



Modality-dependent "what" and "where" preparatory processes in auditory and visual systems

Diaconescu, Andreea Oliviana ; Alain, Claude ; McIntosh, Anthony Randal

Abstract: The present study examined the modality specificity and spatio-temporal dynamics of "what" and "where" preparatory processes in anticipation of auditory and visual targets using ERPs and a cue-target paradigm. Participants were presented with an auditory (Experiment 1) or a visual (Experiment 2) cue that signaled them to attend to the identity or location of an upcoming auditory or visual target. In both experiments, participants responded faster to the location compared to the identity conditions. Multivariate spatio-temporal partial least square (ST-PLS) analysis of the scalp-recorded data revealed supramodal "where" preparatory processes between 300-600 msec and 600-1200 msec at central and posterior parietal electrode sites in anticipation of both auditory and visual targets. Furthermore, preparation for pitch processing was captured at modality-specific temporal regions between 300 and 700 msec, and preparation for shape processing was detected at occipital electrode sites between 700 and 1150 msec. The spatio-temporal patterns noted above were replicated when a visual cue signaled the upcoming response (Experiment 2). Pitch or shape preparation exhibited modality-dependent spatio-temporal patterns, whereas preparation for target localization was associated with larger amplitude deflections at multimodal, centro-parietal sites preceding both auditory and visual targets. Using a novel paradigm, the study supports the notion of a division of labor in the auditory and visual pathways following both auditory and visual cues that signal identity or location response preparation to upcoming auditory or visual targets.

DOI: <https://doi.org/10.1162/jocn.2010.21465>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-55773>

Journal Article

Published Version

Originally published at:

Diaconescu, Andreea Oliviana; Alain, Claude; McIntosh, Anthony Randal (2011). Modality-dependent "what" and "where" preparatory processes in auditory and visual systems. *Journal of Cognitive Neuroscience*, 23(7):1609-1623.

DOI: <https://doi.org/10.1162/jocn.2010.21465>

Modality-dependent “What” and “Where” Preparatory Processes in Auditory and Visual Systems

Andreea Oliviana Diaconescu^{1,2}, Claude Alain^{1,2},
and Anthony Randal McIntosh^{1,2}

Abstract

■ The present study examined the modality specificity and spatio-temporal dynamics of “what” and “where” preparatory processes in anticipation of auditory and visual targets using ERPs and a cue–target paradigm. Participants were presented with an auditory (Experiment 1) or a visual (Experiment 2) cue that signaled them to attend to the identity or location of an upcoming auditory or visual target. In both experiments, participants responded faster to the location compared to the identity conditions. Multivariate spatio-temporal partial least square (ST-PLS) analysis of the scalp-recorded data revealed supramodal “where” preparatory processes between 300–600 msec and 600–1200 msec at central and posterior parietal electrode sites in anticipation of both auditory and visual targets. Furthermore, preparation for pitch process-

ing was captured at modality-specific temporal regions between 300 and 700 msec, and preparation for shape processing was detected at occipital electrode sites between 700 and 1150 msec. The spatio-temporal patterns noted above were replicated when a visual cue signaled the upcoming response (Experiment 2). Pitch or shape preparation exhibited modality-dependent spatio-temporal patterns, whereas preparation for target localization was associated with larger amplitude deflections at multimodal, centroparietal sites preceding both auditory and visual targets. Using a novel paradigm, the study supports the notion of a division of labor in the auditory and visual pathways following both auditory and visual cues that signal identity or location response preparation to upcoming auditory or visual targets. ■

INTRODUCTION

Scene analysis entails identifying (“what”) and localizing (“where”) various objects in the environment. Evidence from electrophysiological recordings in nonhuman primates suggests that identification and localization of visual objects is functionally segregated in ventral and dorsal streams, respectively. More specifically, extrastriate visual cortical areas are broadly organized into two anatomically distinct and functionally specialized pathways: an occipito-temporal pathway for identifying objects and an occipitoparietal pathway for processing spatial relations among objects (Haxby et al., 1991; Livingstone & Hubel, 1988; Mishkin & Ungerleider, 1982) and for integrating vision with action (Goodale & Milner, 1992). Clinical studies in patients with focal brain lesions (Damasio, Damasio, & Van Hoesen, 1982), as well as neuroimaging studies, using PET or fMRI, have revealed the existence of similar segregated streams in humans (Bushara et al., 1999; Andersen, Snyder, Bradley, & Xing, 1997).

In the auditory modality, converging evidence from ERPs and fMRI studies suggests a similar functional segregation for identifying and localizing auditory objects, with temporo-parietal regions responsible for sound location

and anterior temporal–frontal regions responsible for sound identity (Alain et al., 2008; Altmann, Bledowski, Wibral, & Kaiser, 2007; Salmi, Rinne, Degerman, Salonen, & Alho, 2007; Arnott, Grady, Hevenor, Graham, & Alain, 2005; Arnott, Binns, Grady, & Alain, 2004; Alain, Arnott, Hevenor, Graham, & Grady, 2001; Maeder et al., 2001). For instance, Alain et al. (2001) showed that relative to location, pitch identification generated larger fMRI signal in the superior temporal and inferior frontal cortices, and was associated with a sustained potential across inferior fronto-temporal regions. Conversely, sound localization recruited regions in posterior temporal cortex, and in inferior and superior parietal cortices (Alain et al., 2001).

Several studies proposed commonalities between the visual and auditory systems during object processing. In a study of nonhuman primates, Poremba et al. (2003) proposed that anterior temporal regions responsible for pitch processing parallel the unimodal ventral visual pathway. Like the ventral visual pathway, the auditory ventral stream is modality-specific, suggesting that it analyzes sound identity for purposes of pattern recognition much like the visual pathway does for visual stimulus identity (Poremba et al., 2003).

Multisensory integration studies also demonstrated audiovisual interactions during object perception. Perhaps the most documented is the McGurk effect (McGurk & MacDonald, 1976), in which seeing temporally asynchronous

¹Baycrest Centre, Toronto, Ontario, Canada, ²University of Toronto, Ontario, Canada

articulatory gestures paired with speech can modify the percept phonetically. Modality-specific auditory regions, including the superior temporal sulcus (STS), auditory association areas such as the planum temporale (PT), and the superior temporal gyrus (STG), have been shown to respond to audiovisual speech (Allison, Puce, & McCarthy, 2000; Calvert, Campbell, & Brammer, 2000; Puce, Allison, Bentin, Gore, & McCarthy, 1998).

Furthermore, multimodal zones at the borders between unisensory cortical regions, including ventral temporal cortex and lateral occipital cortex, respond to abstract object properties that are accessible via both the auditory and the visual systems (for a review, see Amedi, von Kriegstein, van Atteveldt, Beauchamp, & Naumer, 2005). ERP studies of cross-modal interactions during object processing (Molholm, Ritter, Javitt, & Foxe, 2004; Murray et al., 2004) show evidence of similar time course for processing common object-related pictures and sounds with estimated source locations in the ventral temporal and lateral occipital areas.

Previous ERP studies emphasize the commonalities between the visual and auditory systems with respect to the dorsal, spatial mapping processing stream (Green, Teder-Salejarvi, & McDonald, 2005; Molholm et al., 2002). Intracranial recordings in humans showed that the superior parietal lobule, similarly to the IPL in primates, exhibits audiovisual receptive fields and contributes to the integration of auditory and visual sensory information (Molholm et al., 2006). Neuroimaging studies have also shown that posterior parietal regions are associated with the spatial mapping of auditory (Cohen, Russ, & Gifford, 2005; Weeks et al., 1999) and visual stimuli (Colby & Goldberg, 1999; Heinze et al., 1994). ERP studies of spatial attention reported evidence of bimodal spatio-temporal components during spatial attention tasks (Dale, Simpson, Foxe, Luks, & Worden, 2008; Green et al., 2005; Eimer, van Velzen, & Driver, 2002; Hopf & Mangun, 2000).

Modality-specific responses to object location have also been reported in a study from Bushara et al. (1999) using PET. The group showed that whereas the left inferior frontal and the right middle temporal regions activated specifically in response to sound location, right inferior temporal cortex activated in response to visual stimulus location. The study from Bushara et al. reported unimodal and bimodal networks involved in spatial mapping. However, the study did not use the same type of stimuli in each task, therefore making it difficult to know if the differences were related to the material used or whether they reflected distinct spatio-temporal patterns.

Moreover, most prior research investigating the brain regions engaged in processing object location and object identity in humans used blocked designs in auditory (Alain et al., 2001, 2008; Altmann et al., 2007; Maeder et al., 2001) and visual systems (Shen, Hu, Yacoub, & Ugurbil, 1999; Courtney, Ungerleider, Keil, & Haxby, 1996; Haxby et al., 1991). This makes it difficult to know whether the functional segregation into “what” and “where” pathways can

accommodate a temporally dynamic adjustment of identity versus location task sets. Along similar lines, it would be important to determine the degree of modality specificity of “what” and “where” preparatory processes when these two rules alternate randomly.

The present study investigates the spatio-temporal dynamics of “what” and “where” preparatory processes by employing ERPs and alternating between location and identity cue-driven task sets in anticipation of auditory or visual targets. We hypothesized that preparation to respond to stimulus location and identity is segregated into supramodal and modality-specific processing pathways, respectively, across auditory and visual modalities. In contrast to previous studies that employed different types of auditory and visual stimuli for each identity and location task (cf. Bushara et al., 1999), we sought to examine the spatio-temporal patterns supporting “what” and “where” processes using identical cues and focusing the analyses on the cue–target interstimulus interval (ISI) following the cue and preceding the onset of the target. As such, any task effects observed could be attributed to differences in “what” and “where” processes in anticipation of either auditory or visual targets. This cue–target paradigm also allowed us to investigate audiovisual interactions during “what” and “where” processes. As it has been shown that the auditory system, similarly to the visual system, recruits multimodal posterior parietal regions to map extrapersonal space (cf. Smith et al., 2010; Green et al., 2005), we predicted that preparation for target location will exhibit common spatio-temporal features for auditory and visual systems.

