Grabbing Your Ear: Rapid Auditory–Somatosensory Multisensory Interactions in Low-level Sensory Cortices Are Not Constrained by Stimulus Alignment

Multisensory interactions are observed in species from single-cell organisms to humans. Important early work was primarily carried out in the cat superior colliculus and a set of critical parameters for their occurrence were defined. Primary among these were temporal synchrony and spatial alignment of bisensory inputs. Here, we assessed whether spatial alignment was also a critical parameter for the temporally earliest multisensory interactions that are observed in lower-level sensory cortices of the human. While multisensory interactions in humans have been shown behaviorally for spatially disparate stimuli (e.g. the ventriloquist effect), it is not clear if such effects are due to early sensory level integration or later perceptual level processing. In the present study, we used psychophysical and electrophysiological indices to show that auditory–somatosensory interactions in humans occur via the same early sensory mechanism both when stimuli are in and out of spatial register. Subjects more rapidly detected multisensory than unisensory events. At just 50 ms post-stimulus, neural responses to the multisensory ‘whole’ were greater than the summed responses from the constituent unisensory ‘parts’. For all spatial configurations, this effect followed from a modulation of the strength of brain responses, rather than the activation of regions specifically responsive to multisensory pairs. Using the local auto-regressive average source estimation, we localized the initial auditory–somatosensory interactions to auditory association areas contra-lateral to the side of somatosensory stimulation. Thus, multisensory interactions can occur across wide peripersonal spatial separations remarkably early in sensory processing and in cortical regions traditionally considered unisensory.

Keywords: area CM, cross-modal, event-related potential (ERP), human, LAURA source estimation, redundant signals effect (RSE)

Introduction

Experiencing the external world relies on information conveyed to our different senses. Yet, it is likewise clear that we do not perceive the external world as divided between these senses, but rather as integrated representations. By extension, our perceptions are thus almost always multisensory. Although we are largely unaware of how inputs from the various senses influence perceptions otherwise considered unisensory, multisensory interactions can have dramatic effects on our performance of everyday tasks. For example, speech comprehension in noisy environments can be greatly enhanced by scrutiny of a speaker’s mouth and face (Sumby and Pollack, 1954). Another is theater-goers attributing actors’ voices as coming from the movie screen in front of them, rather than from the sound system surrounding them - an instance of ventriloquism (Klemm, 1909; Thomas, 1941; Radeau, 1994). Similarly, auditory input can influence the perceived texture of surfaces being touched (Jousmäki and Hari, 1998), the perceived direction of visual motion (Sekuler et al., 1997), as well as both qualitative (Stein et al., 1996) and quantitative (Shams et al., 2000) features of visual stimuli.

To achieve these effects, different sensory inputs must be combined, with neurons responding to multiple sensory modalities (Stein and Meredith, 1993). Traditionally, it was believed that such multisensory interactions were reserved for higher cortical levels and occurred relatively late in time, with information along the different senses remaining segregated at lower levels and earlier latencies (e.g. Jones and Powell, 1970; Okajima et al., 1995; Massaro, 1998; Schröger and Widmann, 1998). Recent evidence, however, has demonstrated that multisensory interactions are a fundamental component of neural organization even at the lowest cortical levels. Anatomical tracing studies have revealed direct projections to visual areas V1 and V2 from primary (Falchier et al., 2002) as well as association areas (Falchier et al., 2002; Rockland and Ojima, 2003) of macaque auditory cortex. Others have not only observed similar patterns of projections from somatosensory (Fu et al., 2003) and visual systems (Schroeder and Foxe, 2002) that terminate in belt and parabelt auditory association areas, but also describe the laminar activation profile of multisensory convergence in these auditory regions as consistent with feedforward inputs (Schroeder et al., 2001, 2003; Schroeder and Foxe, 2002). The implication of these studies in nonhuman primates is that the initial stages of sensory processing already have access to information from other sensory modalities. In strong agreement are the repeated observations in humans of nonlinear neural response interactions to multisensory stimulus pairs versus the summed responses from the constituent unisensory stimuli at early (<100 ms) latencies (Giard and Peronnet, 1998; Foxe et al., 2000; Molholm et al., 2002; Lükenhöner et al., 2002; Fort et al., 2002; Gobbelé et al., 2003; see also Murray et al., 2004) and within brain regions traditionally held to be ‘unisensory’ in their function (Macaluso et al., 2000; Calvert, 2001; Foxe et al., 2002).

Despite such evidence for multisensory interactions as a fundamental phenomenon, the principles governing multisensory interactions in human cortex remain largely unsolved, particularly for effects believed to occur at early
sensory-perceptual processing stages. In animals, three such ‘rules’ have been formulated, based largely on electrophysiological recordings of neurons in the superior colliculus of the cat (see Stein and Meredith, 1993). The ‘spatial rule’ states that multisensory interactions are dependent on the spatial alignment and/or overlap of receptive fields responsive to the stimuli. That is, facilitative multisensory interactions can be observed even when stimuli are spatially misaligned in their external coordinates, provided that the responsive neurons contain overlapping representations. If these representations do not overlap, no interaction is seen and in many cases, even response depression is observed (i.e. inhibitory interactions). The ‘temporal rule’ states that multisensory interactions are also dependent on the coincidence of the neural responses to different stimuli (albeit within a certain window). Stimuli with overlapping neural responses yield interactions, whereas those yielding asynchronous responses do not. Finally, the ‘inverse effectiveness’ rule states that the strongest interactions are achieved with stimuli, which when presented in isolation, are minimally effective in eliciting a neural response. Collectively, these rules provide a framework for understanding the neurophysiological underpinnings and functional consequences of multisensory interactions.

