et al., 2002; Olson IR et al., 2002).These findings also 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, with the hypothesis that the decrease in reaction time and facilitation of object recognition should be visible in electrophysiological data. These studies have found significant activity within several latency windows, with the most

surprising results for audio-visual interactions coming at ~50 ms post-stimulus onset, suggesting extremely early processing of audiovisual interactions (Molholm S *et al.*, 2002). In addition, several ERP studies have also evaluated facilitation of bimodal interactions via an additive model (Besle J 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 physical/spectral properties of a modulated stimulus. This response has been found for both visual and auditory signals and has been used extensively for clinical and diagnostic purposes (Sohmer H et al., 1977). Auditory SSRs are generally elicited by amplitude or frequency modulated signals (e.g. (Luo H 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 as well (Müller MM et al., 1997; Ross B 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 KW and PF 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 CE and P Lakatos, 2009; Schroeder CE *et al.*, 2008; Senkowski D *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 (Senkowski D *et al.*, 2008).

The motivation for the current study was to model an ecologically valid audio-visual interaction, namely speech, using artificial signals that incorporate some critical attributes of a multi-sensory speech. The auditory component of speech consists of the frequency and fine spectral components as well as 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 HJM and T 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. Recent MEG evidence supports this generalization: Luo & Poeppel (Luo H and D Poeppel, 2007) and Howard & Poeppel (Howard MF and D Poeppel, 2010) observed that fluctuations in the speech envelope are associated with intrinsic oscillations in the theta frequency band (~4 – 8 Hz). Paired with the auditory signal is a visual component in which facial features -- and especially mouth movements -- aid comprehension, especially in noisy environments (Sumby WH and I Pollack, 1954). We 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 this experiment, the visual signal consists of a

size-modulated disc (to approximate a mouth opening and closing), and the auditory signal consists of an amplitude-modulated sine wave (to approximate the envelope). We hypothesize that that the SSRs elicited by congruent 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 the second, third, and fourth harmonics. The increased signal power of the comodulated conditions relative to unimodal conditions might lead to 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.

Materials and Methods

Participants: Thirteen right-handed (Oldfield RC, 1971)adult subjects (seven female) with normal hearing and normal or corrected-to-normal vision underwent MEG scanning. One person's data set was excluded from all analyses due to insufficient signal-to-noise ratio for all experimental conditions. Age range was 18-41 (mean 27.08 years). Participants were either compensated for their participation or earned course credit in an introductory linguistics course. 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: The experimental stimuli consisted of five types of audio-visual signals presented at two modulation frequencies, for a total of ten signals (Figure 1). The five types were: i) amplitude-modulated sine waves presented concurrently with a static white square on black background; ii) a radius-modulated white disc on black background presented with approximately Gaussian white noise; iii) a radius-modulated disc and an amplitude modulated sine wave at one of three phase relationships (in phase, π/2 radians out of phase, π radians out of phase). The amplitude-modulated sine waves and radius-modulated discs were modulated at either 2.5 Hz or 3.7 Hz with a modulation depth of 24 percent. These values, a little bit lower than the peak of the modulation spectrum for spoken language, were chosen after extensive piloting revealed that higher visual modulation frequencies were very uncomfortable for participants to view for extended periods of time. Two frequencies were chosen to replicate any effects at different, not harmonically related modulation frequencies. The stimuli were four seconds 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.

Figure 1 about here

Auditory signal components were generated with Matlab (v2007b, The Mathworks, Natick, MA) and consisted of a sine wave envelope (either 2.5 Hz or 3.7 Hz modulation frequency) applied to an 800 Hz sine wave carrier signal with 6 ms $\cos^2$ onset and offset ramps presented at approximately 65 dB SPL. The signals were sampled at 44.1

kHz with 16-bit resolution. Signals were generated using the sine, not the cosine function, to eliminate undesired phase effects on onset responses (see below). Visual signal components were generated using GIMP (www.gimp.org). The radius-modulated white discs were centered on a 640 x 480 pixel black background, and ranged from 2.5° visual angle at the minimum diameter and 4° visual angle for the maximum diameter. The individual frames were compiled into .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.

Experimental stimuli were presented in nine blocks, with three repetitions per signal per block. Presentation of conditions was randomized within blocks. The experimental materials were passively attended to; no response to the signals was required. In order to maintain vigilance, a distracter task was incorporated into the experiment. An audio-visual signal (500 or 1500 ms duration) consisting of a crosshair on a black background combined with approximately Gaussian white noise was used as the target and was pseudorandomly presented with the signals (~17% of total trials). Subjects had to press a button in response to the crosshair/noise target; these trials were excluded from analysis.

Delivery: All experimental stimuli were presented using a Dell Optiplex computer with a SoundMAX Integrated HD sound card (Analog Devices, Norwood, MA) via Presentation

stimulus presentation software (Neurobehavioral Systems, Inc., Albany, CA). Stimuli were delivered to the subjects binaurally via Eartone ER3A transducers and non-magnetic air-tube delivery (Etymotic, Oak Brook, IL). The inter-stimulus interval varied pseudo-randomly between 2500 and 3500 ms.

Recording: 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. Data were noise reduced using time-shifted PCA (de Cheveigné A and JZ Simon, 2007) trials averaged offline (artifact rejection ± 2.5 pT), bandpass filtered between .03 - 25 Hz (161 point Hamming window) and baseline corrected over the 700 ms pre-stimulus interval.

Data Analysis

The analysis was performed in sensor space, not source space, to stay as close as possible to the recorded data without making source configuration assumptions. All analyses -- pre-experiment localization parameters, waveform assessment, and the calculation of the magnitude and phase of the SSR as well as significance values -- were performed in Matlab. Statistical analysis of SSR parameters was evaluated using the statistical and probability distribution functions in Matlab's Statistics Toolbox.

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 800 Hz sinusoidal carrier signal,

modulation frequency 7 Hz, modulation depth 100 percent and 11.3 second duration. The visual pre-test consisted of a checkerboard flicker pattern (Fm = 4 Hz), of 240 second 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 Figure 1c). 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 auditory pre-test response was characterized using two distinct methods. The first analysis examined the power spectral density (PSD) of the response and selected the channels with the best (strongest) response (Fourier Transform window: 3 to 5 s), at the modulation frequency. The second analysis examined the maximum field deflection of the M100 response (search window: 80 to 130 ms after stimulus onset) and selected the channels with the maximum response amplitude (both source and sink). The pre-test visual response was characterized only using the PSD, at twice the modulation frequency (the reversal rate), due to the checkerboard pattern not generating a robust onset response permitting an onset response analysis. Since the data were analyzed in sensor space rather than source space, special care was taken to avoid having posterior temporal and occipital sensors overlap. When posterior temporal and occipital

sensors were common to each modality/sensor area, those particular posterior temporal sensors were replaced by the next non-overlapping posterior temporal sensors.

Onset response evaluation and PCA: The signal evaluation window (averaged and filtered sensor data) ranged from 700 ms pre-trigger to 3999 ms post-trigger. Onset peak root-mean-square (RMS), RMS latency, magnetic field deflection and magnetic field deflection latency responses corresponding to the M100 (auditory; search window: 80 to 130 ms after stimulus onset) and M170 (visual; 145 to 195 ms after stimulus onset) for each hemisphere for each condition were collected and averaged across subjects for each stimulus and were plotted topographically to examine the response. The minimum number of trials averaged was twelve and the maximum number was twenty-seven. Since it is hypothesized that the neurophysiological response primarily reflects processing of the envelope for both signal onset and SSRs, an estimation of the envelope was made using principal components analysis (PCA). The preselected sensors were analyzed using PCA and the envelope estimate was calculated using the absolute value of the Hilbert transform for the first principal component, which explained 60 to 80 percent of the total variance, depending on the participant. The channels used for the latency and envelope analysis for the anterior and posterior temporal channels are those from the second analysis (onset analysis described above) of the auditory pre-test data. Congruence between the two sets of pretest channel data was approximately 90%. Occipital channels used were the same as in the pre-test.