The following experiments were designed to examine both the dynamics and the sensory-modality dependence of location and identity task sets. First, to investigate the rapid adjustment of “what” and “where” preparatory processes, response rules (i.e., location vs. identity) alternated randomly, whereas target modality was blocked. Second, to examine whether “what” and “where” preparatory processes differed depending on the sensory modality of the stimuli used, cue and target modalities were also manipulated. Participants were presented with an auditory (Experiment 1) or a visual (Experiment 2) cue that signaled an identity or location response to a subsequent auditory or visual target.

METHODS

Participants

The participants in this study were healthy young right-handed volunteers between the ages of 18 and 35 years (mean age \pm SD, 26 ± 2) with normal or corrected-to-normal vision and hearing. All participants gave formal informed consent in accordance with the joint Baycrest Centre-University of Toronto Research Ethics Committee. Twenty-two volunteers (12 women) participated in Experiment 1, and another group of 12 volunteers (6 women) participated in Experiment 2.

Experimental Design

Each trial consisted of two stimuli (i.e., cue and target) presented for 250 msec, separated by a 1000-msec ISI, and followed by a response period (see Figure 1). The time interval between the end of the target presentation and the beginning of the next trial was between 800 and 1200 msec (equiprobable). The experiment consisted of two blocks comprising two main randomly presented tasks for a total of 240 trials.

Across both tasks, the cue signaled the response rule to the upcoming lateralized target. In one task, participants were cued to respond to the location of the target (left or right) by pressing a button on the same side as the target presentation. On the keyboard, the key “L” corresponded to a response on the right side, and the key “A” to a response on the left side. We refer to this as the “location” task (LOC).

In the second task, participants were cued to respond to the identity of the target, that is, low versus high pitch for an auditory target, or Shape 1 versus Shape 2 for a visual target. The key “S” corresponded to a response to Target 1, and the key “K” to a response to Target 2. Response buttons for this task were counterbalanced between participants. We refer to this as the “identity” task (IDN). We replicated the experiment above using a visual cue to determine whether the spatio-temporal patterns identified depended on the sensory modality of the cue.

Stimuli

For the cue presentation, two kinds of auditory stimuli were used: 100-Hz amplitude modulated buzz and 200-Hz pure

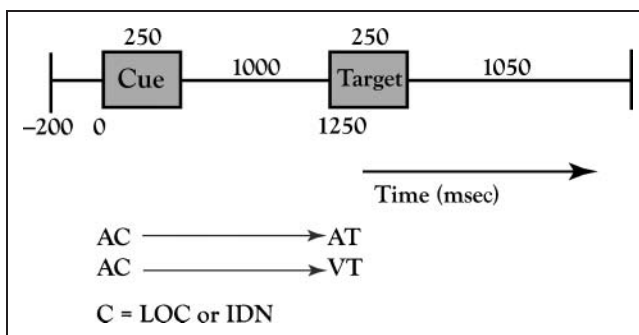


Figure 1. Experiment design. In Experiment 1, two binaural tones (100 Hz buzz and 200 Hz pure tone) served as cues. In the first block, the targets were presented in the auditory modality as two monaural pure tones with distinct pitch (low pitch = 800 Hz; high pitch = 1200 Hz). In the second block, the targets were presented in the visual modality as two lateralized pictures with distinct shapes (Shape 1 = 14.2 × 14.2 cm black-and-white square checkerboard; Shape 2 = same size checkerboard rotated 90°). The time interval between the end of S2 and the beginning of the cue in the next trial was 800–1200 msec, equiprobable ITIs. A 200-msec prestimulus baseline and a 1000-msec ISI interval were used in both experiments. The cue signaled location (LOC) or identity (IDN) response rules to the upcoming targets. In the second experiment, two centrally presented visual stimuli (14.2 cm diameter blue and yellow circles) were used as cues. The experimental design was identical to the first experiment.

tone presented binaurally for 250 msec duration including a 5-msec rise/fall time. The 100-Hz buzz cued an LOC condition, whereas the 200-Hz pure tone an IDN condition. Tone and response rule assignment were counterbalanced between participants. The sensory modality of the target alternated in two blocks. In the first block, the target was presented in the auditory modality. We refer to this as the intramodal (AA) condition. We used low (800 Hz) and high pitched (1200 Hz) tones of 250 msec (including a 5-msec rise and fall time) duration. The two tones were randomized and presented monaurally to the left and right ears. The volume of all tones was initially set to 55 dB (HL), and then adjusted by each participant to be perceptually identical in terms of loudness. Based on the specific cue, participants responded to the location of the target (left vs. right) or to its identity (low vs. high pitch).

In the second block, the target was presented in the visual modality. We refer to this as the cross-modal condition (AV). We used two visual stimuli that matched in terms of luminance and spatial frequency: (1) 14.2 × 14.2 cm black-and-white square checkerboard; (2) same size checkerboard rotated 90° (diamond). Both visual stimuli were presented on the left and right sides of the screen at a viewing distance of 60 cm, and with a 5° visual angle. Based on the cue, participants responded to the location of the target (left vs. right) or to its identity (square vs. diamond). To match the identity and location tasks in terms of difficulty, the distance of the lateralized visual target was set to 6 cm from fixation.

In the second experiment (which used the same experiment design), two kinds of visual stimuli were used as cues: (1) 14.2 cm diameter blue circle, and (2) same size yellow circle with equated luminance levels. Because the same targets followed the cues for the two experiments and because we were interested in the anticipatory sustained potentials prior to target presentation, we used two colored circles as the visual cues in the second experiment. Both visual stimuli were centrally presented at a viewing distance of 60 cm with a 5° visual angle. The blue circle cued an LOC condition, whereas the yellow circle an IDN condition in cross-modal (VA) and intramodal (VV) conditions. Color and response rule assignment were counterbalanced between participants. The targets following the cue were the same as in Experiment 1.

Recording and Preprocessing

The electroencephalogram (EEG) was recorded using NeuroScan 4.32 with a 64-channel ElectroCap according to the standard 10–20 system. Impedances were kept below 5 kΩ. EEG data were digitized at a 250-Hz sampling rate and passed through a 0.01–100 Hz bandpass filter. All electrodes were referenced to Cz during the recording; the data were re-referenced to an average reference off-line, for analysis. Continuous EEG recordings were also notch filtered at 60 Hz to remove line noise.

Data were then epoched into 2.5-sec segments with a 0.2-sec prestimulus baseline. Epoched datasets containing only correct response trials were kept for further analysis. Ocular and muscle artifacts were identified and removed using independent component analysis as implemented in EEGLAB (Delorme & Makeig, 2004). Each participant's artifact-free trials were divided into 16 conditions, according to response rule type, target stimulus location, and target identity for each of the two experiments (2 response rules, 2 target locations, 2 target types). The average range of artifact-free, correct response trials per conditions was between 40 and 50 trials.

Spatio-temporal Partial Least Squares Analysis

The purpose of using spatio-temporal partial least squares (ST-PLS; McIntosh & Lobaugh, 2004; Lobaugh, West, & McIntosh, 2001) was to capture both sensory-specific and multimodal aspects of “what” and “where” processes, and to distinguish “what” and “where” preparatory processes from cortical potentials associated with target processing and motor response preparation or execution. ST-PLS is conceptually related to the analysis ERP difference waveforms (e.g., mismatch negativity), in that it identifies task-related differences in amplitude across all EEG channels by deriving the optimal least-squares contrasts that code for the task differences. As it is a multivariate method, ST-PLS has the advantage of performing this derivation across the entire dataset simultaneously. As such, there is no need to prespecify the time intervals or channels of focus. The statistical assessments are also done in multivariate space, obviating the need for excessive Type I error correction. In terms of multivariate analyses, ST-PLS is most similar to canonical correlation, but ST-PLS focuses on optimizing covariance instead of correlation, and can be applied directly with data that are not full rank.

ST-PLS begins with a data matrix composed of electrode amplitudes with rows corresponding to participants within conditions and columns corresponding to time points within channels. Data are averaged within condition and across subject to obtain task-specific means. The matrix is then mean-centered for each time point and channel and then decomposed using singular value decomposition (SVD). Application of SVD produces a set of mutually orthogonal latent variables (LVs), each consisting of two parts: (i) design salience (or design LV), which identifies the task-dependent contrast between the four experimental conditions: location–intramodal, location–cross-modal, identity–intramodal, and identity–cross-modal; and (ii) the electrode saliences that indicate the location and timing of the task differences identified by the design LV. To link back to the analogy, the electrode saliences are the difference waveform and the design LV is the contrast that produces the difference wave (see Figures 4–7 for such contrasts).

The task- and modality-specific effects were examined prior to target onset, including the cue onset (250 msec)

and the preparation interval (1000 msec). Based on a recent ERP study of cue-driven motor preparation to lateralized visual and auditory targets (Diaconescu, Kovacevic, & McIntosh, 2008), we hypothesized that the spatio-temporal patterns associated with processing the relevant cue will extend approximately 200 msec after cue onset. Because we used a novel paradigm, and because we wanted to capture both sensory-specific and multimodal aspects of “what” and “where” processes, we did not restrict the analysis to a priori selected electrodes, and we included all electrodes and time points into a single analysis.

Statistical Evaluation

Arbitrary decisions regarding the number of LVs to retain and which of the weights to consider important were minimized by providing statistical assessment of the LVs. This was achieved with permutation tests for the LVs and bootstrap estimation of standard errors for the electrode saliences. First, permutation tests were performed by randomly reordering the data matrix rows and calculating a new set of LVs for each reordering to ensure that the task-dependent contrasts identified were significantly different from random data. At each permutation, the LV obtained is compared to the LV from the original data, and it is assigned a probability value based on the number of times the LV from the permuted data exceeds the original value. With permutation tests, we assessed statistical significance of the identified task contrast without relying on distributional assumptions common to conventional parametric statistical methods (McIntosh, Bookstein, Haxby, & Grady, 1996).