In parallel, psychophysical data from humans have similarly begun to describe the circumstances and limitations on the occurrence of multisensory interactions. Such studies generally indicate that multisensory interactions are subject to limitations in terms of the spatial misalignment and temporal asynchrony of stimuli. Some suggest that stimuli must be presented within ~30–40° of each other (though not necessarily within the same hemispace) for effects to occur either in the case of facilitating stimulus detection (Hughes et al., 1994; Harrington and Peck, 1998; Forster et al., 2002; Frassinetti et al., 2002) or in the case of influencing localization judgments (e.g. Caclin et al., 2002), with wider separations failing to generate facilitative interaction effects (see also Stein et al., 1989, for a behavioral demonstration in cats). Others have found interaction effects across wider spatial disparities for tasks requiring visual intensity judgments, sometimes even irrespective of the locus of sounds (Stein et al., 1996; though see Odgaard et al., 2003). Others emphasize the spatial ambiguity of a stimulus over its absolute position, such as in the cases of ventriloquism and shifts in attention (e.g. Bertelson, 1998; Spence and Driver, 2000; Hairston et al., 2003). The variability in these findings raises the possibility that task requirements may influence spatial limitations on, and perhaps also the mechanisms of, multisensory interactions. Another possibility is that some varieties of multisensory interactions occur within neurons and/or brain regions with large spatially insensitive receptive fields. Similarly unresolved is to what extent such behavioral phenomena reflect early (i.e. <100 ms) interactions of the variety described above or later effects on brain responses that already include multisensory interactions - a topic of increasing investigation (Schürmann et al., 2002; Hötting et al., 2003) and speculation (Warren et al., 1981; Stein and Wallace, 1996; Pavani et al., 2000; Slutzky and Recanzone, 2001; Caclin et al., 2002; Odgaard et al., 2003). An important issue, therefore, is to define the limitations on the temporally earliest instantiations of multisensory interactions in human cortex.

It is important to point out at this juncture that the ERP technique is relatively insensitive to subcortical potentials, mainly due to the depth of these generators relative to the scalp electrodes. As such, ERPs are unlikely to index multisensory effects occurring in SC. These effects may well be earlier or simultaneous with any effects we see in cortex. Indeed, multisensory interactions have been seen to onset very early in SC neurons, especially for auditory–somatosensory (AS) combinations (see Meredith et al., 1987). Although it is difficult to relate the timing of single-unit effects in the cat to cortical effects in humans, it is nonetheless likely that multisensory interactions in human SC are relatively rapid and may well precede those that we observe in cortex in our studies. However, it is also worth noting that multisensory integration effects in SC appear to be dependent on cortical inputs (e.g. Stein and Wallace, 1996; Jiang and Stein, 2003), highlighting the importance of understanding the earliest cortical integration effects.

The goal of the present study was to investigate whether early (i.e. <100 ms) AS interactions in humans occur via similar spatiotemporal neural mechanisms, irrespective of the spatial alignment of the stimuli. Resolving this issue carries implications of relating the early timing and loci of neurophysiological instantiations of multisensory interactions with their behavioral and functional consequences. To date, previous studies from our laboratory have begun to detail the cortical circuitry of AS interactions following passive stimulus presentation. In humans, electrophysiological (EEG) measures have demonstrated AS multisensory interactions beginning at ~50 ms (Foxe et al., 2000) that were localized in a subsequent functional magnetic resonance imaging (fMRI) experiment to auditory regions of the posterior superior temporal plane (Foxe et al., 2002). These results are in strong agreement with our findings in monkeys showing feedforward AS multisensory convergence within the caudomedial (CM) belt region surrounding primary auditory cortex (Schroeder et al., 2001, 2003, 2004; Schroeder and Foxe, 2002, 2004; Fu et al., 2003), leading us to propose the localization of a human homolog of macaque area CM. Investigations of AS interactions using magnetoencephalography (MEG) have observed slightly later effects (>75 ms) that were attributed to somatosensory area SII in the hemisphere contralateral to the side of somatosensory stimulation (Lütkenhöner et al., 2002; Gobbélé et al., 2003). However, since stimuli in these previous studies were presented to the same spatial location (Foxe et al., 2000), or otherwise used dichotic (Gobbélé et al., 2003) or binaural auditory presentations (Foxe et al., 2002; Lütkenhöner et al., 2002), the question of spatial dependency remains to be addressed. However, the evidence favoring feedforward AS interactions in ‘auditory’ area CM would suggest that the presence of similarly early and located AS interactions in the case of both spatially aligned and misaligned stimuli would probably rely on the existence of large bilateral auditory receptive fields in this region. Here, we applied the methods of high-density electrophysiological recordings from the human scalp to assess whether spatially aligned and misaligned AS stimuli share a common neural mechanism of multisensory interactions. To do this, we implemented a series of analyses capable of statistically determining the timing and direction (supra- versus sub-additive) of multisensory interactions. Our analyses allow us to differentiate effects due to changes in response strength from effects due to changes in the underlying brain generators. Finally, we applied the local auto-regressive average (LAURA) distributed linear inverse method to estimate the intracranial loci of brain generators underlying the initial AS multisensory interactions for both spatially aligned and misaligned stimulus configurations.
Materials and Methods

Subjects
Twelve healthy, paid volunteers aged 20-34 years participated, with all reporting normal hearing and no neurological or psychiatric illnesses. Eleven were right-handed (Oldfield, 1971). Participants provided written, informed consent to the experimental procedures, which were approved by the local ethics board. Behavioral analyses were based on the data from all 12 participants. However, EEG data from four subjects were excluded due to excessive artifacts. Thus, all EEG analyses were based on a final group of eight subjects (mean age 25.4 years; two women; one left-handed man).