SSR analysis: The magnitude and phase spectra of the SSR were determined using the Fast Fourier Transform (FFT) of the baseline corrected and filtered channel data. The FFT was calculated from stimulus onset (0 ms) to the end of the signal evaluation window (3999 ms). Prior to the calculation of the FT, the data within the signal evaluation window was multiplied by a Kaiser window (length 4000 samples, beta = 13) to minimize the onset and offset responses to the audio-visual signals and minimize spurious frequency contributions. The magnitude of the response was calculated using the RMS of the FT 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 subject response power was determined by calculating the mean of the individual subject power vectors. To determine the across subject phase response, the mean direction of the individual mean directions was calculated. The trials analyzed for the SSR analysis were the same as those analyzed for the onset responses. Figure 2 illustrates waveform recordings for pre-windowed data over the entire analysis frame, the onset responses in detail, and the sensor layout.

Figure 2 about here

SSR cross-modal control analysis: To determine the validity of the sensor selection from the pre-experiment localization, unimodal modulation data were analyzed using the sensors from the other modality. This analysis confirmed that the responses recorded from the unimodal modulation truly reflected that particular modality. This particular analysis does not necessarily indicate that the unimodal visual condition had an effect

on the unimodal auditory condition whereas the converse was not true; rather it means that the neurophysiological signals generated and recorded were truly present in the recorded magnetic field data.

Across-subject response averaging: The across-subject responses were computed by collecting the individual subject field deflections (source and sink field deflections and RMS) and calculating the mean response amplitudes and the RMS of the subject RMS values. The aggregate waveforms peaks and latencies were characterized in the same search windows as described above. However, the data were not subject to windowing to preserve the onset responses. A similar procedure was used for the Fourier transformed data. Individual subject vectors for response power (squared magnitude) and phase were collected the relevant statistics calculated.

Statistical analyses: To assess the effect of signal manipulation on the underlying neurophysiological computations, the mean latency and peak values (for both magnetic field sink and source deflections and RMS) for onset responses were analyzed using mixed measures ANOVA (SPSS 16.0, SPSS Inc., Chicago, IL). A full factorial design was employed, with amplitude (peak RMS and maximum response) and latency as the dependent measures and Hemisphere (RH vs. LH), Frequency (2.5 Hz vs. 3.7 Hz) and Phase (in phase, π/2 radians out of phase, π radians out of phase) as factors. Planned comparisons using Wilcoxon sign rank tests compared unimodal modulation against each comodulated condition for i) peak RMS, ii) peak RMS latency, iii) peak source and sink deflection, iv) source and sink deflection latency, v) envelope onset period, vi) peak

of envelope onset and vi) envelope periodicity if and only if the ANOVA results were found to be significant.

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 (Dobie & Wilson 1996). The signal evaluation window used gave a frequency resolution of 0.25 Hz and gave the exact response at Fm = 2.5 Hz, but not at 3.7 Hz. To evaluate the response at Fm = 3.7 Hz, the bin next closest in frequency (3.75 Hz) was used. The significance of the phase of the response was assessed using Rayleigh's phase coherence test on the mean direction (Fisher NI, 1996). Individual subject responses at each modulation frequency for each condition were assessed using an *F* test to determine if the response was significant and whether or not a particular subject should be excluded due to lack of a response or exhibiting a response other than at the modulation frequencies and harmonics of interest. For the across-subject data, *F* tests were performed on the power of the SSR at the modulation frequency, two subharmonics in the delta band, one harmonic in the theta band and one in the alpha band (see e.g., Jones & Powell 1970; Senkowski et al. 2008 for review of frequency band descriptions). The power at individual harmonic components of the modulation frequency in different frequency bands across conditions was compared using Wilcoxon sign rank tests (Matlab v7). Two sets of sign rank tests were performed: the first compared the mean unimodal modulation magnitudes against the mean comodal modulation magnitudes for a given sensor area (e.g. LH anterior temporal unimodal auditory vs. LH anterior temporal comodal, phi = π) and the second compared the

comodulated conditions (e.g. RH occipital, phi = zero vs. RH occipital, phi = π/2. A mixed effects ANOVA implemented in R (Baayen 2008; R Development Core Team 2008) assessed any possible differences in modulation frequency and hemisphere.

Dipole estimation: SSR source estimation was performed on data from seven subjects who exhibited a SSR response. Since we did not have structural MR images for our participants, source estimations were not anatomically constrained. Single equivalent current dipole estimates with a GOF < 80% and not localized to the hemisphere from which the channels were selected were excluded from subsequent statistical analyses. A simple spherical head model was used to determine the source of the SSR (x,y,z axes) as well as the dipole angles (theta and phi) using 35 sensors per hemisphere with the greatest PSD (ten channels each from anterior and temporal sensor divisions; fifteen from occipital sensor divisions per hemisphere). The sensors selected came from both the auditory and visual sensor division described previously. To perform the fit to the desired components found significant by the *F* test (see *Results*), the real part of the Fourier transformed data for the component of interest was multiplied by $\cos(2*\pi*F*t)$ and the imaginary part of the data by $\sin(2*\pi*F*t)$, where F and t are the frequency of interest and the time vector, respectively. The resulting vectors were then algebraically added and fits were performed on the peaks of the variance of the resulting sinusoidal waveform. Peaks corresponding to the magnetic field topography reflected by the response magnitudes were used for the source estimation (see *Results*). Dipole estimations from a single peak was recorded and entered for subsequent statistical analysis. Statistical significance of the values of x,y,z, theta and

phi were assessed using a mixed effects ANOVA in R using the 'languageR' statistical package. Wilcoxon tests were performed on the values of theta and phi both across and within subjects to assess any potential differences in the source orientation of the SSR (R 2.81).

Results

Figure 3 illustrates the response for one participant for both unimodal modulation conditions, Fm = 2.5 Hz. This characteristic pattern demonstrates that the majority of evoked activity occurs primarily at the modulation frequency. The least amount of response power was observed to be in the sensors overlying the anterior temporal lobe (analyzed for auditory alone condition only), with greater magnitude in the posterior temporal sensors for auditory alone and for the occipital sensors in visual alone unimodal conditions. The overall response fit well with that of a prototypical SSR, namely the response was elicited robustly at the modulation frequency and occasionally visible at the first few harmonics.

Figure 3 about here

The response topography for a representative subject is illustrated in Figure 4. The complex-valued magnetic field response profile reflects the SSR response power at the modulation frequency for each condition, Fm = 2.5 Hz, as measured at one peak in the sinusoidal waveform used for dipole localization (discussion below). The topography

reflects the whole-head response power measured for each condition. For the unimodal modulation conditions, the whole-head response resembles the topography of a visual response. This is congruent with the observation that the visual response was greater than the auditory response in the unimodal modulation conditions (see Figure 3 and Figure 6). The bimodal condition magnetic field topographies reflect the combined auditory and visual cortical computations underlying their generation. The increasing contribution of auditory cortex to cortical processing of the bimodal signals can be seen in the spatial configuration of the magnetic fields recorded. The contribution of the auditory cortex can be observed most clearly when the signal component envelopes are initially orthogonal to one another (Figure 4d; $\Phi = \pi/2$). Magnetic field sink-source orientation reverses at each peak for this condition.

Figure 4 about here

Across-Subject Power Analysis

Figure 5 displays the across subject response power for Fm = 3.7 Hz, plotted with a linear scale for frequency and a logarithmic scale for response power, shown here for right hemisphere sensors only. Across conditions it is evident that the anterior channels capture the underlying SSR activity in a less compelling manner; posterior temporal and occipital channels, on the other hand, reveal very clear patterns. Response power was concentrated at the modulation frequency and the second harmonic, with some activity centered also around 10 Hz, as was clear for the representative subject shown above. Figure 5 shows the grand-averaged response power across subjects for the right

hemisphere only, $\Phi = 0$, $\Phi = \pi/2$ and $\Phi = \pi$, Fm = 3.7 Hz. Response power significance for all bimodal conditions (as determined by Wilcoxon sign-rank tests) compared to the unimodal modulation conditions show that the responses are significantly greater in bimodal than unimodal responses at the frequencies found significant by the ANOVA.

Figure 5 about here

Several observations merit highlighting. First, the majority of the activity is reflected in the sensors overlying the posterior temporal lobes and occipital lobes. Second, for the AV comodulated condition in which the signal envelopes are at the same initial phase, the response power is greatest at the modulation frequency, localized to the sensors overlying the posterior temporal lobes. Third, as the difference in the relative phase increases, the response power decreases, although the response is still greater than that of unimodal modulation condition (see Figure 6). This change is indexed primarily by the response power as measured at the modulation frequency.