Second, statistical reliability of task effects was determined using bootstrap estimation of standard errors for each time point salience. The primary purpose of the bootstrap estimation is to determine those portions of the ERP waveforms that show reliable experimental effects across subjects. Bootstrap estimation of standard errors involves (i) randomly sampling subjects with replacement while keeping the assignment of experimental conditions fixed for all observations, (ii) performing SVD on the resampled matrix, and (iii) computing the standard error of the task contrasts calculated and expressed at each time point (Lobaugh et al., 2001). The time points where the salience was greater than three times the standard error (i.e., a bootstrap ratio ≥ 3) are indicated above or below the plots of the grand-averaged waveforms in Figures 4–9. The bootstrap iterations were also used to derive confidence intervals on the design LV (see Figures 4A–7A).

These two resampling techniques provide complementary information about the statistical strength of the task contrast observed and its reliability across participants. Statistical evaluation of experimental effects was performed using an optimal number of 500 permutations and 300 bootstrap iterations (cf. McIntosh et al., 1996; Efron & Tibshirani, 1986).

Source Waveform Analysis

To obtain additional spatio-temporal precision of the “what” and “where” preparation processes, we performed source modeling using brain electric source analysis (BESA) software (version 5.2.4.48) to confirm the modality specificity of “what” and “where” preparatory processes. Grand-averaged ERPs were calculated for an average of 45 trials for each condition. The source waveform analysis assumed a four-shell ellipsoidal head model from BESA (for other applications of ST-PLS on source waveforms derived from BESA, refer to Alain, McDonald, Kovacevic, & McIntosh, 2009; Lobaugh et al., 2001).

We used a hybrid source montage that was designed to capture auditory and visual evoked potentials, as well as multimodal central, central-parietal, and frontal regional sources. To enhance signal-to-noise ratio, the model was created using grand-averaged ERPs that included the cue onset (250 msec) and the preparation interval (1000 msec), a time interval identical to the analysis performed in electrode space. The model was composed of 12 orthogonal dipoles, including three bilateral sets of orthogonal dipoles to account for all directions of current flow in the auditory system (tangential, radial, anterior/posterior) and three bilateral sets of orthogonal dipoles to account for all directions of current flow in the visual system (radial, basal, inferior). Symmetry and orthogonality constraints were maintained. To ensure that multimodal patterns underlying “what” and “where” processes were also captured, the hybrid model also included 10 regional sources in central (left, right, medial), parietal (left, right, medial), frontal (left, right, medial), and frontal polar (medial) regions. In the grand-average data, source location was kept constant and dipole orientation was allowed to change in order to minimize the signal-to-noise ratio. For each participant, the resulting hybrid model was held fixed and was used as a spatio-temporal filter to derive source waveforms for each condition for both auditory and visual cueing experiments.

In the original version of ST-PLS, SVD is used to identify the experimental effects that capture the largest percentage of cross-block covariance in the data. For the source waveform analysis, we used a nonrotated version of ST-PLS, in which a priori contrasts restrict the task-dependent contrasts to hypothesized experimental effects (McIntosh & Lobaugh, 2004). This version of ST-PLS has the advantage of allowing a direct mapping of the experimental effects from electrode to source space. There is, however, no guarantee that these effects represent the largest percent of cross-block covariance between the experimental design and neuroelectrical brain activity that can be identified using the original version of ST-PLS with SVD (see above). To investigate the experimental effects identified in electrode space, the following task contrasts for location-intramodal, location-cross-modal, identity-intramodal, and identity-cross-modal conditions were employed using nonrotated ST-PLS: (i) $[1 \ -1 \ 1 \ -1]$ for the effect of tar-

get modality, and (ii) $[1 \ 1 \ -1 \ -1]$ for the effect of response rule. Statistical assessment of experimental effects was performed using 500 permutations and 300 bootstrap iterations.

RESULTS

Behavioral Results

Percent accuracy and RT data were analyzed using repeated measures ANOVA, in which Response rule, Target modality, and Target presentation (left vs. right presentation, and Target 1 vs. Target 2) were the within-subject factors. Separate analyses were done for the auditory and visual cue experiments.

The results of the RT analysis for the auditory cue experiment are illustrated in Figure 2. Two main effects of Target modality [$F(1, 87) = 93.5, p < .001$] and Task (i.e., identity vs. location) [$F(1, 87) = 441.6, p < .001$], and an interaction between the two [$F(1, 87) = 141.3, p < .001$], were observed. First, cross-modal conditions (AV) were performed faster than intramodal ones (AA) by an average of 80 msec. Second, conditions in which the cue signaled a location response were faster compared to identity conditions. Finally, an interaction between task and target modality was observed such that location-cross-modal (AV) conditions were significantly faster than intramodal (AA) ones. Identity conditions did not differ between intramodal (AA) and cross-modal (AV) pairings.

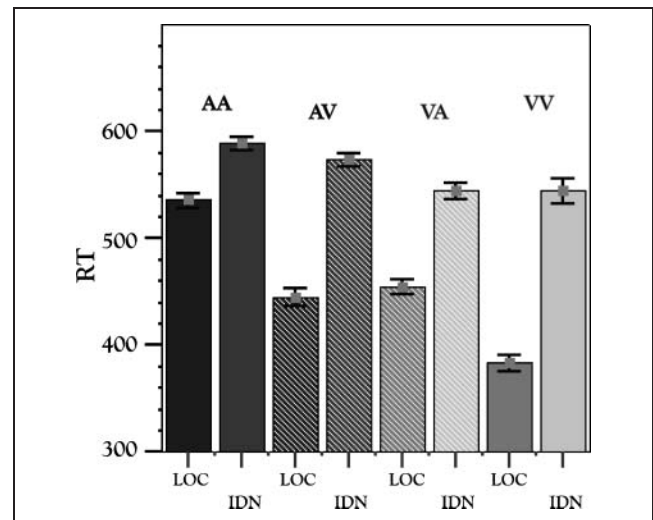


Figure 2. Reaction time data. The two experiments are presented side by side, auditory cue-auditory and visual targets (AA and AV) and visual cue-auditory and visual targets (VA and VV). The first two columns on the left represent mean RTs for the AA conditions alternating between location and identity tasks. The next two columns represent RTs for the AV conditions alternating between location and identity tasks. The next set of columns from the left represents the VA conditions alternating between location and identity tasks, followed by the RTs from the VV conditions in the rightmost column. Error bars represent standard error of the mean.

Upon replicating the experiment using a visual cue, two main effects of target modality [$F(1, 47) = 14.5, p < .001$] and task [$F(1, 47) = 301.3, p < .001$], and an interaction between the two, were observed [$F(1, 47) = 76.1, p < .001$]. In this case, intramodal (VV) conditions were faster than cross-modal (VA) ones by an average of 70 msec. Participants were also faster in the location compared to the identity conditions. An interaction between task and target modality was also observed. Intramodal (VV) location conditions were significantly faster than cross-modal (VA) ones. Identity conditions, however, did not differ greatly between VV and VA pairs. Across both experiments, overall, participants were faster at responding to the location of visual lateralized targets.

Accuracy was analyzed to measure the extent to which cued response rules differed in difficulty. In the auditory cueing experiment, an interaction between Task and Target modality was observed [$F(1, 87) = 21.83, p < .001$]. Participants were at ceiling during crossmodal location conditions (LOC AV). Participants made the most errors during the intramodal identity cued condition (IDN AA), responding with an average of 87% accuracy. No differences in accuracy were observed in the second experiment.

ERP Results

Figure 3 illustrates grand-average ERPs for the entire epoch including both cue- and target-related evoked potentials

in Experiments 1 and 2 for inspection of the basic task-dependent effects. The ERPs in the figure were averaged over task (i.e., location and identity response rules) and target laterality to emphasize cue- and target-related responses. The ERPs comprised N1 and P2 waves elicited by cue onset, followed by a sustained cortical potential that was largest at central and parietal sites and extended between 250 and 800 msec following cue onset. The target-elicited modality-specific evoked responses were followed by a late positive complex reflecting response selection and execution.

To capture the spatio-temporal patterns supporting “what” and “where” processes following identical cues in anticipation of either auditory or visual targets, the epoch that was analyzed included only the cue presentation and the ISI prior to target presentation. For each experiment, a single mean-centered ST-PLS analysis was performed for four conditions: (i) identity–intramodal, (ii) identity–cross-modal, (iii) location–intramodal, and (iv) location–cross-modal.

Experiment 1

In Experiment 1, an effect of Task (LV1 = 49.05% cross-block covariance, $p < .012$; Figure 4A) and Target modality (LV2 = 40.50% cross-block covariance, $p < .04$; Figure 5A) were observed (Table 1A).