Stimuli and Task
Subjects were presented with the following stimulus conditions: (a) somatosensory stimuli alone, (b) auditory stimuli alone, (c) spatially 'aligned' AS stimulation where both stimuli were simultaneously presented to the same location (e.g. left hand and left-sided speaker), and (d) spatially 'misaligned' AS stimulation presented to different locations (e.g. left hand and right-sided speaker). In total, there were eight configurations of stimuli such that both left- and right-sided presentations were counterbalanced (Fig. 1). Left- and right-sided stimuli were separated by $-100^\circ$ within azimuth. Somatosensory stimuli were driven by DC pulses ($+5 V$; $-685 Hz$; $15 ms$ duration) through Oticon A 100 $\Omega$ bone conduction vibrators (Oticon Inc., Somerset, NJ) with 1.6 x 2.4 cm surfaces held between the thumb and index finger of each hand and away from the knuckles to prevent bone conduction of sound. To further ensure that somatosensory stimuli were inaudible, either the hands were wrapped in sound-attenuating foam ($n = 6$, with four contributing to the EEG group analyses) or earplugs were worn ($n = 6$, with four contributing to the EEG group analyses). Auditory stimuli were 30 ms white noise bursts ($70 dB$; $2.5 ms$ rise/fall time) delivered through a stereo receiver (Kenwood, model no. VR205) and speakers (JBL, model no. CM142) located next to the subjects' hands (Fig. 1). Sample trials prior to the experiment verified that sounds were clearly audible even when earplugs were worn. All subjects were tested before the experiment began to ensure that they were easily able to localize the sound stimuli to either the left or right speaker. Given the $>100^\circ$ separation between sound sources, this proved trivial for all participants. [Sound localization ability can be degraded by the use of earplugs or other ear protection devices. However, while some attenuation of fine localization ability may have occurred for those wearing earplugs in the present experiment, the wide separation used between sound sources was sufficient to make the judgment of sound location trivial for all subjects. A recent examination of the effects of horizontal and vertical electro-oculography. An artifact rejection criterion of $\pm 60 \mu V$ was used to exclude trials with excessive muscle or other noise transients. The mean ± SD acceptance rate of EEG epochs for any stimulus condition was 92.3 ± 3.8%. Accepted trials were epoched from $-100 ms$ pre-stimulus to $300 ms$ post-stimulus, and baseline activity was defined over the $-100 ms$ to $0 ms$ epoch. Event related potentials (ERPs) were computed for each of the eight stimulus configurations for each subject. For each of these ERPs, data at any artifact channels were omitted, and the remaining data were interpolated to a 117-channel electrode array that excluded the electro-oculographic and mastoid channels (3D spline; Perrin et al., 1997). Each subject's ERPs were recalculated against the average reference and normalized to their mean global field power (GFP; Lehmann and Skrandies, 1980) prior to group-averaging. GFP is equivalent to the spatial standard deviation of the scalp electric field, yields larger values for stronger fields, and is calculated as the square root of the mean of the squared value recorded at each electrode (versus the average reference).

To identify neural response interactions, ERPs to multisensory stimulus pairs were compared with the algebraic sum of ERPs to the constituent unisensory stimuli presented in isolation (Foxe et al., 2000; Murray et al., 2001; Molholm et al., 2002, 2004). The summed ERP responses from the unisensory presentations (‘sum’) should be equivalent to the ERP from the same stimuli presented simultaneously (‘pair’) if neural responses to each of the unisensory stimuli are independent. Divergence between ‘sum’ and ‘pair’ ERPs indicates nonlinear interaction between the neural responses to the multisensory stimuli. It should be noted that this approach is not sensitive to areas of multisensory convergence wherein responses to two sensory modalities might occur, but sum linearly. Next we detail how we statistically determined instances of such divergence, and by extension, nonlinear interactions.

‘Pair’ and ‘sum’ ERPs for each spatial configuration were compared using two classes of statistical tests. The first used the instantaneous GFP for each subject and experimental condition to identify changes in electric field strength. The analysis of a global measure of the ERP was in part motivated by the desire to minimize observer bias that can follow from analyses restricted to specific selected electrodes. GFP area measures were calculated (versus the $0 \mu V$ baseline) and submitted to a three-way repeated-measures ANOVA, using within-subjects factors of pair versus sum, aligned versus misaligned, and somatosensory stimulation of the left versus right hand. Given that GFP yields larger values for stronger electric fields, we were also able to determine whether any divergence was due to a larger or smaller magnitude response to the multisensory pair for each experimental condition. Observation of a GFP modulation does not exclude the possibility of a contemporaneous change in the electric field topography. Nor does it rule out the possibility of topographic modulations that nonetheless yield statistically indistinguishable GFP values. However, the observation of a GFP modulation in the absence of a ‘pair’ versus ‘sum’ topographic change is most parsimoniously explained by amplitude modulation of statistically indistinguishable generators across experimental conditions. Moreover, the direction of any amplitude change further permits us to classify multisensory interactions as facilitating/enhancing (i.e. ‘pair’ greater than the ‘sum’) or interfering with (i.e. ‘pair’ less than the ‘sum’) response magnitude.