Statistical summary

The significance of the SSR was calculated at the modulation frequency, as well as two subharmonics, and the second and third harmonics. Significance was determined by means of an *F* test on the power of the SSR at each particular frequency as described by Dobie and Wilson ((Dobie RA and MJ Wilson, 1996) - see *Methods*). All subjects elicited a statistically significant response for the SSR at each envelope modulation frequency. Within-subject response significance was restricted to evaluation at the

modulation frequency (see *Methods*) with degrees of freedom (df) df1 = 2, df2 = 12 and α = 0.05. The across-subject significance for subharmonics was assessed using *df* = 2,4 and significance for the modulation frequency and second and third harmonics were assessed using *df* = 2,12.

SSR power at subharmonics was not found to be statistically significant. Statistically significant responses were observed at the modulation frequency, as well as second and third harmonics. The difference between the observed statistical significance for subharmonics and the second and third harmonics may be attributable to the decreased degrees of freedom for df2.

Results of Rayleigh's test on the mean direction of the SSR vectors (at the frequencies observed to be significant by the *F* test) found the phase angle directions to be statistically significant at α = 0.05

Figure 6 about here

SSR power comparisons

Table 1 summarizes the SSR power for each modulation frequency, modulation condition and sensor/cortical area as well as the interactions found to be significantly significant as a result of Wilcoxon sign-rank tests (see *Methods*). For both envelope modulation frequencies, several statistically significant responses are held in common. First, both modulation frequencies exhibit statistically significant responses power at the

modulation frequency for all comodulated conditions and this interaction is largely limited to the sensors overlying the posterior temporal lobe for both hemispheres. Second, there were significant interactions at the second harmonic for $\Phi = 0$ and $\Phi = \pi$; both modulation frequencies held this interaction in common in the LH sensors overlying the posterior temporal lobe. One last interaction was common to both modulation frequencies for the third harmonic for $\Phi = 0$ in the LH sensors overlying the occipital lobe. Several other statistically significant interactions were found to be unique to each modulation frequency; these perhaps inconsistent interactions may be a result of data variance (see *Discussion*). No statistical difference was observed for SSR power between the three bimodal conditions. Linear mixed effects models with modulation frequency and hemisphere as factors found no significant statistical interactions. Wilcoxon sign-rank tests were performed on the incidental power centered around 10 Hz to determine if it was significant. Results of the tests across conditions yielded no significant results.

Table 1 about here

Dipole source estimation

Based on the literature, we assumed several cortical locations to be implicated for each condition of the experiment. The response for the unimodal auditory condition should be localized to early auditory cortex (Ross B *et al.*, 2000) the response for the unimodal visual condition to occipital visual cortex (Müller MM *et al.*, 1997) For the bimodal conditions, we assumed an SSR source localization to superior temporal sulcus or

438 superior parietal lobule, both regions previously suggested to underlie AV integration
439 (e.g. (Molholm S *et al.*, 2006)). However, equivalent current dipole estimation was
440 largely unsuccessful. The majority of localization estimates had a goodness of fit values
441 < 80% and for estimates that met our criteria, localization was more successful for the
442 LH than for the RH. When dipole localization was successful at all, dipoles were
443 localized to the LH, corresponding perhaps to the parietal lobe area (Molholm S *et al.*,
444 2006). Because this study is explicitly about the detection and modulation of a specific
445 electrophysiological response, the SSR, we did not design the study with source
446 localization in mind (and indeed, we do not have subject structural MRIs to permit
447 adequate source localization). Although the unimodal SSR responses are well
448 characterized in the literature and their superior temporal and occipital localizations not
449 controversial, it would be helpful to be able to localize the bimodal SSR here, but we
450 were not able to with sufficient accuracy.
451
452 *PCA (onset) data for envelopes and response amplitudes*
453 Statistical evaluation of the first principal component yielded no significant interactions.
454 However, several potential interesting possibilities warranting further studies were
455 observed. First, the rise time of the onset response may vary with the relative phase of
456 the components. When both signal component envelopes are completely synchronized,
457 the rise time appeared to be fastest; when completely desynchronized the rise time
458 appeared to be slowest. There are potential differences between hemispheres and
459 sensor/cortical divisions. Additionally, the initial part of the onset response seems to
460 reflect the nature of the signals presented: there is initial sinusoidal activity that reflects

the both the AM envelope of the auditory signal component as well as the increases and decreases in the radius of the visual signal component. Occurring prior to signal entrainment and the SSR, this onset activity has roughly the same duration as the period of the modulation frequency.

As with the PCA, no statistically significant interactions were observed for the mean magnetic field deflections. However, as for PCA, there were several interactions that warrant follow-up. Though no latency or peak field values were found to be significant, the spatial configuration of the observed magnetic fields may vary based the type of modulation (unimodal vs. bimodal) as well as the phase of the signal components. Contributions from the posterior temporal lobe/sensors result in the magnetic field displaying a more 'auditory' or heteromodal character. As with the previously described SSR topographies, the observed magnetic field was a mixture of auditory and visual magnetic field topographies.

Discussion

The study examines the steady state response properties to multisensory signals. We demonstrate that the oscillatory activity in the human auditory and visual cortices entrains to periodic stimuli as reflected in increased power in the SSR at the modulation frequencies. Given that auditory SSRs and visual SSRs have been utilized as robust diagnostic tools, this result is not surprising. However, the types SSR studies differ from the current experiment in three important ways. First, our visual stimulus differed from

484 the usual transient (checkerboard or flicker) stimuli used in SSVEPs. We used gradually
485 growing and shrinking discs in an effort to approximate the opening and closing of a
486 human mouth producing speech and were successful in eliciting steady state responses
487 at the modulation frequencies employed. The modulation frequencies chosen here
488 were based on previous findings regarding the temporal modulation of the speech
489 envelope (Chandrasekaran C et al., 2009)
490
491 A second important difference is the use of a *combination* of auditory and visual signals
492 to elicit the SSR and to evaluate whether entrainment to comodulated, multisensory
493 stimuli differs from unimodally modulated stimuli. We find that the multisensory,
494 comodulated signals elicit a SSR with greater power than either unimodal signal. This
495 experiment was designed to model the potential tracking of speech amplitude
496 envelopes at an approximately syllabic rate (~ 4 Hz). The neurophysiological
497 mechanisms and behavioral consequences of crossmodal integration may be related to
498 that of the auditory phenomenon known as comodulation masking release (CMR)
499 (Grant KW and PF Seitz, 2000). This additional power at the modulation frequency may
500 reflect the neural advantage behind the perceptual boost gained from the release from
501 masking in bimodal speech.
502
503 In CMR, an auditory signal can be detected even at poor signal to noise ratios due to
504 the presence of a comodulated stimulus. The degree of release from masking is
505 greatest when the masker bandwidth is large and has a high spectrum level, the
506 modulation frequency of the signal is low, has a high modulation depth and regular

envelope (Verhey JL et al., 2003). This has obvious parallels with audio-visual speech signals where background noise (approximately white noise) can mask the auditory signal of interest, the speech envelope is modulated at low frequencies and has a relatively high modulation depth. The mechanisms involved for such an auditory-only process could logically be extended not only to other modalities, but also to crossmodal computations.

Third, this study investigates levels of asynchrony to evaluate the *tolerance* of this SSR to shifts in phase between the modalities. We did not, however, show statistically significant differences in the power of the responses at the modulation frequency as a function of the phase shift between the signals in the bimodal conditions.

We hypothesize that the initial differences in phase give rise to distinct representations of the bimodal signals. For the condition in which both signal component envelopes are completely synchronous, the AV signal is computed (integrated) as being a single object. When the starting phases are orthogonal to each other, the AV signal is alternately evaluated as being either two signals or one signal, as the initial phase difference causes the component envelopes to synchronize and desynchronize over the duration of the signal. And when the signal component envelopes are π radians out of phase, each component is evaluated as being a separate object. When the signal component envelopes are completely synchronized, this is analogous to a highly correlated statistical regularity in the bimodal signal. As the signal components are desynchronized, these correlations and redundancies in the bimodal signal decrease,

modifying the processing and representation of the percept. Chandrasekaran et al. (2009) employed 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.

