Research report

Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study

Sophie Molholm\textsuperscript{a,b}, Walter Ritter\textsuperscript{a,b}, Micah M. Murray\textsuperscript{a}, Daniel C. Javitt\textsuperscript{a,e}, Charles E. Schroeder\textsuperscript{a,c}, John J. Foxe\textsuperscript{a,b,c,d,*}

\textsuperscript{a}Cognitive Neurophysiology Laboratory, Program in Cognitive Neuroscience and Schizophrenia, Nathan Kline Institute for Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA
\textsuperscript{b}Department of Psychology, The City College of the City University of New York, 1338th St., Convent Avenue, New York, NY 10031, USA
\textsuperscript{c}Department of Neuroscience, 1300 Morris Park Avenue, Albert Einstein College of Medicine, Bronx, NY 10461, USA
\textsuperscript{d}Department of Psychiatry and Behavioral Science, 1300 Morris Park Avenue, Albert Einstein College of Medicine, Bronx, NY 10461, USA
\textsuperscript{e}Department of Psychiatry, New York University Medical Center, 550 1st Avenue, New York, NY 10016, USA

Abstract

Integration of information from multiple senses is fundamental to perception and cognition, but when and where this is accomplished in the brain is not well understood. This study examined the timing and topography of cortical auditory–visual interactions using high-density event-related potentials (ERPs) during a simple reaction-time (RT) task. Visual and auditory stimuli were presented alone and simultaneously. ERPs elicited by the auditory and visual stimuli when presented alone were summed (`sum' ERP) and compared to the ERP elicited when they were presented simultaneously (`simultaneous' ERP). Divergence between the `simultaneous' and `sum' ERP indicated auditory–visual (AV) neural response interactions. There was a surprisingly early right parieto-occipital AV interaction, consistent with the finding of an earlier study [J. Cogn. Neurosci. 11 (1999) 473]. The timing of onset of this effect (46 ms) was essentially simultaneous with the onset of visual cortical processing, as indexed by the onset of the visual C1 component, which is thought to represent the earliest cortical visual evoked potential. The coincident timing of the early AV interaction and C1 strongly suggests that AV interactions can affect early visual sensory processing. Additional AV interactions were found within the time course of sensory processing (up to 200 ms post stimulus onset). In total, this system of AV effects over the scalp was suggestive of both activity unique to multisensory processing, and the modulation of `unisensory' activity. RTs to the stimuli when presented simultaneously were significantly faster than when they were presented alone. This RT facilitation could not be accounted for by probability summation, as evidenced by violation of the `race' model, providing compelling evidence that auditory–visual neural interactions give rise to this RT effect. © 2002 Elsevier Science B.V. All rights reserved.

Theme: Neural basis of behavior

Topic: Cognition

Keywords: Multisensory; Visual; Auditory; Electrophysiology; ERPs

1. Introduction

Everyday tasks involve the seemingly automatic integration of information from multiple sensory modalities. For instance, driving a car involves the synthesis of visual (seeing the road), auditory (hearing the car engine; a passing car), somatosensory (feeling the steering wheel), and motor (depressing the gas pedal) activity. The combination of inputs from different senses can function to reduce perceptual ambiguity (e.g. Ref. [32]) and enhance stimulus detection (e.g. Ref. [53]). Despite the fundamental role that sensory integration plays in performance and perception, how and when information from separate sensory modalities comes together in the human neocortex is not well understood. The bulk of our knowledge on the mechanisms of multisensory integration in the brain comes from the pioneering research of Stein, Meredith, and co-workers (see Ref. [54] for a review) on multisensory processing in the superior colliculus (of anaesthetized cats), a sub-cortical structure involved in orienting to auditory, visual, and somatosensory stimuli.
visual, and somatosensory stimuli. However, the extent to which the multisensory mechanisms defined in the superior colliculus generalize to cortical processes remains to be fully elucidated. Knowledge of the timing and anatomical distribution of cortical multisensory processing is essential to determining the roles that it plays in information processing.

Generally, it has been assumed that cortical multisensory processing occurs relatively late, following extensive processing of sensory inputs, and that it occurs in higher order cortical areas specialized for this purpose (e.g. Ref. [33]). This assumption can be partially attributed to: (1) a bias resulting from the tradition of studying sensory systems in isolation, and (2) animal studies that reveal multisensory convergence in higher-order regions of the parietal (e.g. Refs. [17,28,50]), temporal (e.g. Refs. [5,12,27]), and frontal lobes (e.g. Refs. [3,61]) along with a general lack of corresponding studies demonstrating convergence in lower-tier cortical areas. However, recent evidence suggests that multisensory processing occurs during initial sensory transmission, and in cortical areas that are usually held to be unisensory. An investigation by Schroeder et al. [47] in the caudomedial (CM) belt area of the auditory association cortex of awake behaving macaque monkeys, which gets direct input from primary auditory cortex (A1), showed auditory–somatosensory co-representation. Critically, both the auditory and somatosensory inputs to CM had characteristic feed-forward patterns, with both inputs arriving first in layer 4 at about 12 ms post stimulus onset, strongly suggesting bottom-up multisensory integration that occurs early in the sensory processing hierarchy. Functional imaging studies have suggested multisensory effects in what have been classically considered unisensory cortical areas [7,31], although the prevailing opinion is that these interactions represent feedback from higher-tier multisensory onto the lower-tier unisensory areas. Direct empirical evidence of feedback-mediated multisensory convergence in classical sensory cortex is sparse but supports this possibility (see Ref. [49]).