The second class of analysis tested the data in terms of the spatiotemporal characteristics of the global electric field on the scalp
(maps). The topography (i.e. the spatial configuration of these maps) was compared over time within and between conditions. Changes in the map configuration indicate differences in the active neuronal populations in the brain (Fender, 1987; Lehmann, 1987). The method applied here has been described in detail elsewhere (Michel et al., 2001; Murray et al., 2004). Briefly, it included a spatial cluster analysis (Pascual-Marqui et al., 1995) to identify the most dominant scalp topographies appearing in the group-averaged ERPs from each condition over time. We further applied the constraint that a given scalp topography must be observed for at least 20 ms duration in the group-averaged data. From such, it is possible to summarize ERP data by a limited number of maps. The optimal number of such maps that explains the whole group-averaged data set (i.e. the group-averaged ERPs from all tested conditions, collectively) was determined by a modified cross-validation criterion (Pascual-Marqui et al., 1995). The appearance and sequence of these maps identified in the group-averaged data was then statistically verified in those from each individual subject. The moment-by-moment scalp topography of the individual subjects’ ERPs from each condition was compared with each map by means of strength-independent spatial correlation and labeled with that yielding the highest value (Michel et al., 2001). This fitting procedure yields the total amount of time a given map was observed for a given condition across subjects (i.e. its frequency over a given time period) that was then tested with a repeated-measures ANOVA, using within subjects factors of pair versus sum, aligned versus misaligned, somatosensory stimulation of the left versus right hand, and map. This analysis reveals if the ERP from a given experimental condition was more often described by one map versus another, and therefore if different generator configurations better account for particular experimental conditions.

Source Estimation

As a final step, we estimated the sources in the brain demonstrating AS multisensory interactions, using the LAURA distributed linear inverse solution (Grave de Peralta et al., 2001, 2004). This inverse solution selects the source configuration that better mimics the biophysical behavior of electric vector fields. That is, the estimated activity at one point depends on the activity at neighboring points according to electromagnetic laws. Since LAURA belongs to the class of distributed inverse solutions, it is capable of dealing with multiple simultaneously active sources of a priori unknown location. The lead field (solution space) was calculated on a realistic head model that included 4024 nodes, selected from a 6 x 6 x 6 mm grid equally distributed within the gray matter of the average brain provided by the Montreal Neurological Institute. Transformation between the Montreal Neurological Institute’s coordinate system and that of Talairach and Tournoux (1988) was performed using the MNI2TAL formula (www.mrc-cbu.cam.ac.uk/imaging). The results of the GFP and topographic pattern analyses were used for defining time periods of AS multisensory neural response interactions with stable scalp topographies for which intracranial sources were estimated. That is, the GFP analysis was used to define when AS multisensory neural response interactions occurred, and the topographic analyses was used to determine whether and when stable scalp topographies were present in the ERP from each condition. It is important to note that these estimations provided visualization of the likely underlying sources and do not themselves represent a statistical analysis.

Results

Behavioral Results

Subjects readily detected stimuli of each modality. On average, subjects detected 98.7 ± 2.0% of auditory stimuli, 97.0 ± 3.4% of somatosensory stimuli and 99.0 ± 1.3% of multisensory stimulus pairs. For both spatially aligned and misaligned configurations, mean reaction times were faster for AS multisensory stimulus pairs than for the corresponding unisensory stimuli (Fig. 2a). This facilitation of reaction times is indicative of a redundant signals effect for multisensory stimuli (Miller, 1982; Schröger and Widmann, 1998; Molholm et al., 2002) and was assessed via two separate ANOVAs. The first tested for a redundant signals effect with spatially aligned stimulus pairs. The within-subjects factors were stimulus type (auditory-alone, somatosensory-alone, AS multisensory pair) and side of space (left, right). There was a main effect of stimulus type [F(2,22) = 36.22; P < 0.0001]. The interaction between factors of type and side of space was also significant [F(2,22) = 4.69; P < 0.05]. This followed from a difference in the relative advantage of multisensory stimuli versus each unisensory stimulus, though the overall pattern and facilitative effect was present for both sides of space (see Fig. 2). The second ANOVA tested for a redundant signals effect with spatially misaligned stimulus pairs. The within-subjects factors were stimulus type (auditory-alone, somatosensory-alone, AS multisensory pair) and side of somatosensory stimulation. There were main effects of both stimulus type [F(2,22) = 37.32; P < 0.0001] and side of somatosensory stimulation [F(1,11) = 4.90; P < 0.05]. However, the interaction between factors of type and side was not significant (P = 0.46). Follow-up planned comparisons (paired t-tests) confirmed that for both ‘aligned’ and ‘misaligned’ stimulus pairs, reaction times to AS multisensory stimulus pairs were significantly faster than to either single sensory modality (Table 1). This constitutes demonstration of an RSE with AS multisensory stimulus pairs. A third ANOVA was conducted to determine if mean reaction times were faster for spatially aligned versus misaligned AS stimulus pairs. The within-subjects factors were type (aligned versus misaligned) and side of somatosensory stimulation (left versus right). Neither factor nor the interactions between the factors yielded a significant difference, indicating similar reaction times for all multisensory stimulus pairs.