Preparation for processing the location of an auditory target was expressed maximally at left central–parietal

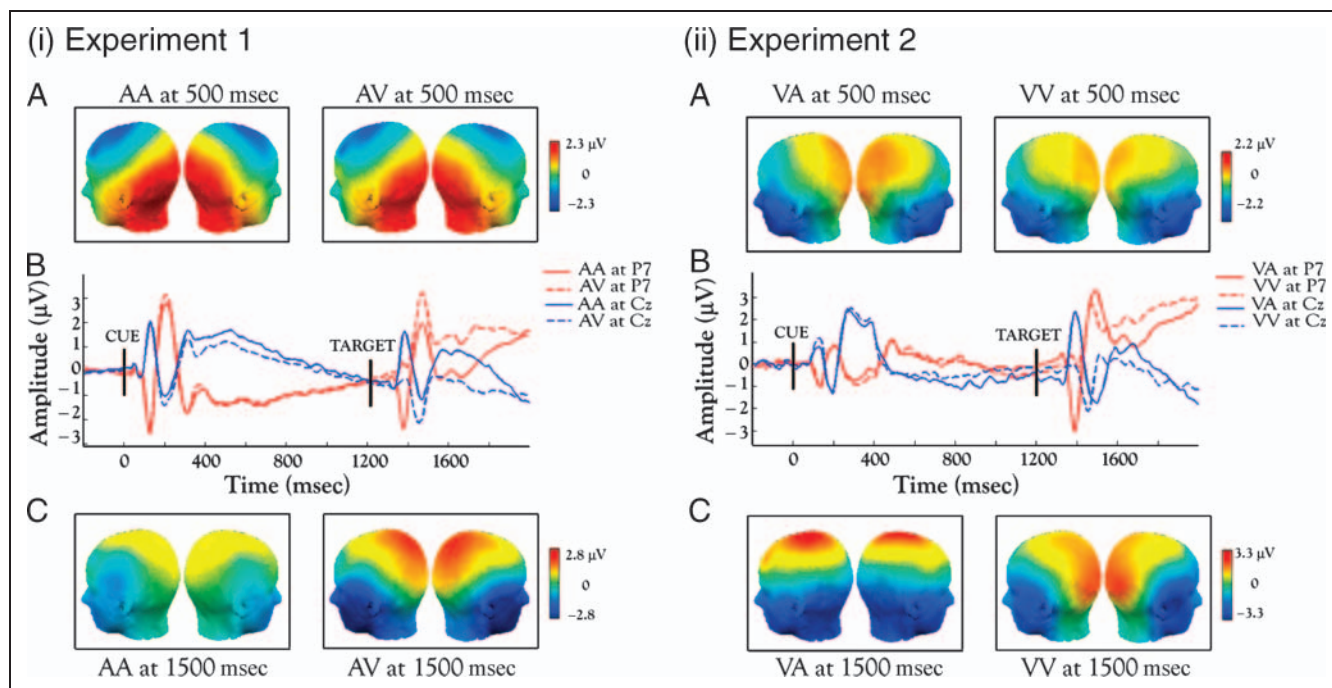


Figure 3. Grand-average ERPs. (i) Experiment 1: (A) Scalp distributions for the AA (left) and AV (right) conditions averaged over several conditions, namely, location/identity, left/right target presentation, and target type (i.e., Target 1/Target 2) are shown at 500 msec (A) and 1500 msec (C). Time courses extending from cue onset till the end of the trial for the two conditions at the midline central scalp electrode (Cz) and left parietal scalp electrode (P7) are shown in (B). The same time course is illustrated for Experiment 2 in (ii).

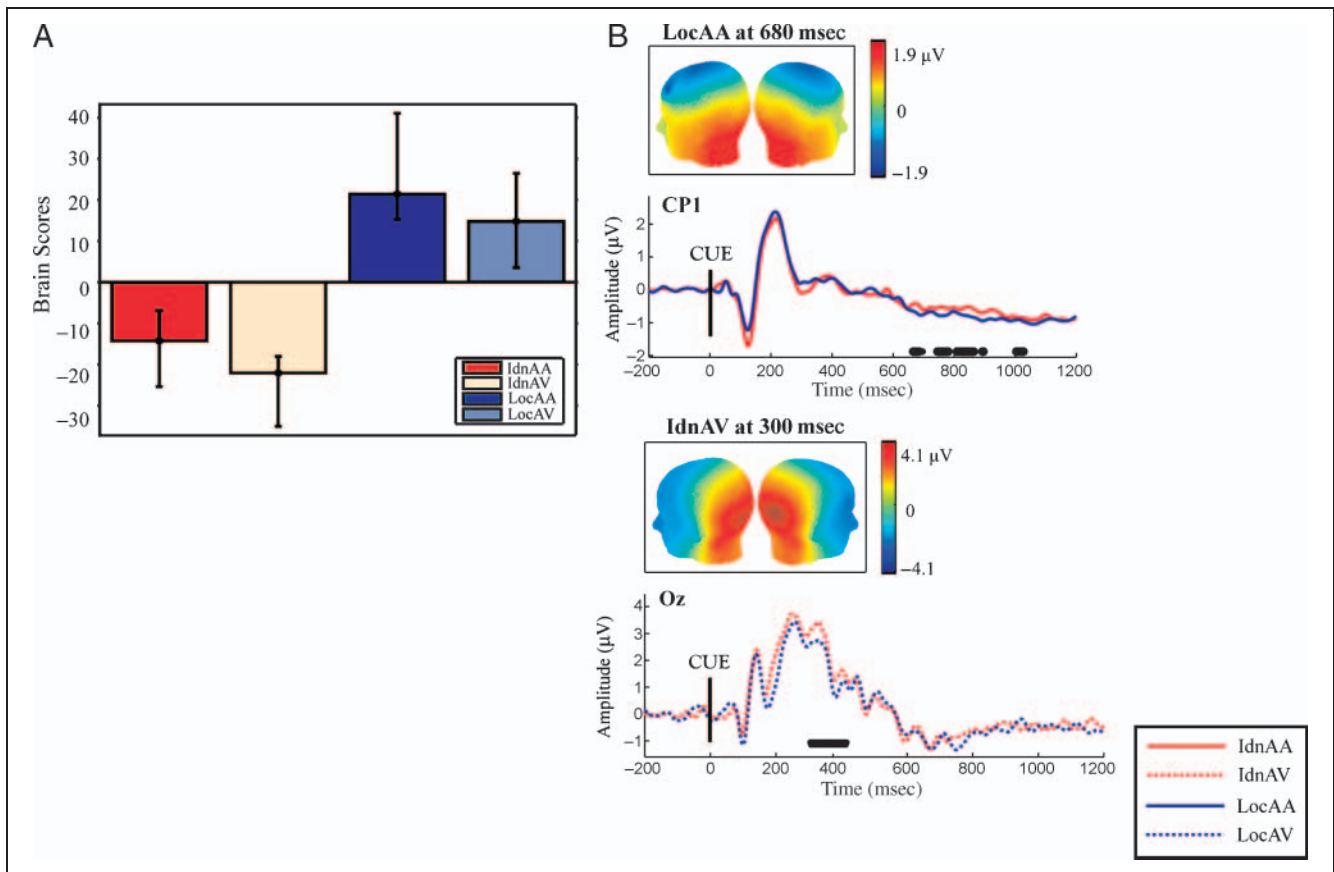


Figure 4. Location versus identity preparation (Experiment 1). The bar graph (A) represents the task-dependent contrast for the first LV (49.05% cross-block covariance, $p < .012$, location vs. identity). The topography map for the condition with the largest amplitude deflection for the particular task-dependent contrast (A) was displayed. Preparation for location processing was maximally expressed at CP1 electrodes with larger amplitudes for location compared to identity conditions between 750 and 1000 msec in anticipation of an auditory target (B). Preparation for shape processing was associated with larger amplitudes between 300 and 500 msec at central occipital channels. Bootstrap ratios ≥ 3 (approx. 99% confidence interval) are displayed below the grand-averaged waveforms for each pair of conditions.

electrodes with a more negative amplitude between 600 and 1000 msec when compared to identity AA conditions (Figure 4B). Preparation for shape processing was captured at occipital electrodes between 300 and 500 msec with a more positive amplitude for identity compared to the location AV conditions (Figure 4C).

Preparation for pitch processing in contrast to shape processing was captured maximally at left temporo-parietal channels with increased positivity between 300 and 750 msec (Figure 5B). Comparing location preparation for AA and AV conditions, we found that location AA conditions exhibited larger, more positive amplitude deflections across posterior parietal electrodes between 300 and 550 msec (Figure 5C). Preparation for shape compared to pitch processing was associated with a more positive amplitude between 1000 and 1200 msec at occipital channels (Figure 5D).

Experiment 2

In Experiment 2, an effect of Target modality (LV1 = 46.31% cross-block covariance, $p < .008$; Figure 6A) and Task (LV2 =

34.91% cross-block covariance, $p < .017$; Figure 7A) were observed (Table 1B).

Preparation for pitch processing was captured at right frontal electrodes with a more positive amplitude deflection extending between 700 and 1150 msec for identity VA compared to identity VV conditions (Figure 6B). Shape processing was detected at left posterior-occipital channels between 750 and 1150 msec with a larger positive amplitude for identity VV compared to identity VA conditions (Figure 6D). By contrasting location preparation processing for VA and VV conditions, we found that location processing in anticipation of an auditory target was associated with a more positive amplitude between 450 and 600 msec at right posterior parietal electrodes (Figure 6B).

Preparing to respond to the location compared to the identity of an auditory target in cross-modal VA conditions was associated with a more positive amplitude deflection between 300 and 450 msec across central channels (Figure 7B). In the intramodal VV condition, preparation for location processing was expressed maximally across posterior parietal electrodes between 900 and 1200 msec with a larger, more negative amplitude for location compared to identity conditions (Figure 7C).