While we have demonstrated that the feasibility of eliciting a bimodal SSR and that the SSR indexes congruency, the indexing performed by the SSR, at least with the conditions in this particular experiment, are limited. For all bimodal conditions we observed that their response power was greater than that of the unimodal conditions and that the further the signal components were separated in phase, the response power decreased. However the congruency indexed by the phase separation may have practical limits. There is evidence that integration of bimodal signals, with the auditory signal component leading, takes place within a 40-60 ms duration window (van Wassenhove V et al., 2007). For the modulation frequencies employed in this experiment, the incongruity between signal components did not fall within this integration window. It is entirely possible that the incongruity tolerance is dependent on the modulation frequency. Higher envelope modulation rates (e.g. 7-11 Hz) will yield phase separations that can test the tolerance between signal components. 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 D *et al.*, 2008; van Wassenhove V *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. Future studies could investigate the converse situation, where the modulation envelope of the visual signal component reaches the maximum prior to that of the auditory component. Further iterations of this experimental paradigm could investigate a combination of these factors: the impact modulation frequency and phase has on bimodal integration, using a wider range of phase separations and determining to what extent which signal component reaches the maximum envelope value first affects indexing of congruity.

Subsequent experiments can improve on the paradigm introduced here in several important ways. First, the modulation depth of the auditory signal component might be made more variable. In this study, we aimed to have as much congruency as possible between the auditory and visual components of the signal. To create a more ecologically valid AV signal, the modulation depth should correspond to the conditions occurring in natural human speech, where the mouth opens and closes fully (modulation depth ranging from 0 to 100 per cent). Secondly, the analysis of the signals could be improved in several ways. The amount of trials analyzed were sufficient to characterize the SSR, but not the signal onsets. It may be possible to analyze both the SSR and the signal onsets using PCA by increasing the number of trials while decreasing the trial

duration. The SSR may still be recovered by concatenating the trials (Xiang 2008) while to recover the onset information, the trials can be averaged, as with the analysis of M100 or M170 responses. Several potential hypotheses suggest that conducting a dual analysis in this manner would be potentially useful. First, it has been previously argued that the RH shows a slight preference in extracting envelope information in auditory processing (Luo H and D Poeppel, 2007) even though we found no difference in power between hemispheres. The LH might play a larger role in analyzing the onset responses and the increase in trial numbers might provide a more clear set of data to analyze.

An additional way the signal could be improved is by using ellipsoids rather than circles to simulate mouth movements. This shape is not only much closer to that of a human mouth, but by modulating only the shorter of the radii, this would yield a more natural modulation motion. There was also a technical issue in implementing the current study. We desired to use modulation rates in the 4 – 16 Hz range; however these modulation rates caused shape and color change effects in the visual signal component. Pilot versions of the current study found that these rates and the perceptual effects they caused made the experiment undoable, except at slower modulation rates. Ultimately, we chose modulation rates that agreed with previous research and that did not induce undesired perceptual effects. The human visual system can track changes faster than those of 16 Hz without such issues occurring and thus there may be a local vs. global problem when using circular stimuli as employed in this study. Using the elliptical shape may reduce these effects and allow for the creation and investigation of more

599 ecologically valid signals. Adding 'jitter' or noise to the signal component envelopes
600 may also yield a yet more ecologically valid set of stimuli for further experimentation.
601 This adds the variability inherent in ecological speech, while retaining the modulation
602 information of the signal component envelopes.
603
604 In summary, we demonstrate – to our knowledge for the first time -- that an
605 experimental technique previously solely applied to unimodal signals, the SSR, can be
606 applied to signals of a bimodal nature. Furthermore, the paradigm reported (as well as
607 potential modifications) yields a useful index of signal component congruity that can be
608 applied to the study of speech and other ecologically valid crossmodal interactions.
609
610 **Acknowledgements**
611
612 This project originated with a series of important discussions with Ken W. Grant
613 (Auditory-Visual Speech Recognition Laboratory, Army Audiology and Speech Center,
614 Walter Reed Army Medical Center). The authors would like to thank him for his
615 extensive contributions to the conception of this work. The authors would like to thank
616 Mary F. Howard and Philip J. Monahan for critical reviews of this manuscript. We would
617 also like to thank Jonathan Z. Simon for outstanding analytical assistance of the
618 experimental data and Jeff Walker for technical assistance in data collection. This work
619 was supported by the National Institutes of Health (2R01DC05660 to DP) and the
620 National Institute on Deafness and Other Communication Disorders of the National
621 Institutes of Health (Training Grant DC-00046 support to JJIII and AER).

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.

Baumann O, Greenlee MW (2007) Neural Correates 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. Current Biology 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.

Chandrasekaran C, Trubanova A, Stillittano 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: Cambridge University Press.

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.

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.

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 N, K. (2008) Visual Modulation of Neurons in Auditory Cortex. Cereb Cortex 18: 1560-1574.

666 Kelly SP, Gomez-Ramirez M, Foxe JJ (2008) Spatial Attention Modulates Initial Afferent
667 Activity in Human Primary Visual Cortex. Cereb Cortex 18: 2629-2636.
668 Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of
669 Neuronal Oscillations as a Mechanism of Attentional Selection. Science 320:
670 110.
671 Lalor EC, Kelly SP, Pearlmutter BA, Reilly RB, Foxe JJ (2007) Isolating endogenous
672 visuo-spatial attentional effects using the novel visual-evoked spread spectrum
673 analysis (VESPA) technique. Eur J Neurosci 26: 3536--3542.
674 Luo H, Poeppel D (2007) Phase Patterns of Neuronal Responses Reliably Discriminate
675 Speech in Human Auditory Cortex. Neuron 54: 1001-1010.
676 Luo H, Wang Y, Poeppel D, Simon JZ (2006) Concurrent Encoding of Frequency and
677 Amplitude Modulation in Human Auditory Cortex: MEG Evidence. J Neurophysiol
678 96: 2712-2723.
679 Macaluso E, Driver J (2005) Multisensory spatial interactions: a window onto functional
680 integration in the human brain. Trends Neurosci 28: 264-271.
681 Mesulam MM (1998) From sensation to cognition. Brain 121: 1013-1052.
682 Miller BT, D'Esposito M (2005) Searching for "the Top" in Top-Down Control. Neuron
683 48: 535-538.
684 Molholm S, Martinez A, Shpaner M, Foxe JJ (2007) Object-based attention is
685 multisensory: co-activation of an object's representations in ignored sensory
686 modalities. Eur J Neurosci 26: 499-509.
687 Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory Visual-Auditory Object
688 Recognition in Humans: a High-density Electrical Mapping Study. Cereb Cortex
689 14: 452-465.
690 Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002)
691 Multisensory auditory-visual interactions during early sensory processing in
692 humans: a high-density electrical mapping study. Cognitive Brain Research 14:
693 115-128.
694 Molholm S, Sehatpour P, Mehta AD, Shpaner M, Gomez-Ramirez M, Ortigue S, Dyke
695 JP, Schwartz TH, Foxe JJ (2006) Audio-Visual Multisensory Integration in
696 Superiour Parietal Lobule Revealed by Human Intracranial Recordings. J
697 Neurophysiol 96: 721-729.
698 Müller MM, Teder W, Hillyard SA (1997) Magnetoencephalographic recording of
699 steadystate visual evoked cortical activity. Brain Topography 9: 163-168.
700 Murray MM, Foxe JJ, Wylie GR (2005) The brain uses single-trial multisensory
701 memories to discriminate without awareness. NeuroImage 27: 473-478.
702 Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh
703 inventory. Neuropsychologia 9: 97-113.
704 Olson IR, Gatenby JC, Gore JC (2002) A comparison of bound and unbound audio-
705 visual information processing in the human cerebral cortex. Cognitive Brain
706 Research 14: 129-138.
707 Ross B, Borgmann C, Draganova R, Roberts LE, Pantev C (2000) A high-precision
708 magnetoencephalographic study of human auditory steady-state responses to
709 amplitude modulated tones. J Acoust Soc Am 108: 679-691.
710 Schroeder CE, Foxe J (2005) Multisensory contributions to low-level, 'unisensory'
711 processing. Curr Opin Neurobiol 15: 454-458.