Two broad classes of models could explain instances of the redundant signals effect: race models and coactivation models. In race models (Raab, 1962), neural interactions are not
required to obtain the redundant signals effect. Rather, stimuli independently compete for response initiation and the faster of the two mediates behavior for any trial. Thus, simple probability summation could produce the effect, since the likelihood of either of two stimuli yielding a fast reaction time is higher than that from one stimulus alone. In coactivation models (Miller, 1982), neural responses from stimulus pairs interact and are pooled prior to behavioral response initiation, the threshold for which is met faster by stimulus pairs than single stimuli. We tested whether the RSE exceeded the statistical facilitation predicted by probability summation using Miller’s inequality (Miller, 1982). Detailed descriptions of this analysis are described in previous reports from our laboratory (Murray et al., 2001). In all cases, we observed violation of the race model (i.e. values greater than zero) over the fastest quartile of the reaction time distribution, supporting coactivation accounts of the present redundant signals effect (Fig. 2b).

**Electrophysiological Results**

As in our previous study (Foxe et al., 2000), visual inspection of the group-averaged ERP waveforms revealed a difference between responses to multisensory stimulus pairs and the summed responses from the constituent unisensory stimuli starting at ~50 ms at lateral central scalp sites over the hemisphere contralateral to the stimulated hand (Fig. 3a, arrows). This suggestion of nonlinear neural response interaction for AS multisensory stimuli was observed here for all spatial combinations and manifested as a supra-additive response. These interaction effects observed at a local scale (i.e. at specific electrodes) were also evident at a global scale. That is, these effects were likewise evident in the corresponding GFP (see Materials and Methods) waveforms from all conditions (Fig. 3a).

Moreover, these latter measures have the advantage of being a single, global index of the electric field at the scalp (i.e. GFP is not biased by the experimenter’s selection of specific electrodes for analysis). For both the aligned and misaligned configurations of AS multisensory stimuli, responses to stimulus pairs again appeared to be of larger amplitude than the summed responses from the corresponding unisensory stimuli over the 50–150 ms period and of smaller amplitude over the 200–300 ms period. These modulations were statistically tested in the following manner.

For each ‘pair’ and ‘sum’ condition as well as each subject, the GFP responses over both the 50–150 ms and 200–300 ms periods were submitted to separate three-way ANOVAs, using within-subjects factors of pair versus sum, aligned versus misaligned, and left versus right somatosensory stimulation. For the 50–150 ms period, only the main effect of pair versus sum was significant \[F(1,7) = 26.76, P < 0.003\]. Follow-up comparisons (paired \(t\)-tests) confirmed that for each stimulus configuration there was a larger GFP over the 50–150 ms period in response to multisensory stimulus pairs than to the summed responses from the constituent unisensory stimuli [aligned left \(t(7) = 4.92, P < 0.002\); aligned right \(t(7) = 4.41, P < 0.003\); misaligned left \(t(7) = 3.61, P < 0.009\); misaligned right \(t(7) = 2.85, P < 0.025\); see Fig. 3b, left]. All other main effects and interactions failed to meet the 0.05 significance criterion. In other words, both spatially aligned and misaligned multisensory stimulus conditions yielded brain responses over the 50–150 ms period that were stronger than the sum of those from the constituent unisensory stimuli and that did not statistically differ from each other. For the 200–300 ms period, there was a significant main effect of pair versus sum \([F(1,7) = 14.09, P < 0.007]\). In addition, there was a significant interaction between aligned versus misaligned configurations and the side of somatosensory stimulation \([F(1,7) = 9.733, P = 0.017]\), owing to the generally larger GFP for the ‘misaligned right’ condition. Nonetheless, follow-up comparisons (paired \(t\)-tests) confirmed that for each stimulus configuration there was a smaller GFP over the 200–300 ms period in response to multisensory stimulus pairs than to the summed responses from the constituent unisensory stimuli [aligned left \(t(7) = 3.27, P < 0.014\); aligned right \(t(7) = 5.16, P < 0.001\); misaligned left \(t(7) = 2.47, P < 0.05\); misaligned right \(t(7) = 3.28, P < 0.013\); see Fig. 3b, right].

However, given that the average speed of reaction times was ~530 ms (see Table 1), it is likely that this GFP difference follows from the contemporaneous summation of two motor responses in calculating the ‘sum’ ERP (i.e. one from each unisensory ERP). As such, the ‘pair’ versus ‘sum’ comparison is intermixed with the comparison of a single versus a double motor response. Such notwithstanding, this analysis thus indicates the presence of supra-additive nonlinear neural response interactions between auditory and somatosensory modalities that onset at ~50 ms post-stimulus irrespective of whether the stimuli were presented to the same spatial location.