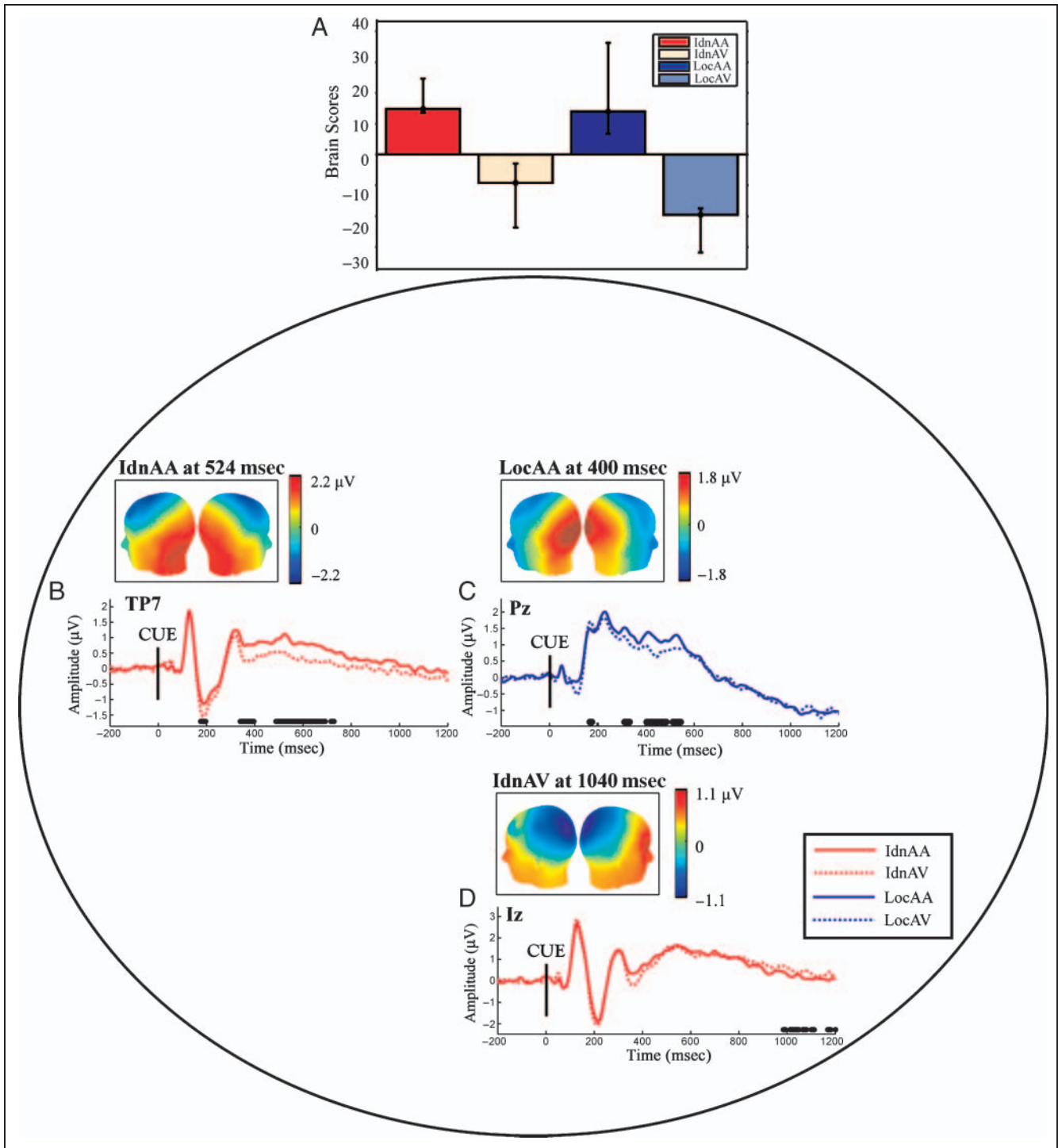


Figure 5. Effect of target modality (Experiment 1). The bar graph (A) represents the task-dependent contrast for the second LV (40.50% cross-block covariance, $p < .040$, auditory vs. visual). The topography map for the condition with the largest amplitude deflection for the particular task-dependent contrast (A) was displayed. Grand-average ERPs for identity AA compared to AV conditions with more positive amplitude deflections for AA conditions at TP7 between 300 and 750 msec are displayed in (B). Preparation for location processing was maximally expressed at posterior parietal electrodes with larger amplitudes for AA compared to AV conditions between 300 and 550 msec (C). Preparation for shape in contrast to pitch processing was associated with larger amplitudes between 1000 and 1200 msec at central occipital channels (D). Bootstrap ratios ≥ 3 (approx. 99% confidence interval) are displayed below the grand-averaged waveforms for each pair of conditions.

Table 1. Latent Variable Values, Percent Cross-block Covariance, and Significance Values for ST-PLS Analysis

	<i>Target Modality</i>	<i>Task</i>
A. Auditory cue (Experiment 1)	LV2 = 23.9, 40.50% cross-block covariance, $p < .040$	LV1 = 26.3, 49.05% cross-block covariance, $p < .012$
B. Visual cue (Experiment 2)	LV1 = 36.2, 46.31% cross-block covariance, $p < .008$	LV2 = 31.4, 34.91% cross-block covariance, $p < .017$

For both Experiments 1 and 2, we observed (1) an effect of target modality: intramodal vs. cross-modal (40.50% cross-block covariance, $p < .040$ for Experiment 1; 46.31% cross-block covariance, $p < .008$ for Experiment 2), and (2) an effect of Task (i.e., location vs. identity) and Target modality (49.05% cross-block covariance, $p < .012$ for Experiment 1; 34.91% cross-block covariance, $p < .017$ for Experiment 2).

Source Analysis Results

In source space, a main effect of response rule was observed (Experiment 1, $p < .05$; Experiment 2, $p < .006$). Location processing in anticipation of both auditory and visual targets was captured at the centro-medial regional source between 600 and 1100 msec with a larger positivity for location compared to identity conditions in Experiment 1 (Figure 8A), and at the parietal medial regional source between 600 and 1050 msec, with increased positivity for location compared to identity conditions in Experiment 2 (Figure 9A).

Preparation for pitch processing was associated with a larger, more positive amplitude at the right auditory radial dipole between 400 and 600 msec for identity compared to the location AA condition (Figure 8B). In the second experiment, pitch processing was also captured at the right auditory radial dipole between 300 and 600 msec for identity in contrast to the location VA condition (Figure 9B).

Shape processing was associated with a larger, more negative amplitude at the left visual central–inferior dipole between 700 and 1050 msec for identity compared to the location AV condition (Figure 8C) and, in Experiment 2, also between 700 and 1200 msec with larger, more negative amplitude for identity compared to the location VV condition (Figure 9C).

The timing of the experimental effects identified with bootstrap estimations of standard errors in source space differed from those calculated in electrode space. There are several interpretations for the observed temporal differences between the two analysis techniques.

In general, ST-PLS is less able to detect stable experimental effects in situations in which there is large variability between subjects and within conditions. In the present application of source modeling, the hybrid model was held fixed for each participant and was used as a spatio-temporal filter to derive the source waveforms for each condition. The model was adjusted for each participant to minimize percent residual variance (i.e., maintain it below 3%); however, the model fit may have been variable between participants, leading to fewer time points where the salience was greater than three times the standard error (i.e., exhibiting a bootstrap ratio ≥ 3). Furthermore, because electrodes captured a summed signal from simultaneously active brain sources, the effects identified may be the result of multiple neural sources (Lobaugh et al., 2001) and may differ from those identified using the spatial filters de-

rived from BESA. Finally, another reason for the temporal differences between electrodes and source waveform data relates to the use of the nonrotated ST-PLS technique. The nonrotated version of ST-PLS applied in source space restricted the contrasts to hypothesized experimental effects. Thus, these results obtained in source space are not mutually independent, and do not represent the largest percent of cross-block covariance between the experimental design and neuroelectrical brain activity identified with the original version of ST-PLS with SVD (see above).

DISCUSSION

The present study examined spatio-temporal patterns underlying “what” and “where” preparatory processes. On the behavioral side, participants’ RTs were fastest in response to location conditions, in particular, when the cue was followed by the visual target (Figure 2). Accuracy was high for both experiments, and was highest for location conditions. Although no differences in accuracy were detected for Experiment 2, we acknowledge that there is a difference in the difficulty of response mapping between location and identity conditions.

Faster RTs to visual targets preceded by valid auditory or visual cues have been reported previously for spatial discrimination tasks (Diaconescu et al., 2008; Spence & Driver, 1997; Proctor, Dutta, Kelly, & Weeks, 1994). Furthermore, valid auditory and visual cues have been shown to facilitate performance to visual targets in particular, suggesting that the auditory system might be in the service of the visual–motor system during “where” processes, but not vice-versa (Kubovy & Van Valkenburg, 2001). In the auditory modality, RTs to location conditions were significantly faster than those to identity conditions. Previous research investigating cross-modal spatial attention also showed that, in contrast to the visual and somatosensory modalities, auditory spatial tasks were aided by a spatial cue, whereas nonspatial auditory cueing tasks (e.g., target detection, pitch discrimination) did not generate robust behavioral cue facilitation effects (Spence & Driver, 1994). Spence and Driver (1994) proposed that this effect reflects the differences in the fundamental properties of sensory modalities with the visual and tactile receptor coding intrinsically spatial and the coding of the auditory system intrinsically tonotopic.