712 Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of
713 sensory selection. Trends Neurosci 32: 9-18.
714 Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillaions
715 and visual amplification of speech. Trends Cogn Sci 12: 106-113.
716 Senkowski D, Molholm S, Gomez-Ramirez M, Foxe JJ (2006) Oscillatory Beta Activity
717 Predicts Response Speed during a Multisensory Audiovisual Reaction Time
718 Task: A High-Density Electrical Mapping Study. Cereb Cortex 16: 1556-1565.
719 Senkowski D, Schneider TR, Foxe JJ, Engel AK (2008) Crossmodal binding through
720 neural coherence: implications for multisensory processing. Trends Neurosci 31:
721 401-409.
722 Sohmer H, Pratt H, Kinarti R (1977) Sources of frequency following response (FFR) in
723 man. Electroencephalogr Clin Neurophsyiol 42: 656-664.
724 Steeneken HJM, Houtgast T (1980) A physical method for measuring speech-
725 transmission quality. J Acoust Soc Am 67: 318-326.
726 Stein BE, Meredith MA, Huneycutt WS, McDade L (1989) Behavioral Indices of
727 Multisensory Integration: Orientation to Visual Cues is Affected by Auditory
728 Stimuli. J Cogn Neurosci 1: 12-24.
729 Sumby WH, Pollack I (1954) Visual Contribution to Speech Intelligibility in Noise. J
730 Acoust Soc Am 26: 212-215.
731 van Wassenhove V, Grant KW, Poeppel D (2007) Temporal window of integration in
732 auditory-visual speech perception. Neuropsychologia 45: 598-607.
733 Verhey JL, Pressnitzer D, Winter IM (2003) The psychophysics and physiology of
734 comodulation masking release. Exp Brain Res 153: 405-417.
735
736

737

Figure captions

Figure 1a. Schematic of stimuli employed in experiment. Upper panel illustrates the movement of the visual signal component throughout the duration of stimulus (4 seconds – see *Methods* for details). Lower panel illustrates the auditory signal component for the duration of the stimulus. The stimuli were presented at one of two modulation frequencies (Fm = 2.5 and 3.7 Hz), modulation depth was 24 per cent, carrier frequency for the auditory signal component was 800 Hz. Synchronous condition is pictured: maximum radius of circle corresponds to maximum envelope value for auditory component.

Figure 1b. Visual signal components were centered on a 640 x 480 black background and presented from 2.5° - 4° visual angle approximately 30 cm from the subjects' nasion. As in Fig 1a, maximum radius of the circle corresponds to maximum envelope value for auditory component when both component envelopes are synchronized.

Figure 1c. 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 *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.

Figure 2a. Butterfly plot of MEG waveform pre-windowing from a single subject (see *Methods* for details). Recorded magnetic field deflections are in black, root-mean-square (RMS) in red. This illustrates the recorded field deflections (onset and SSR) prior to multiplication by the Kaiser window in the post-signal onset time domain.

Figure 2b. Magnification of data shown in panel 2a, focusing on the recorded onset response (prior to multiplication by Kaiser window in the time domain). This illustrates the necessity of minimizing the onset responses to avoid contamination of the evaluation of the steady-state portion of the signal.

Figure 2c. Sensor configuration of whole-head biomagnetometer. The waveforms displayed are from the steady-state portion of the response from 2810 to 3010 ms post-stimulus onset.

Figure 3. Single subject response power plots. Shown is the response at Fm = 2.5 Hz, for each of the unimodal modulation conditions Left column plots the recorded response for the LH, right column the response for the RH. Top row illustrates the response from the anterior temporal sensors, middle row posterior temporal sensors (these rows correspond to unimodal A condition), bottom row occipital sensors (corresponds to unimodal V condition). Gray shading highlights the response power at modulation frequency and second and third harmonics. Responses were found to be significant for the first, second and third harmonics.

Figure 4a. Complex-valued magnetic field topography for a single subject, Fm = 2.5 Hz, unimodal auditory modulation. The measured response topography resembles that of a visual response, possibly because the occipital response power is greater than that of the temporal response overall (for comparison, see Figure 3).

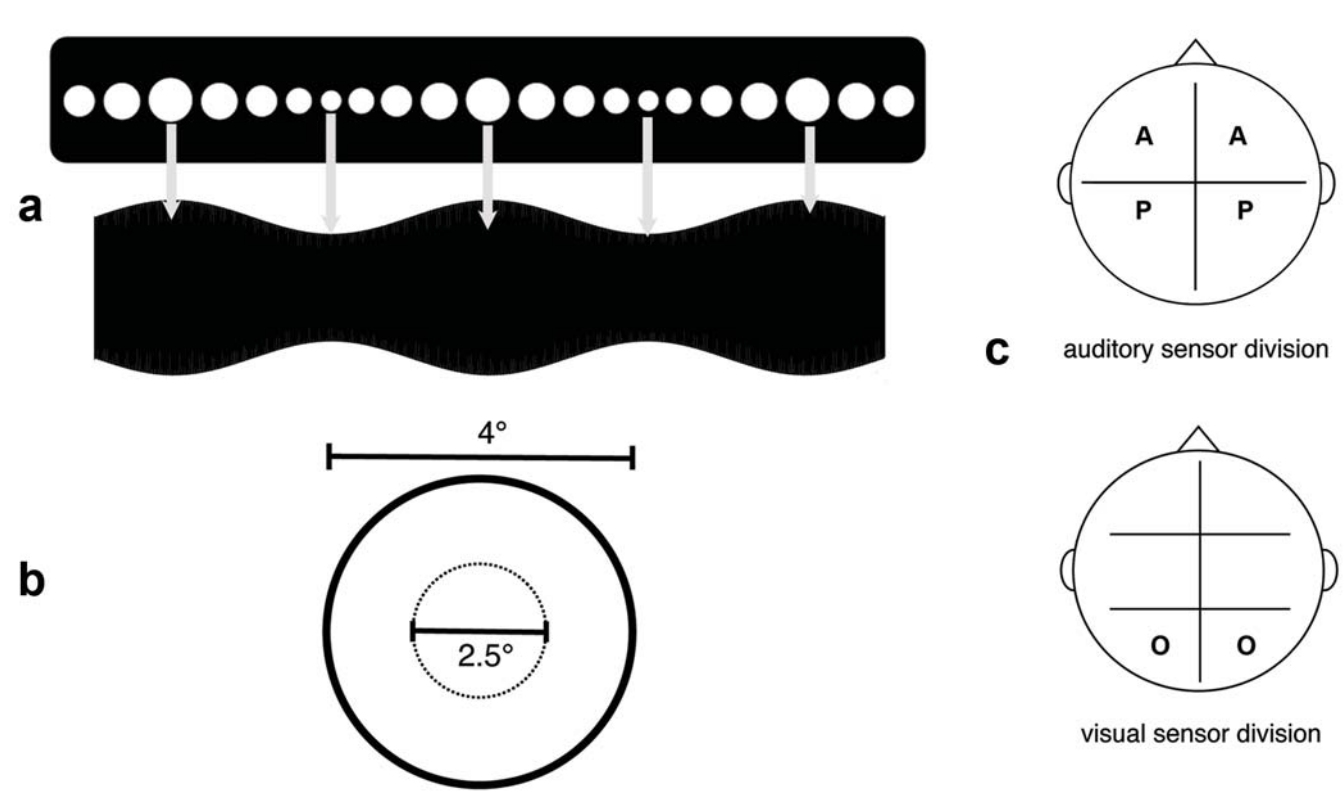
Figure 4b. Complex-valued magnetic field topography, unimodal visual modulation. As in (a), the observed power mirrors that of a primarily visual response.