In order to determine whether or not these AS neural response interactions were explained by alterations in the underlying generator configuration, we submitted the data to a topographic pattern analysis. This procedure revealed that six different scalp topographies optimally described the cumulative 300 ms post-stimulus periods across all eight conditions. Moreover, in the group-averaged data, we found that some of these scalp topographies were observed in some conditions, but not others over roughly the same time periods as when GFP modulations were identified (see shaded bars in Fig. 4, top). Specifically, over the 54–94 ms period, the scalp topography appeared to vary across conditions according to the side of somatosensory stimulation (the red-framed scalp topography for responses to left-sided somatosensory stimuli versus the
gold-framed scalp topography for responses to right-sided somatosensory stimuli, but did not vary between AS multisensory stimulus pairs and the summed responses from the constituent unisensory stimuli. In contrast, over the ~200–300 ms period the scalp topography appeared to vary between AS multisensory stimulus pairs (green-framed scalp topography) and the summed responses from the constituent unisensory stimuli (blue-framed scalp topography), irrespective of the spatial alignment of the auditory and somatosensory stimuli. We would note that this observation, which is statistically verified below, highlights the sensitivity of these analyses to effects following from the spatial attributes of the stimuli. More importantly, these findings are in solid agreement with the pattern observed in the selected ERP as well as GFP waveforms shown in Figure 3.

The appearance of these topographies was statistically verified in the ERPs of the individual subjects using a strength-independent spatial correlation fitting procedure, wherein each time point of each individual subject’s ERP from each condition was labeled with the map with which it best correlated. From this fitting procedure we determined both when and also the total amount of time a given topography was observed in a given
condition across subjects (Fig. 4, bar graphs). These latter values were subjected to a repeated measures ANOVA, using within subjects factors of pair versus sum, aligned versus misaligned, left versus right somatosensory stimulation, and topographic map. Over the 54–94 ms period, there was a significant interaction between pair versus sum and map \[ F(1,7) = 31.5, P < 0.009 \], indicating that one map was more often observed in the ERPs to multisensory stimulus pairs, whereas another map was more often observed in the ERPs to summed unisensory stimuli, irrespective of the spatial alignment of the auditory and somatosensory stimuli. No other main effect or interaction reached the 0.05 significance criterion for any of these tested time periods. We would note here that we are hesitant to overly interpret GFP or topographic modulation during this later (200–300 ms) period, given that mean reaction times to AS multisensory pairs occurred at ~350–375 ms. That is, this period likely includes brain activity reflecting motor preparation, rather than multisensory interactions per se, and is further exacerbated by the fact that the ‘sum’ ERP includes two motor responses (i.e. that from each of the corresponding auditory alone and somatosensory alone conditions).

Such notwithstanding, both the analysis of the GFP as well as of the scalp topography indicate that AS multisensory neural response interactions are present over the 54–94 ms period and are explained by a single, stable scalp topography that varies according to the hand stimulated and not as a function of paired versus summed unisensory conditions. We therefore performed our source estimation over this 54–94 ms period. We first averaged the ERP for each subject and each of the eight experimental conditions over this time period, thereby generating a single data point for each subject and condition. We then calculated the difference between these single-point per subject ERPs between the ‘pair’ and ‘sum’ conditions for each spatial configuration. LAURA source estimations were then performed and subsequently averaged across subjects. Figure 5 displays these averages (shown on the MNI template brain), which reflect the group-averaged source estimation of the ‘pair’ minus ‘sum’ ERP difference averaged for each subject over the 54–94 ms period. In each case, the LAURA source estimation revealed a robust activation pattern within the general region of posterior auditory cortex and the posterior superior temporal gyrus of the hemisphere contralateral to the hand stimulated. The coordinates (Talairach and Tournoux, 1988) of the source estimation current density maximum for conditions involving somatosensory stimulation of the left hand, both when spatially aligned and misaligned with the auditory stimulus, were 53, –29, 17 mm. Those for conditions involving somatosensory stimulation of the right hand, both when spatially aligned and misaligned with the auditory stimulus, were –54, –29, 17 mm. These coordinates are in close agreement both with the locus of AS interactions identified in our previous fMRI results (Foxe et al., 2002), as well as with the location of area LA as defined on anatomical criteria in humans (Rivier and Clarke, 1997).

**Discussion**

The present study investigated whether AS multisensory interactions for spatially aligned and misaligned stimulus configurations share a common spatiotemporal neural mechanism. Both the behavioral and electrophysiological data provide evidence that such facilitative interactions occur not only when

![Figure 4](image-url)
Rapid Auditory-Somatosensory Multisensory Interactions

Auditory and somatosensory stimuli are presented to the same location, but also when they are presented unambiguously to disparate locations. Behavioral as well as electrophysiological indices of AS interactions were equivalent for spatially aligned and misaligned combinations. Specifically, for all spatial combinations, reaction times were facilitated for the detection of AS multisensory pairs to a degree that exceeded simple probability summation (i.e. the race model). Likewise, electrophysiological indices of nonlinear neural response interactions between auditory and somatosensory stimuli were evident at just 50 ms post-stimulus in all cases and were manifest as a strength modulation (i.e. response enhancement) of statistically indistinguishable generators. Source estimations of these interactions were localized to auditory association cortices in the hemisphere contralateral to the hand being stimulated, regardless of the location of the auditory stimulus. In what follows, we discuss the implications of these findings on our understanding of multisensory interactions and spatial representations.