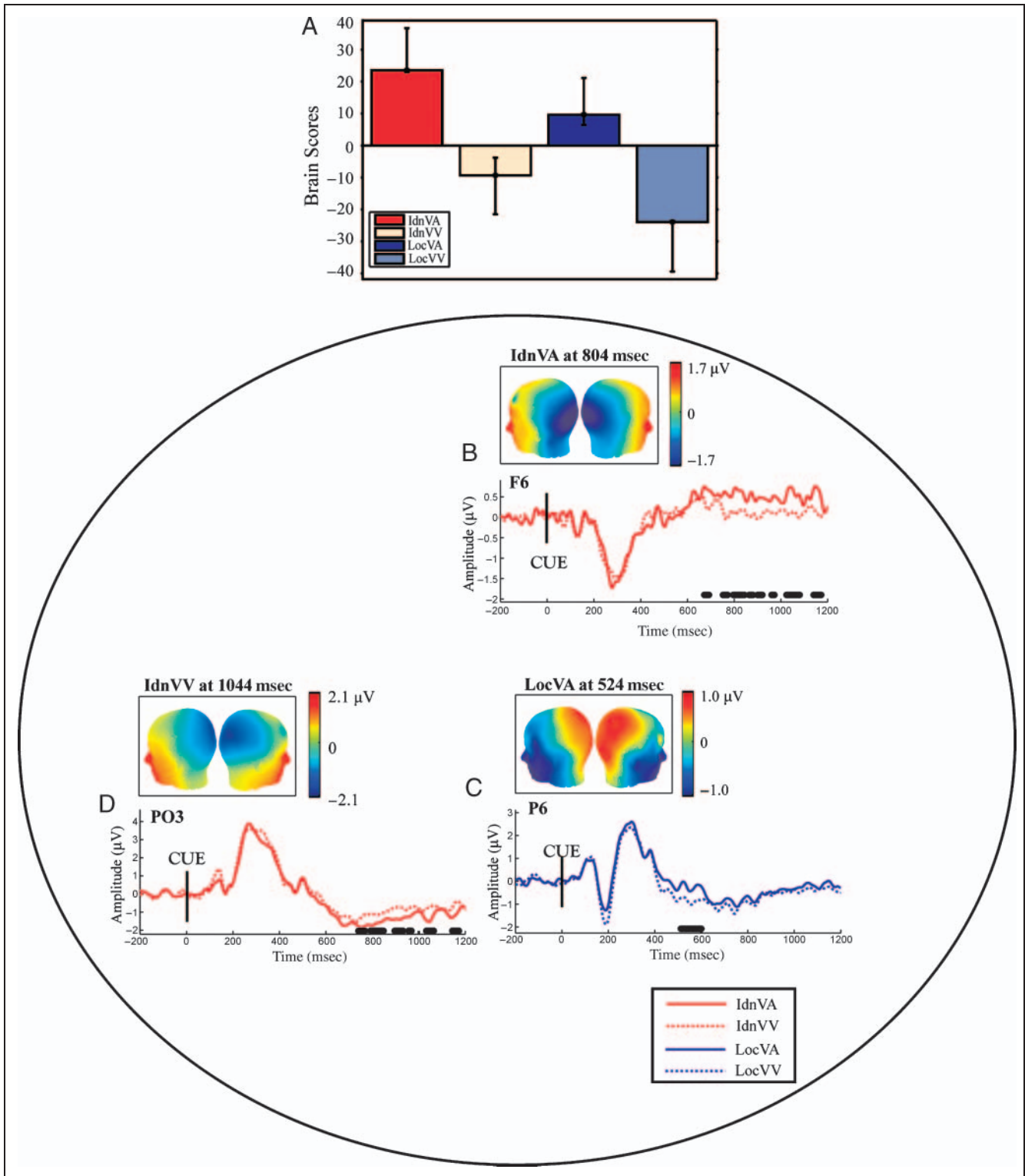


Figure 6. Effect of target modality (Experiment 2). The bar graph (A) represents the task-dependent contrast for the first IV (46.31% cross-block covariance, $p < .008$, auditory vs. visual). The topography map for the condition with the largest amplitude deflection for the particular task-dependent contrast (A) was displayed. Grand-average ERPs for identity VA compared to VV conditions with more positive amplitude deflections for VA conditions across frontal channels between 700 and 1150 msec are displayed in (B). Preparation for location processing was maximally expressed at right posterior parietal electrodes with more positive amplitudes for VA compared to VV conditions between 450 and 600 msec (C). Preparation for shape in contrast to pitch processing was captured between 750 and 1150 msec at left posterior-occipital channels (D). Bootstrap ratios ≥ 3 (approx. 99% confidence interval) are displayed below the grand-averaged waveforms for each pair of conditions.

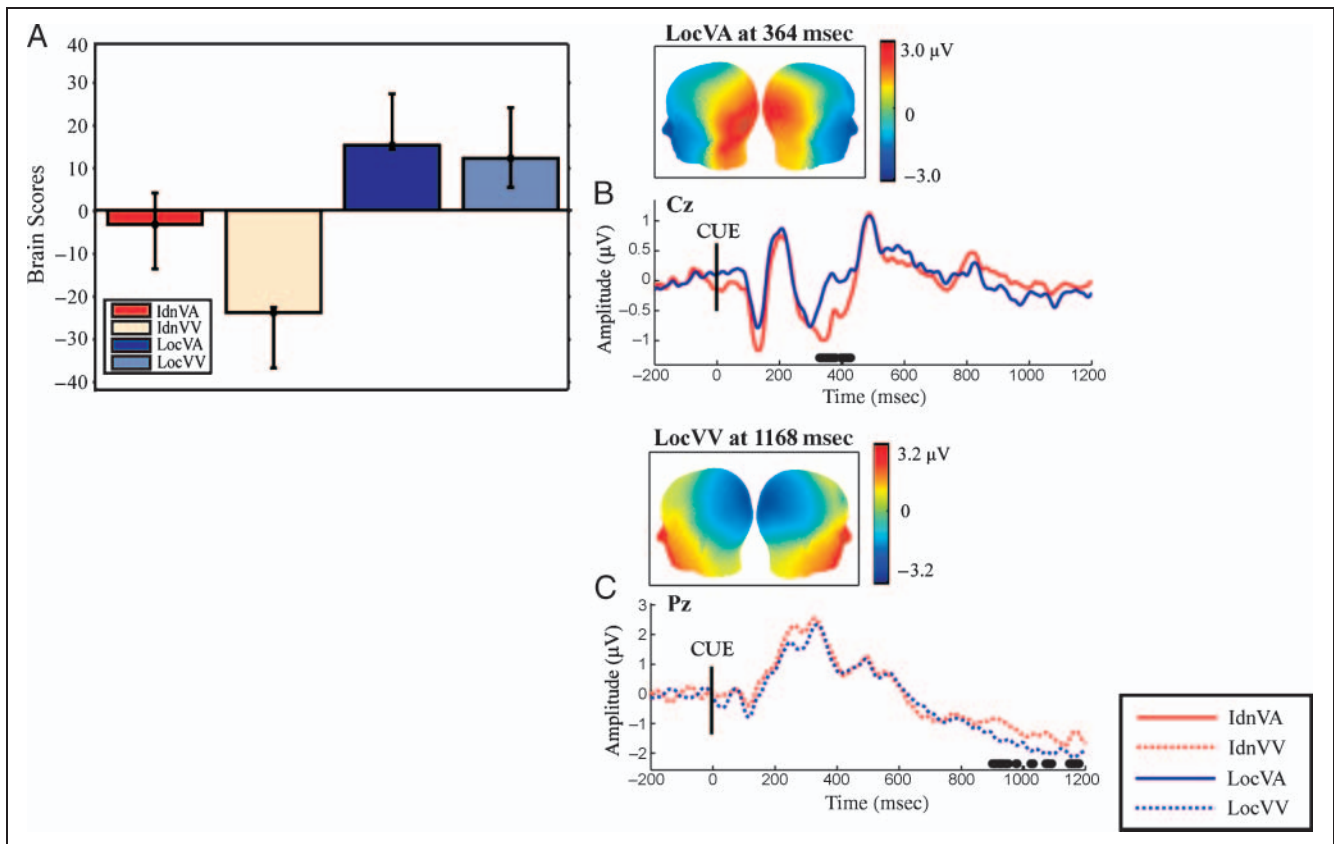


Figure 7. Location versus identity preparation (Experiment 2). The bar graph (A) represents the task-dependent contrast for the second LV (34.91% cross-block covariance, $p < .017$, location vs. identity). The topography map for the condition with the largest amplitude deflection for the particular task-dependent contrast (A) was displayed. Preparation for location processing was maximally expressed at Cz and Pz electrodes, with larger amplitudes for location compared to identity conditions between 300 and 450 msec in anticipation of an auditory target (B), and between 900 and 1200 msec in anticipation of a visual target (C). Bootstrap ratios ≥ 3 (approx. 99% confidence interval) are displayed below the grand-averaged waveforms for each pair of conditions.

The novelty of the present study is that it demonstrates the emergence of modality-specific and supramodal cue-driven “what” versus “where” preparation processes in anticipation of visual and auditory targets using identical cues. It is important to emphasize that the task effects reported were captured in the interval between the cue and the target onset. Previous studies that explored differences between “what” and “where” processes in the visual and auditory systems used blocked designs in which participants responded to a particular task set: stimulus location or stimulus identity (Alain et al., 2001, 2009; Altmann et al., 2007; Maeder et al., 2001; Shen et al., 1999; Courtney et al., 1996; Haxby et al., 1991). In the present study, participants were required to prepare to respond to target location or target identity using rapidly alternating cues. The cues were maintained constant while target modality was blocked. The results of the present study suggest that, despite the use of identical cues, preparatory processes differ on the basis of the preceding response rules (“location” vs. “identity”).

Preparation for location processing was associated with increased amplitudes at central and parietal channels between 600 and 1200 msec for Experiments 1 and 2 using

cues of distinct sensory modalities. Source modeling confirms that preparation for spatial localization is associated with a more positive amplitude deflection across centro-parietal areas between 600 and 1100 msec in anticipation of both auditory and visual targets.

Several timing differences were observed in the second experiment when comparing location cross-modal and intramodal conditions at posterior parietal electrode sites (i.e., between 300 and 500 msec in Experiment 1 following an auditory cue, and between 500 and 600 msec in Experiment 2 following a visual cue). This finding suggests that the latency of “what” versus “where” preparatory processes is modulated by the sensory modality of the cue stimulus used (i.e., pure tones vs. colors). In addition, it is worth pointing out that the interval between cue onset and the target stimulus was always 1250 msec in the current paradigm. It is possible that, if shorter intervals were used, an earlier intramodal versus cross-modal effect may have been detected.

Contrasting location to identity spatio-temporal patterns showed that preparation to respond to both auditory and visual target location was associated with increased cortical activity at central and posterior parietal electrodes.

Several ERP and functional neuroimaging studies have jointly shown that posterior parietal areas activate during selective attention to spatial location for both auditory and visual modalities (Salmi et al., 2007; Green et al., 2005; Jacoboni, Woods, & Mazziotta, 1998).

Although a posterior parietal network has consistently been implicated in the control of visual spatial attention, recent studies have shown that cued orienting to the location of auditory targets also recruits a distributed posterior parietal network including the bilateral precuneus (BA 7), superior parietal lobule (Smith et al., 2010; Wu, Weissman, Roberts, & Woldorff, 2007; Shomstein & Yantis, 2006), and the intraparietal sulcus (Cohen et al., 2004). Using ERPs and cue–target spatial localization tasks, Green et al. (2005) mapped cue-driven electrical potentials to temporo-parietal channels in preparation for both auditory and visual targets.