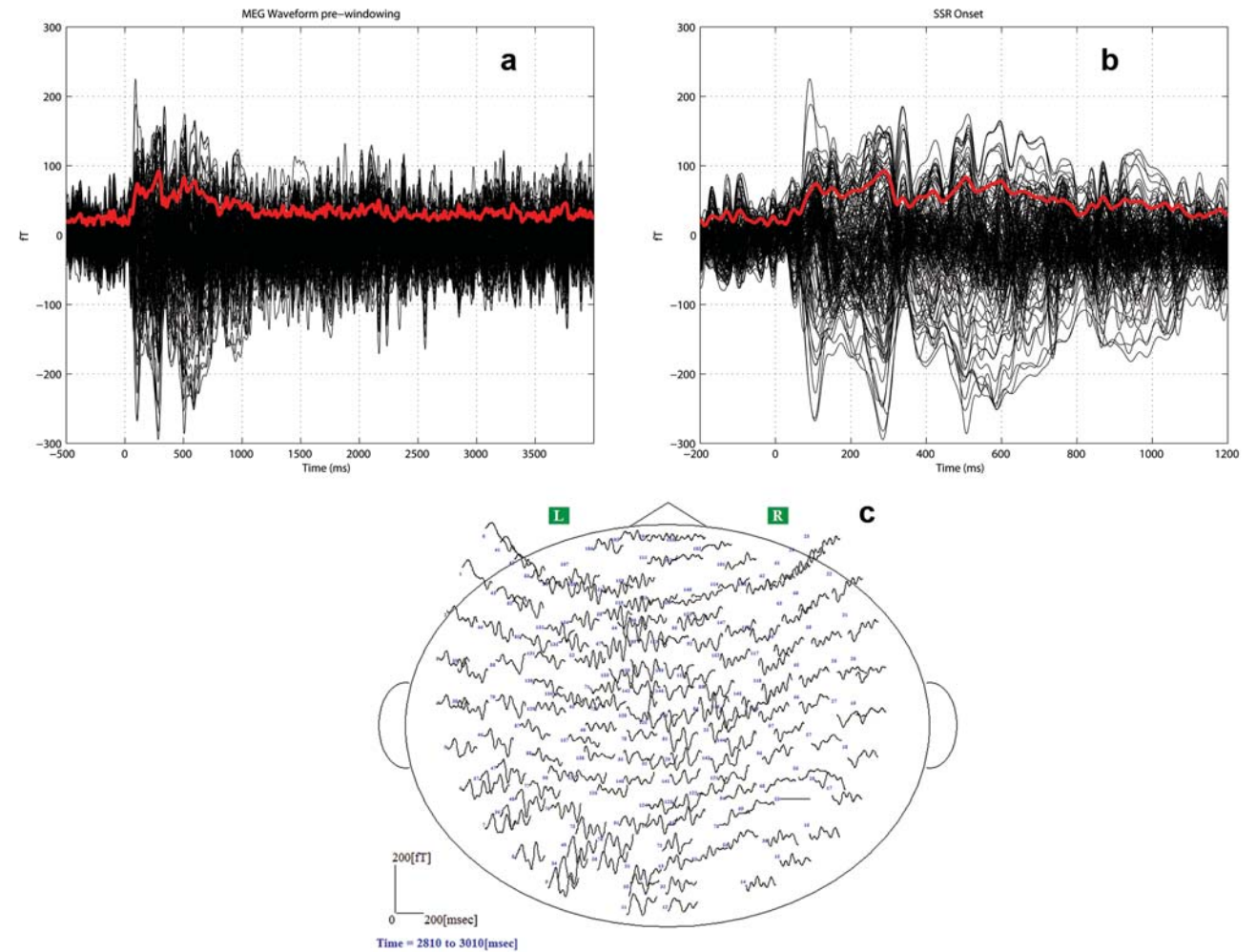
Figure 4c – Figure 4e. Complex valued magnetic field topography for the bimodal conditions. The sink/source pattern shows increasing auditory cortex contribution to bimodal processing observed in the magnetic field spatial distribution. Panel (c) plots the response for $\Phi = 0$, panel (d), $\Phi = \pi/2$, panel (e) $\Phi = \pi$.

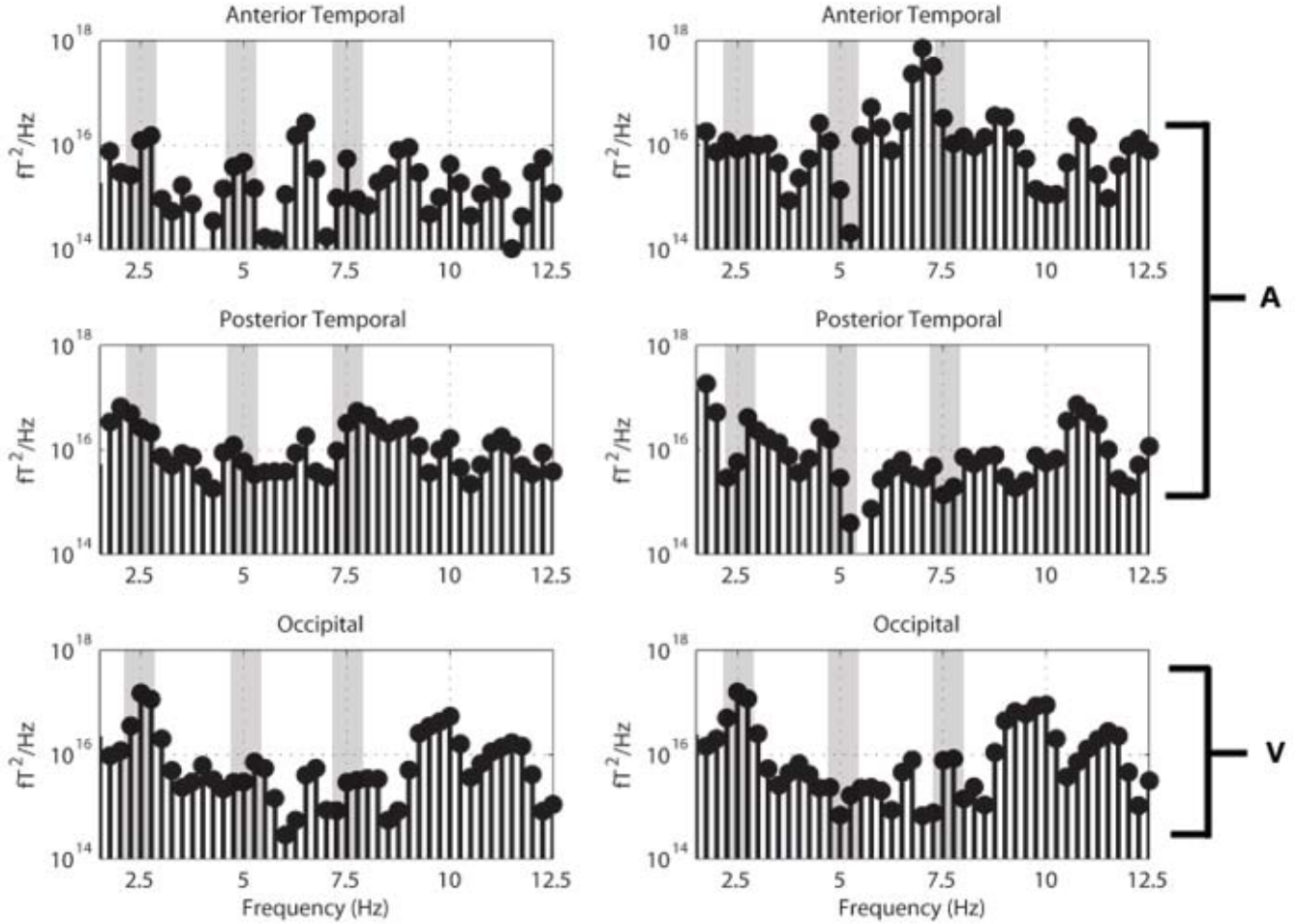
Figure 5. Across subject response power for all three bimodal conditions, Fm = 3.7 Hz. Mean response power across subjects for RH sensors only. Frequency (Hz) is plotted on a linear scale; response power (fT^2/Hz) is plotted on a logarithmic scale. Gray shading indicates power at the modulation frequency and second harmonic. SSR power peaked at the modulation frequency and the second harmonic, with some incident activity centered around 10 Hz. Response power is concentrated in the sensors overlying the posterior temporal and occipital lobes (middle and bottom rows). Logarithmic plotting on the y-axis is employed to give a sense of the range of the