Behavioral Equivalence of AS Interactions

All spatial combinations yielded reaction time facilitation for multisensory stimulus pairs versus corresponding unisensory stimuli that exceeded simple probability summation. This facilitation is referred to as the redundant signals effect and has been observed both with unisensory (e.g. Murray et al., 2001) as well as multisensory stimulus pairs (e.g. Molholm et al., 2002). To our knowledge, this is the first demonstration of this phenomenon with AS stimuli, though reaction time studies using combined auditory and somatosensory stimuli do exist (Todd, 1912; Sarno et al., 2003). In addition, that the redundant signals effect in all cases exceeded probability summation argues against an attention-based explanation of the present results. According to such an account, subjects would have selectively attended to either the auditory or somatosensory modality, and the selectively attended modality would be stimulated on 50% of unisensory trials and 100% of multisensory trials. By this account, probability summation would suffice to account for any behavioral facilitation. Rather, this was not the case, and each condition exceeded probability summation over the fastest quartile of the reaction time distribution. The implication is that, in all cases, neural responses to the auditory and somatosensory stimuli interact in a (behaviorally) facilitative manner. In addition, reaction times were also statistically indistinguishable between stimulus pairs from the aligned and misaligned configurations. This pattern raises the possibility of the equivalence of AS interaction effects across all spatial configurations.

Equivalent Spatiotemporal Mechanisms of AS Interactions

The electrophysiological data provide strong support for this possibility. Robust nonlinear AS interactions were revealed for both aligned and misaligned combinations over the 50–95 ms post-stimulus. That is, responses to the multisensory ‘whole’ were greater than the summed responses from the unisensory ‘parts’. This replicates and extends our previous EEG evidence where similarly early interactions were observed following passive presentation of monaural (left ear) sounds and left median nerve electrical stimulation (Foxe et al., 2000). Furthermore, the present analysis approach provided a statistical means of identifying field strength as well as topographic (i.e. generator) modulations (see Murray et al., 2004, for details). Our analyses indicate that this early interaction is attributable to amplitude modulation of statistically indistinguishable scalp topographies. That is, these early AS interactions do not lead to activity in a new area or network of areas, but rather modulate the responses within already active generators. Modulation of the scalp topography predominating over this period was instead observed as a function of the hand stimulated, regardless of whether the sound was presented to the same or opposite side of space. Application of the LAURA distributed linear inverse solution to this period for all spatial configurations yielded sources in and around posterior auditory cortices and the superior temporal gyrus in the hemisphere contralateral to the hand stimulated.

This localization is in close agreement with our previous fMRI investigation of AS interactions (Foxe et al., 2002), where nonlinear blood oxygen level dependent (BOLD) responses were observed in posterior superior temporal regions. We interpreted this region as the homologue of macaque area CM, a belt area of auditory cortex situated caudo-medial to primary auditory cortex, which our previous electrophysiological studies in awake macaques have shown to exhibiting timing and laminar profiles of activity consistent with feedforward mechanisms of AS interactions (Schroeder et al., 2001, 2003;
Either to the same position or loci separated by indistinguishable mechanisms when stimuli are presented localized sources and occur at identical latencies and via multisensory interactions are not restricted to spatially co-
other sensory modalities. Of sensory processing already have access to information from other sensory modalities.

Most critically, the present data indicate that facilitative AS multisensory interactions are not restricted to spatially co-
spatial alignment. In other words, no behavioral task was required when another stimulus (typically of the same sensory

Possibility of Dynamic Spatial Representations
Another important consideration is the possibility that these early effects are mediated by higher-order cognitive processes — i.e. a ‘top-down’ influence. That is, in performing the task, subjects could adopt an overarching strategy that the presence of auditory, somatosensory, or AS multisensory stimuli will be used to determine when to respond, irrespective of whether or not the two stimuli are spatially aligned. In other words, temporal coincidence of the auditory and somatosensory stimuli suffices to facilitate reaction times (Stein and Meredith, 1993). Thus, one could envisage a situation where top-down influences reconfigure early sensory mechanisms to emphasize temporal properties and minimize spatial tuning. Such a mechanism would likely involve dynamically modifying spatial receptive fields, a notion that derives some support from studies within the visual system (e.g. Worgotter and Eysel, 2000; Tolias et al., 2001). It is also worth remembering that the relationship of auditory and somatosensory receptive fields are always in flux — i.e. we constantly move our limbs and body relative to our head (and ears). For instance, one’s left hand can operate entirely in right space to produce sounds there. Further examples of this include using tools and playing instruments where we generate somatosensory and auditory sensations that are spatially disparate (Obayashi et al., 2001; Maravita et al., 2003; see also Macaulay et al., 2002, for a similar discussion on visual-somatosensory interactions). Indeed, evidence for dynamic shifts in auditory receptive fields have been documented in the primate superior colliculus (e.g. Jay and Sparks, 1987) and parietal cortex (see Andersen and Buneo, 2002). The ecological importance of AS interactions can also be considered in the context of mechanisms for detecting and responding to hazardous events in a preemptive manner (e.g. Romanski et al., 1993; Cooke et al., 2003; Farnè et al., 2003). For example, vibrations and sounds can provide redundant information regarding the imminence of unseen dangers.