In contrast to location, identity preparation processes were detected at modality-specific temporal and occipital

sites 300–750, 300–500, and 1000–1200 msec in anticipation of auditory and visual targets, respectively. Modality-specific preparation processes for target identity were replicated in the second experiment, following a visual cue. In source space, we found that preparation for pitch was captured at the auditory radial dipole between 300 and 600 msec, and preparation for shape processing was detected across the visual central inferior dipole between 750 and 1150 msec in both Experiments 1 and 2.

Previous ERP and functional neuroimaging studies also reported a double dissociation for sound identity (i.e., pitch processing) and sound localization with anterior STG, STS, planum polare, and inferior frontal gyrus responsible for “what” processes and PT, posterior STG, and inferior and superior parietal cortices responsible for sound localization (Altmann et al., 2007; Alain et al., 2001). A recent study from Hill and Miller (2010) investigated the cocktail party effect and demonstrated that

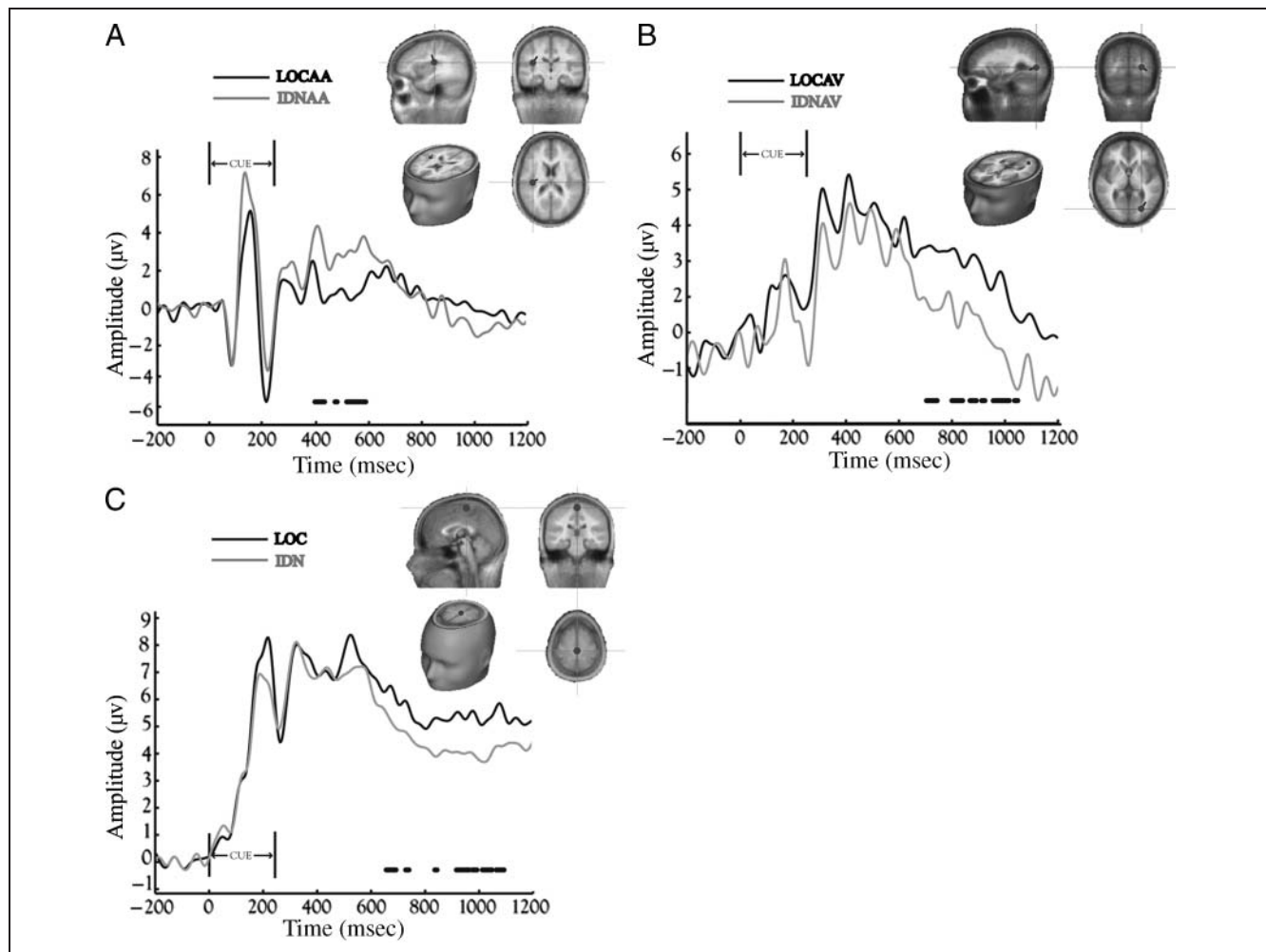


Figure 8. Source waveforms location versus identity preparation (Experiment 1). Increased positivity at the right auditory radial dipole between 400 and 600 msec was observed for identity compared to the location AA conditions (A). Increased negativity at the left visual central-inferior dipole between 700 and 1050 msec was observed for identity compared to the location AV conditions (B). Source waveforms averaged for AA and AV conditions exhibit larger positive amplitude deflections for location compared to identity conditions at the centro-medial regional source between 600 and 1100 msec (C). Bootstrap ratios > 3 (approx. 99% confidence interval) are displayed above the grand averaged source waveforms for each pair of conditions.

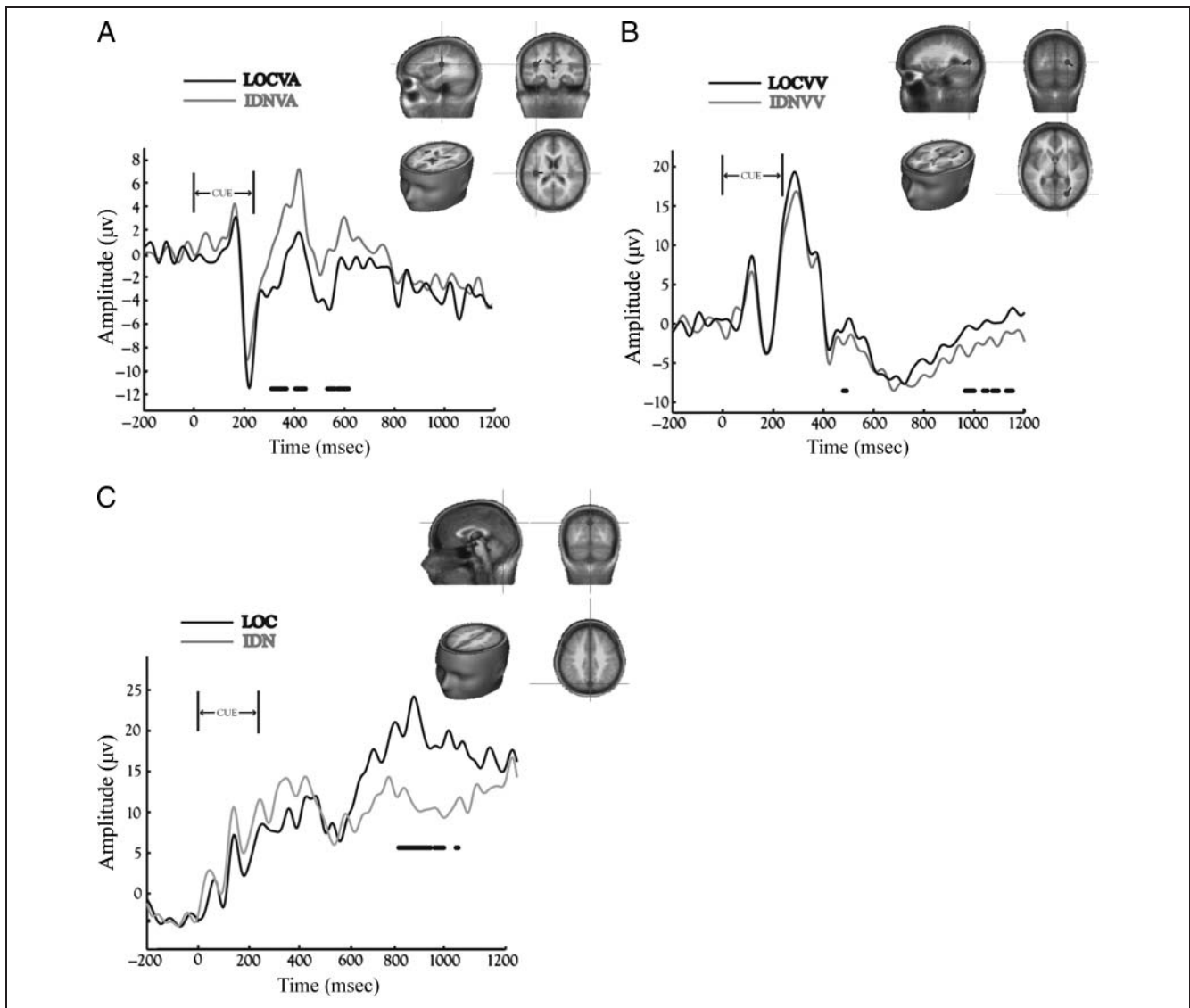


Figure 9. Source waveforms location versus identity preparation (Experiment 2). Increased positivity at the right auditory radial dipole between 300 and 600 msec was observed for identity in contrast to location VA conditions (A). Increased negativity at the left visual central-inferior dipole between 700 and 1200 msec was observed for identity compared to location VV conditions (B). Source waveforms averaged for VA and VV conditions exhibit increased positivity for location compared to identity conditions at the parietal medial regional between 600 and 1050 msec (C). Bootstrap ratios > 3 (approx. 99% confidence interval) are displayed above the grand averaged source waveforms for each pair of conditions.

selective attention to the location of speech modulated BOLD activity in the intraparietal sulcus, whereas selective attention toward pitch activated the STS (Hill & Miller, 2010). Selective attention to visual attributes of a stimulus, such as color or shape, enhances activity in extrastriate cortex (Chawla, Rees, & Friston, 1999; Hillyard & Anllo-Vento, 1998).