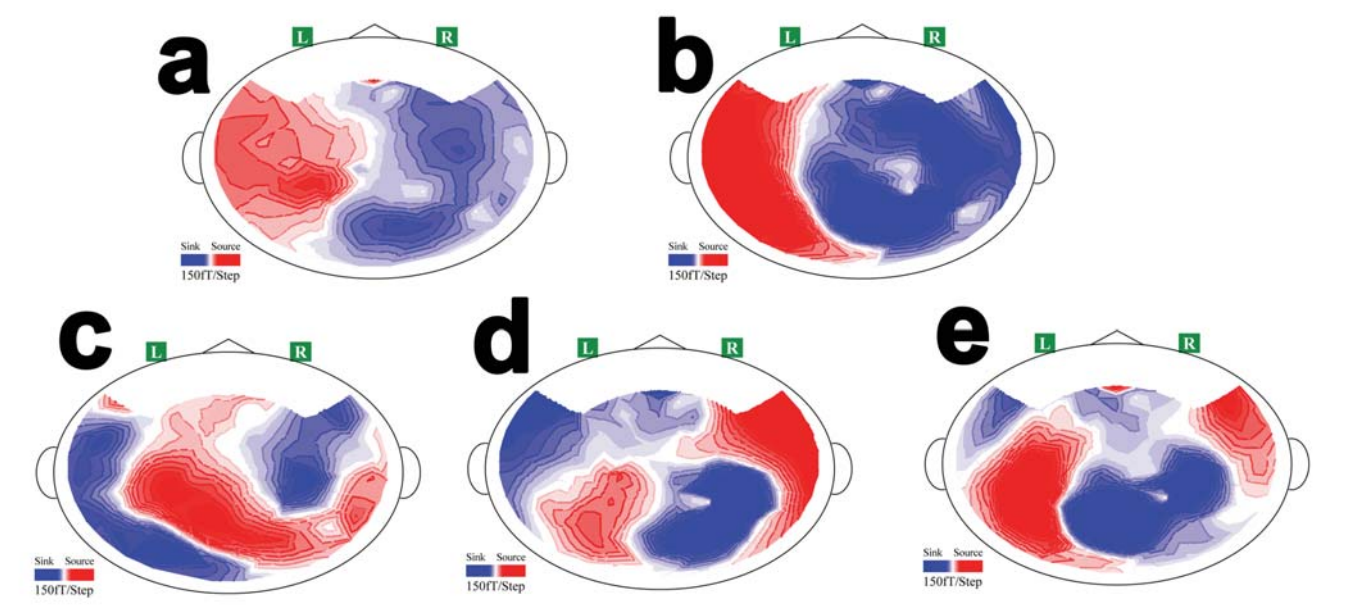
response power for all sensor areas; however, it skews the data somewhat in that the powers found to be statistically significant by the *F* test do not appear to be significant.

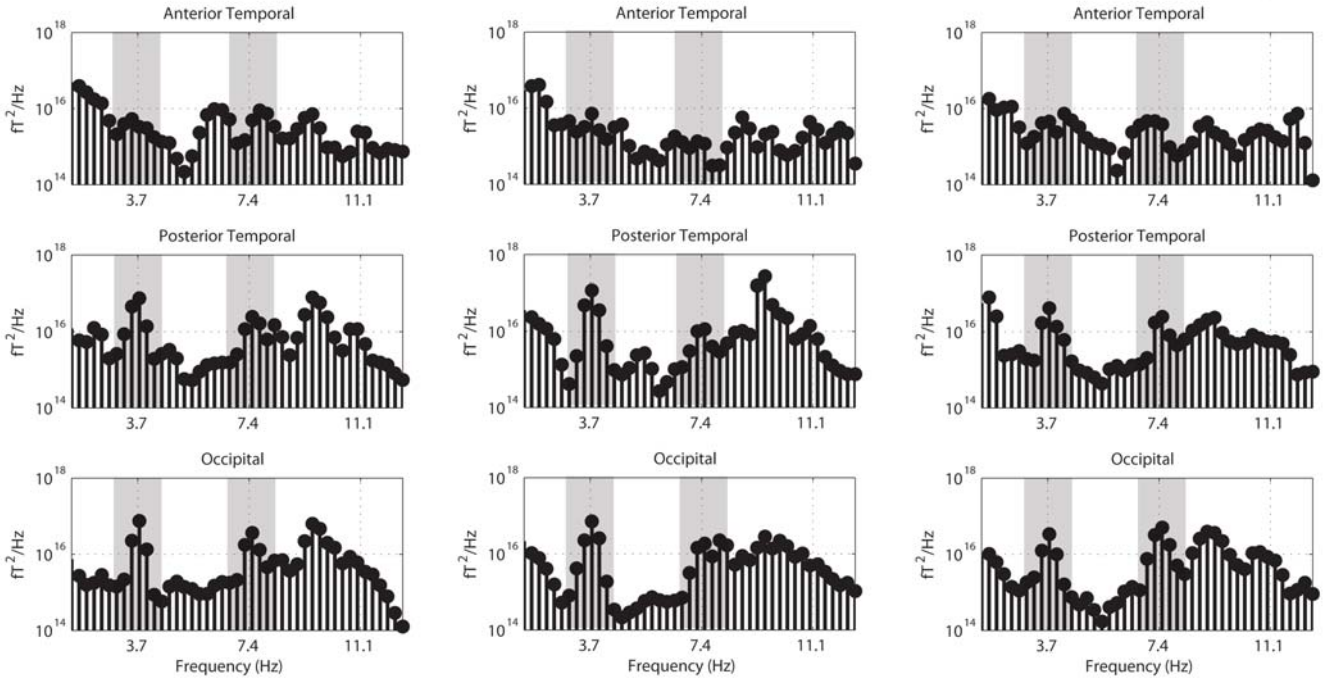
Figure 6. Magnitude response plots at modulation frequency for each modulation frequency (top row shows Fm = 2.5 Hz, bottom row shows Fm = 3.7 Hz) and each hemisphere (left panels show LH, right panels show RH), grouped by experimental condition. Magnitude is displayed logarithmically on the y-axis (fT^2/Hz), and gray shading indicates sensor division. Response power for comodulated conditions is greater than response power for unimodal conditions. Greatest response power is seen in the posterior temporal sensors when the envelopes are initially orthogonal ($\Phi = \pi/2$).

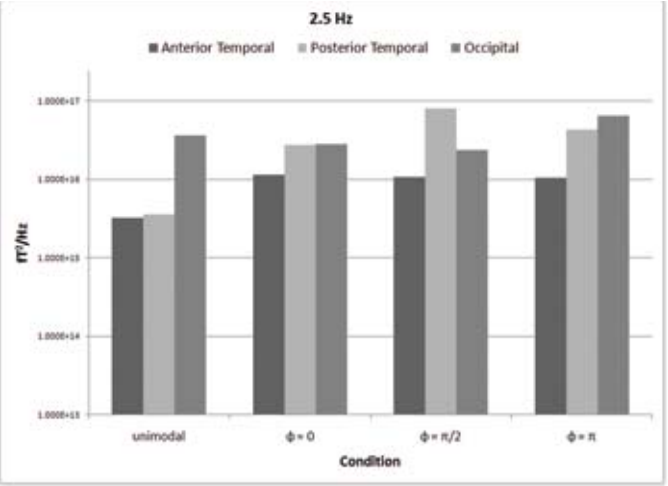

a
b
4°
2.5°
c
A
A
P
P
auditory sensor division
O
O
visual sensor division

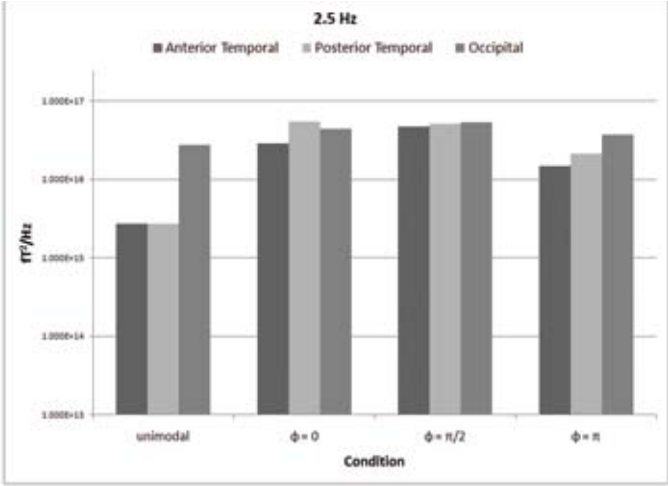

MEG Waveform pre−windowing
a
fT
Time (ms)
SSR Onset
b
fT
Time (ms)
L
R
c
200[fT]
0
200[msec]
Time = 2810 to 3010[msec]