The spatial aspect of our results can also be interpreted in relation to clinical observations of multisensory interactions. In their studies of patients with somatosensory extinction, Farnè, Lādavas and colleagues (Farnè and Lādavas, 2002; Farnè et al., 2003) demonstrate that a touch on the contralesional side of space can be extinguished by an auditory or visual stimulus on the opposite (ipsilesional) side of space. Extinction refers to the failure to report stimuli presented to the contralesional side/ space when another stimulus (typically of the same sensory modality) is simultaneously presented ipsilesionally. The prevailing explanation of multisensory extinction is that this auditory or visual stimulus activates a somatosensory representation on the ipsilesional side whose activity in turn competes with the actual somatosensory stimulus on the contralesional side. An alternative that would be supported by the present data is that there is a direct interaction between AS stimuli presented to different sides of space.

Resolving Discrepancies with Prior MEG Studies
Lastly, contrasts between the present results (as well as our previous findings) and those of the MEG laboratory of Ritta Hari and colleagues (Lütkenhöner et al., 2002; Gobbelé et al., 2003) are also worth some discussion. Specifically, these MEG studies reported AS interactions at slightly longer latencies than those from our group and attributed them to somatosensory area SII, rather than auditory association cortices. In addition to the variance that might be attributed to the relative sensitivity of EEG and MEG to radial and tangential generator configurations, several paradigmatic and analytical differences are also worth noting. For one, no behavioral task was required of subjects in either MEG study, though we hasten to note that this was also the case our earlier work (Foxe et al., 2000), thereby providing no means of assessing whether the stimuli were perceived as separate or conjoint. Likewise, the MEG study by Gobbelé et al. (2003) used separate blocks to present unisensory and multisensory trials while varying spatial location within blocks, albeit always with spatially aligned stimuli. Most critically from a paradigmatic perspective, both of these MEG studies used a fixed rate of stimulus presentation. As has

Schroeder and Foxe, 2002; Fu et al., 2003). Other areas in the vicinity of CM have likewise been implicated in AS interactions, including the temporo-parietal (Tpt) parabelt region (Leinonen et al., 1980), situated posteriorly to CM. In addition to providing further evidence for the localization of homologous areas in humans, the timing of the present AS interactions support a model of multisensory interactions wherein even initial stages of sensory processing already have access to information from other sensory modalities.
been previously discussed in the case of auditory-visual multisensory interactions (Molholm et al., 2002; Teder-Salejarvi et al., 2002), a fixed or predictable timing between successive stimuli can result in anticipatory slow-wave potentials that will effectively ‘contaminate’ the pair versus sum comparison. Such was not the case in the present study where a randomized interstimulus interval ranging from 1.5–4 s was used. Likewise, the application of a source estimation approach using a limited, predetermined number of dipolar sources that were moreover fixed in their positions further obfuscates a direct comparison between the results of our laboratories. Such being said, we would note that somatosensory responses in the human temporal lobe have indeed been observed with MEG at latencies of ~70 ms post-stimulus onset (Tesche, 2000). Experiments directly combining EEG and MEG while parametrically varying task demands will likely be required to fully resolve these discrepancies.

Summary and Conclusions

In summary, electrical neuroimaging and psychophysical measures were collected as subjects performed a simple reaction time task in response to somatosensory stimulation of the thumb and index finger of either hand and/or noise bursts from speakers placed next to each hand. There were eight stimulus conditions, varying unisensory and multisensory combinations as well as spatial configurations (Fig. 1). In all spatial combinations, AS multisensory stimulus pairs yielded significant re-action time facilitation relative to their unisensory counterparts that exceeded probability summation (Fig. 2), thereby providing one indication of similar interaction phenomena at least at a perceptual level. Moreover, equivalent electrophysiological AS interactions were observed at ~50 ms post-stimulus onset with both spatially aligned and misaligned stimuli. Interaction effects were assessed by comparing the responses to combined stimulation with the algebraic sum of responses to the constituent auditory and somatosensory stimuli. These would be equivalent if neural responses to the unimodal stimuli were independent, whereas divergence indicates neural response interactions (Figs 3 and 4). Lastly, LAURA distributed linear source estimations (Grave de Peralta et al., 2001, 2004) of these early AS interactions yielded sources in auditory regions of the posterior superior temporal plane in the hemisphere contralateral to the hand of somatosensory stimulation (Fig. 5). Collectively, these results demonstrate the equivalence of mechanisms for early AS multisensory interactions in humans across space and suggest that perceptual–cognitive phenomena such as capture and ventriloquism manifest at later time periods/stages of sensory-cognitive processing.

Notes

We thank Deirdre Foxe and Beth Higgins for their ever-excellent technical assistance, Denis Brunet for developing the Cartool EEG analysis software used herein, and Rolando Grave de Peralta and Sara Andino for the LAURA inverse solution software. Stephanie Clarke, Charles Spence, as well as three anonymous reviewers, provided detailed and insightful comments on earlier versions of the manuscript for which we are indebted. This work was supported by grants from the NIH to JF (MH65350 and MH63434). We would also like to extend a hearty welcome to Oona Maedh Nic an tSionnagain.

Address correspondence to John J. Foxe, The Cognitive Neurophysiology Lab, Nathan S. Kline Institute, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA, email: foxe@nki.fmh.org or Micah Murray, Division Autonome de Neuropsychologie et Service Radiodiagnostic et Radiologie Interventionnelle, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland, email: micah.murray@hospvd.ch.

References