ERP studies investigating cue-related deployment of attention to auditory or visual stimuli demonstrated a late, sustained, sensory-specific activation 400 msec after cue onset biasing auditory and visual processes in preparation for the upcoming targets (Foxe, Simpson, Ahlfors, & Saron, 2005). We observed a similar biasing of activity across auditory areas between 300 and 750 msec in preparation for pitch processing relative to shape processing. Preparation

for shape processing was associated with increased amplitudes over occipital channels between 700 and 1200 msec and 750 and 1200 msec following auditory and visual cues, respectively.

In summary, the present study suggests that “what” and “where” processes can be distinguished following rapidly alternating cues that signal participants to respond to the identity or location of an upcoming target. The study demonstrated in two sets of experiments that location preparation modulated cortical activity across central and parietal regions in anticipation of both visual and auditory targets. Conversely, preparation for target identity was associated with increased cortical activity in occipital areas in anticipation of visual targets and fronto-temporal areas in anticipation of auditory targets.

Using a novel paradigm, the study supports the notion of a division of labor in auditory and visual sensory systems when participants are cued to respond to the location or identity of an upcoming target. In contrast to pitch or shape preparation, which exhibited modality-specific spatio-temporal patterns, preparation for target localization showed supramodal, spatio-temporal patterns with increased cortical activity at multimodal, centroparietal sites in anticipation of both auditory and visual targets.

Acknowledgments

This research was supported by grants from the Canadian Institutes of Health Research, the Natural Sciences and Engineering Research Council of Canada, and the JS McDonnell Foundation.

Reprint requests should be sent to Andreea Oliviana Diaconescu or Anthony Randal McIntosh, Rotman Research Institute, Baycrest Centre, 3560 Bathurst Street, Toronto, Ontario, Canada M6A 2E1, or via e-mail: adiaconescu@rotman-baycrest.on.ca; rmcintosh@rotman-baycrest.on.ca.

REFERENCES

- Alain, C., Arnott, S. R., Hevenor, S., Graham, S., & Grady, C. L. (2001). "What" and "where" in the human auditory system. *Proceedings of the National Academy of Sciences, U.S.A.*, *98*, 12301–12306.
- Alain, C., McDonald, K. L., Kovacevic, N., & McIntosh, A. R. (2009). Spatiotemporal analysis of auditory "what" and "where" working memory. *Cerebral Cortex*, *19*, 305–314.
- Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences*, *4*, 267–278.
- Altmann, C. F., Bledowski, C., Wibral, M., & Kaiser, J. (2007). Processing of location and pattern changes of natural sounds in the human auditory cortex. *Neuroimage*, *35*, 1192–1200.
- Amedi, A., von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S., & Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. *Experimental Brain Research*, *166*, 559–571.
- Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neuroscience*, *20*, 303–330.
- Arnott, S. R., Binns, M. A., Grady, C. L., & Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *Neuroimage*, *22*, 401–408.
- Arnott, S. R., Grady, C. L., Hevenor, S. J., Graham, S., & Alain, C. (2005). The functional organization of auditory working memory as revealed by fMRI. *Journal of Cognitive Neuroscience*, *17*, 819–831.
- Bushara, K. O., Weeks, R. A., Ishii, K., Catalan, M. J., Tian, B., Rauschecker, J. P., et al. (1999). Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nature Neuroscience*, *2*, 759–766.
- Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*, 649–657.
- Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience*, *2*, 671–676.
- Cohen, Y. E., Russ, B. E., & Gifford, G. W., III (2005). Auditory processing in the posterior parietal cortex. *Behavioural and Cognitive Neuroscience Reviews*, *4*, 218–231.
- Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, *22*, 319–349.
- Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1996). Object and spatial visual working memory activate separate neural systems in human cortex. *Cerebral Cortex*, *6*, 39–49.
- Dale, C. L., Simpson, G. V., Foxe, J. J., Luks, T. L., & Worden, M. S. (2008). ERP correlates of anticipatory attention: Spatial and non-spatial specificity and relation to subsequent selective attention. *Experimental Brain Research*, *188*, 45–62.
- Damasio, A. R., Damasio, H., & Van Hoesen, G. W. (1982). Prosopagnosia: Anatomic basis and behavioral mechanisms. *Neurology*, *32*, 331–341.
- Diaconescu, A. O., Kovacevic, N., & McIntosh, A. R. (2008). Modality-independent processes in cued motor preparation revealed by cortical potentials. *Neuroimage*, *42*, 1255–1265.
- Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Statistical Science*, *1*, 54–77.
- Eimer, M., van Velzen, J., & Driver, J. (2002). Cross-modal interactions between audition, touch, and vision in endogenous spatial attention: ERP evidence on preparatory states and sensory modulations. *Journal of Cognitive Neuroscience*, *14*, 254–271.
- Foxe, J. J., Simpson, G. V., Ahlfors, S. P., & Saron, C. D. (2005). Biasing the brain's attentional set: I. Cue driven deployments of intersensory selective attention. *Experimental Brain Research*, *166*, 370–392.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*, 20–25.
- Green, J. J., Teder-Salejarvi, W. A., & McDonald, J. J. (2005). Control mechanisms mediating shifts of attention in auditory and visual space: A spatio-temporal ERP analysis. *Experimental Brain Research*, *166*, 358–369.
- Haxby, J. V., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., et al. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *88*, 1621–1625.
- Heinze, H. J., Mangun, G. R., Burchert, W., Hinrichs, H., Scholz, M., Munte, T. F., et al. (1994). Combined spatial and temporal imaging of brain activity during visual selective attention in humans. *Nature*, *372*, 543–546.
- Hill, K. T., & Miller, L. M. (2010). Auditory attentional control and selection during cocktail party listening. *Cerebral Cortex*, *20*, 583–590.
- Hillyard, S. A., & Anillo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences, U.S.A.*, *95*, 781–787.
- Hopf, J. M., & Mangun, G. R. (2000). Shifting visual attention in space: An electrophysiological analysis using high spatial resolution mapping. *Clinical Neurophysiology*, *111*, 1241–1257.
- Iacoboni, M., Woods, R. P., & Mazziotta, J. C. (1998). Bimodal (auditory and visual) left frontoparietal circuitry for sensorimotor integration and sensorimotor learning. *Brain*, *121*, 2135–2143.
- Kubovy, M., & Van Valkenburg, D. (2001). Auditory and visual objects. *Cognition*, *80*, 97–126.
- Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science*, *240*, 740–749.

- Lobaugh, N. J., West, R., & McIntosh, A. R. (2001). Spatiotemporal analysis of experimental differences in event-related potential data with partial least squares. *Psychophysiology*, *38*, 517–530.
- Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., et al. (2001). Distinct pathways involved in sound recognition and localization: A human fMRI study. *Neuroimage*, *14*, 802–816.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.
- McIntosh, A. R., Bookstein, F. L., Haxby, J. V., & Grady, C. L. (1996). Spatial pattern analysis of functional brain images using Partial Least Squares. *Neuroimage*, *3*, 143–157.
- McIntosh, A. R., & Lobaugh, N. J. (2004). Partial least squares analysis of neuroimaging data: Applications and advances. *Neuroimage*, *23*(Suppl. 1), S250–S263.
- Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural Brain Research*, *6*, 57–77.
- Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory visual–auditory object recognition in humans: A high-density electrical mapping study. *Cerebral Cortex*, *14*, 452–465.
- Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory–visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Brain Research, Cognitive Brain Research*, *14*, 115–128.
- Molholm, S., Sehatpour, P., Mehta, A. D., Shpaner, M., Gomez-Ramirez, M., Ortigue, S., et al. (2006). Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. *Journal of Neurophysiology*, *96*, 721–729.
- Murray, M. M., Michel, C. M., Grave de Peralta, R., Ortigue, S., Brunet, D., Gonzalez Andino, S., et al. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *Neuroimage*, *21*, 125–135.
- Poremba, A., Saunders, R. C., Crane, A. M., Cook, M., Sokoloff, L., & Mishkin, M. (2003). Functional mapping of the primate auditory system. *Science*, *299*, 568–572.
- Proctor, R. W., Dutta, A., Kelly, P. L., & Weeks, D. J. (1994). Cross-modal compatibility effects with visual–spatial and auditory–verbal stimulus and response sets. *Perception & Psychophysics*, *55*, 42–47.
- Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience*, *18*, 2188–2199.
- Salmi, J., Rinne, T., Degerman, A., Salonen, O., & Alho, K. (2007). Orienting and maintenance of spatial attention in audition and vision: Multimodal and modality-specific brain activations. *Brain Structure and Function*, *212*, 181–194.
- Shen, L., Hu, X., Yacoub, E., & Ugurbil, K. (1999). Neural correlates of visual form and visual spatial processing. *Human Brain Mapping*, *8*, 60–71.
- Shomstein, S., & Yantis, S. (2006). Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. *Journal of Neuroscience*, *26*, 435–439.
- Smith, D. V., Davis, B., Niu, K., Healy, E. W., Bonilha, L., Fridriksson, J., et al. (2010). Spatial attention evokes similar activation patterns for visual and auditory stimuli. *Journal of Cognitive Neuroscience*, *22*, 347–361.
- Spence, C., & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 555–574.
- Spence, C., & Driver, J. (1997). On measuring selective attention to an expected sensory modality. *Perception & Psychophysics*, *59*, 389–403.
- Weeks, R. A., Aziz-Sultan, A., Bushara, K. O., Tian, B., Wessinger, C. M., Dang, N., et al. (1999). A PET study of human auditory spatial processing. *Neuroscience Letters*, *262*, 155–158.
- Wu, C. T., Weissman, D. H., Roberts, K. C., & Woldorff, M. G. (2007). The neural circuitry underlying the executive control of auditory spatial attention. *Brain Research*, *1134*, 187–198.