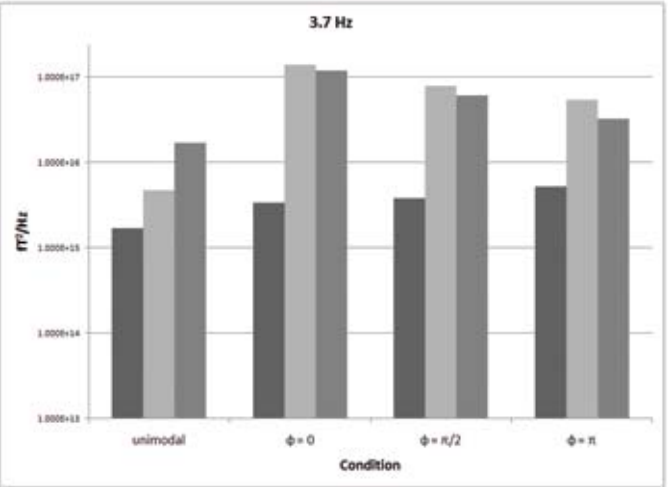

Anterior Temporal
Posterior Temporal
Occipital
Anterior Temporal
Posterior Temporal
Occipital
fT²/Hz
10¹⁸
10¹⁶
10¹⁴
2.5
5
7.5
10
12.5
Frequency (Hz)
A
V

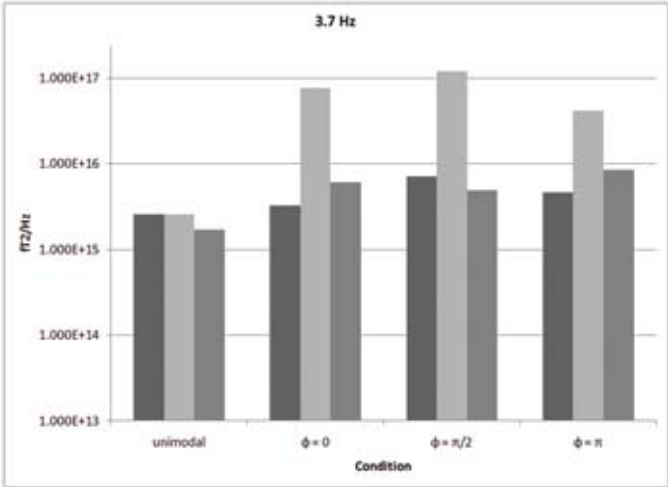

a
L
R
Sink
Source
150fT/Step
b
L
R
Sink
Source
150fT/Step
c
L
R
Sink
Source
150fT/Step
d
L
R
Sink
Source
150fT/Step
e
L
R
Sink
Source
150fT/Step


Fm = 3.7 Hz, phi = 0
Fm = 3.7 Hz, phi = pi/2
Fm = 3.7 Hz, phi = pi
Anterior Temporal
Posterior Temporal
Occipital
fT^2/Hz
10^18
10^16
10^14
3.7
7.4
11.1
Frequency (Hz)


2.5 Hz
Anterior Temporal
Posterior Temporal
Occipital
fT²/Hz
1.000E+17
1.000E+16
1.000E+15
1.000E+14
1.000E+13
unimodal
ϕ = 0
ϕ = π/2
ϕ = π
Condition


2.5 Hz
Anterior Temporal
Posterior Temporal
Occipital
fT²/Hz
1.000E+17
1.000E+16
1.000E+15
1.000E+14
1.000E+13
unimodal
ϕ = 0
ϕ = π/2
ϕ = π
Condition


3.7 Hz
fT²/Hz
1.000E+17
1.000E+16
1.000E+15
1.000E+14
1.000E+13
unimodal
ϕ = 0
ϕ = π/2
ϕ = π
Condition


3.7 Hz
fT2/Hz
1.000E+17
1.000E+16
1.000E+15
1.000E+14
1.000E+13
unimodal
ϕ = 0
ϕ = π/2
ϕ = π
Condition

a.

Fm = 2.5 Hz								
Cond	**Unimodal**		**Φ = 0**		**Φ = π/2**		**Φ = π**	
Hem	**LH**	**RH**	**LH**	**RH**	**LH**	**RH**	**LH**	**RH**
Ant Tem	Fm=3.253e15 2nd=9.985e14 3rd=1.783e15	Fm=2.750e15 2nd=1.180e15 3rd=4.888e14	Fm=1.136e16 2nd=1.294e15 3rd=8.316e14	Fm=2.866e16 2nd=2.797e15 3rd=1.092e15	Fm=1.077e16 2nd=6.617e14 3rd=8.486e14	***Fm=4.741e16*** 2nd=1.305e15 3rd=8.950e14	Fm=1.043e16 2nd=8.746e14 3rd=1.409e15	***Fm=1.494e16*** 2nd=1.827e15 3rd=3.193e15
Pos Tem	Fm=3.630e15 2nd=1.539e15 3rd=1.199e16	Fm=2.750e15 2nd=1.577e15 3rd=5.518e16	***Fm=2.739e16*** ***2nd=5.645e15*** ***3rd=7.867e16***	***Fm=5.489e16*** ***2nd=1.298e16*** 3rd=2.005e16	***Fm=8.04e16*** 2nd=1.676e15 3rd=6.539e16	***Fm=5.112e16*** 2nd=1.581e15 3rd=2.867e16	***Fm=4.281e16*** ***2nd=4.668e15*** 3rd=3.021e16	***Fm=2.149e16*** ***2nd=4.7875e15*** 3rd=4.981e16
Occ	Fm=3.654e16 2nd=4.213e15 3rd=1.910e16	Fm=2.760e16 2nd=4.772e15 3rd=2.160e16	Fm=2.821e16 2nd=8.426e15 ***3rd=1.068e17***	Fm=4.403e16 2nd=6.545e15 3rd=3.176e16	Fm=2.376e16 2nd=3.606e15 3rd=1.291e17	Fm=5.317e16 2nd=9.348e15 3rd=4.404e16	Fm=6.533e16 2nd=9.061e15 3rd=2.918e16	Fm=3.680e16 2nd=1.229e16 3rd=1.046e16

b.[1]

[1] (a) Observed SSR power, Fm=2.5 Hz. Shown are the mean SSR power values for each experimental condition, hemisphere and sensor division for the frequencies found to be significant via an *F* test (see *Methods* for details). Italicized and bolded entries indicate the comparisons found to be significant via Wilcoxon sign-rank tests between unimodal and bimodal modulation conditions. While bimodal SSR power was greater than that of the unimodal condition, not all interactions were found to be significant. The majority of significant comparisons were localized to the posterior temporal sensors for the frequency values at the modulation frequency and the theta band harmonic. (b) Identical data as in (a), Fm=3.7 Hz. Conventions are the same as described previsouly.

Fm = 3.7 Hz

Cond	Unimodal		Φ = 0		Φ = π/2		Φ = π	
Hem	**LH**	**RH**	**LH**	**RH**	**LH**	**RH**	**LH**	**RH**
Ant Tem	Fm=1.687e15 2nd=1.094e15 3rd=1.787e15	Fm=2.575e15 2nd=1.476e15 3rd=9.166e15	Fm=3.372e15 2nd=8.640e15 3rd=9.878e14	Fm=3.260e15 2nd=4.971e15 3rd=2.482e15	Fm=3.823e15 2nd=4.821e15 3rd=4.257e15	Fm=7.108e15 2nd=1.160e15 3rd=4.3526e15	***Fm=5.176e15*** 2nd=2.792e15 3rd=1.620e15	***Fm=4.581e15*** 2nd=3.791e15 3rd=2.560e15
Pos Tem	Fm=4.701e15 2nd=4.938e15 3rd=1.170e16	Fm=2.575e15 2nd=5.001e15 3rd=6.959e15	***Fm=1.384e17*** ***2nd=2.196e17*** 3rd=5.496e15	***Fm=7.631e16*** 2nd=2.451e16 3rd=1.156e16	***Fm=7.857e16*** 2nd=3.819e16 3rd=1.797e16	***Fm=1.186e17*** 2nd=1.134e16 3rd=9.994e15	***Fm=5.400e16*** ***2nd=4.334e16*** 3rd=1.585e16	Fm=4.149e16 2nd=2.452e16 3rd=5.488e15
Occ	Fm=1.686e16 2nd=2.994e16 3rd=1.422e15	Fm=3.645e16 2nd=3.942e16 3rd=1.713e15	***Fm=1.194e17*** 2nd=1.344e17 ***3rd=1.205e16***	Fm=7.506e16 2nd=3.581e16 3rd=6.079e15	Fm=6.044e16 2nd=3.952e16 ***3rd=1.571e16***	Fm=7.048e16 2nd=1.835e16 3rd=4.895e15	Fm=3.236e16 2nd=1.150e17 ***3rd=1.693e16***	Fm=3.318e16 2nd=5.044e16 3rd=8.471e15