Very recently two event-related potential (ERP) studies found surprisingly early multisensory effects that, in light of their scalp topography, appear to indicate the early integration of sensory information in traditionally held unisensory cortex. In Giard and Peronnet [24], auditory–visual (AV) effects were found to onset at just 40 ms over right parieto-occipital scalp; this is consistent with generators in early visual cortices, although the spatial resolution of ERPs does not allow the contribution of the abutting multisensory areas in posterior parietal cortex or superior temporal sulcus (STS), to be ruled out. And in Foxe et al. [20], auditory–somatosensory effects onset at about 50 ms over central/post-central scalp, consistent with generators in early somatosensory cortex; and at 70 ms over scalp areas consistent with neural activity from posterior auditory areas, in line with the findings of Schroeder et al. [47].

The finding of an AV effect that onsets at 40 ms over parieto-occipital scalp [24] suggests that AV effects can occur at about the same time that initial activation of primary visual cortex (V1) is usually assumed to occur (45–55 ms as represented by the onset of the earliest cortical visual evoked potential, C1: e.g. Refs. [10,11,21]). This surprisingly early latency finding suggests a model of auditory–visual interaction in the cortex where auditory input, which reaches the cortex in less than half the time of visual input (9 to 15 ms: [9,59]), is transmitted from auditory cortices to visual or nearby visually dominant cortical areas, and consequently affects the early sensory processing of visual input.

The purpose of the present study was to advance the understanding of cortical multisensory processing by placing early AV interactions within the temporal and topographical framework of cortical sensory processing of the individual auditory and visual inputs. We first endeavored to determine if the early AV effect reported in Giard and Peronnet [24] would be elicited using a simple task and basic stimuli. In Giard and Peronnet a relatively complicated task and stimulus set were employed: on each trial, one of two tones was presented, and/or a permanently placed circle morphed into a horizontal or vertical ellipse. Subjects made forced two-choice classifications of the six randomly occurring stimulus conditions. In contrast, in the present study, single visually presented disks and auditory pure tones were presented either alone or simultaneously, and subjects performed a speeded simple reaction-time task. Elicitation of the effect under these conditions, in conjunction with its elicitation in the very different conditions of Giard and Peronnet [24], would suggest that early cortical multisensory processing for auditory and visual stimuli that onset simultaneously may be present for a variety of stimuli and tasks. We then compared the onset of the earliest AV effect to the onset of C1 in response to the visual stimulus alone. We expected that the initial cortical response to the visual stimulus would precede any AV effects, reflecting that cortical unisensory processing of the visual input began prior to cortical multisensory processing.

The technique of high-density electrical mapping (from 128 scalp electrodes) was used to establish the spatio-temporal dynamics of auditory–visual multisensory processing in relation to activation across a distributed sensory processing network. To assess multisensory processing, ERPs to the ‘visual alone’ and ‘auditory alone’ stimulus conditions were summed (hereafter referred to as the ‘sum’ ERP) and compared to the ERP to the simultaneously presented auditory and visual stimuli (the ‘simultaneous’ ERP). If neural responses to the auditory and visual inputs were processed in the same way when they were presented simultaneously as when they were presented alone, then, based on the principle of superposition of electrical fields, the ‘simultaneous’ ERP would be equivalent to the ‘sum’ ERP. However, if the neural responses to the simultaneous-
ly presented auditory and visual stimuli interacted during processing, the ‘simultaneous’ and ‘sum’ ERPs would diverge. This method of measuring multisensory processing is valid when neural responses reflect sensory processing unique to the stimulus, and do not reflect processes common to all three stimulus types such as target (e.g. the P3) or response (e.g. motor cortex activity) related neural activity. Several forms of interaction effects have been reported from this comparison (e.g. Refs. [20,24,38,40,55]). Although our primary focus was on the earliest AV interaction, AV interactions up to 200 ms were considered.

We also tested whether multisensory processing was reflected in our behavioral data. Simple reaction-times are generally facilitated when location concordant stimuli are simultaneously presented. This has been called the ‘redundant signal effect’ (RSE) (e.g. Refs. [29,37]). There are two classes of models to explain this effect: race models and coactivation models. In race models each stimulus of a pair independently competes for response initiation, and the faster of the two mediates the response for any trial. According to this model probability summation produces the RSE, since the likelihood of either of the two stimuli yielding a fast reaction-time is higher than that from one stimulus alone. In coactivation models, the interaction of neural responses to the simultaneously presented stimuli facilitates response initiation and produces the RSE. We tested whether the RSE exceeded the statistical facilitation predicted by the race model, and thereby provided evidence for the contribution of AV neural interactions to RT facilitation.

2. Materials and methods

2.1. Subjects

Twelve neurologically normal, paid volunteers participated (mean age 23.8±2.69 S.D.; five female; 11 right-handed), all reported normal hearing and normal or corrected-to-normal vision. The Institutional Review Board of the Nathan Kline Institute for Psychiatric Research approved the experimental procedures, and each subject provided written informed consent. Data from two additional subjects were excluded, one for excessive blinking, and the other for failure to maintain central fixation.

2.2. Stimuli

2.2.1. Auditory alone

A 1000 Hz tone (60 ms duration; 75 dB SPL, 5 ms rise/fall times) was presented from a single JBL speaker located atop the monitor on which the visual stimulus was presented.

2.2.2. Visual alone

A disk (60 ms duration), subtending 1.2° in diameter (143 cm viewing distance) and appearing red on a black background, was presented on a monitor (Iiyama Vision Master Pro 502, model #A102GT), peripheral to central fixation. The location of the circle was chosen on an individual subject basis after completion of the following procedure. In a pre-test session, the red circle was presented 100 times to each of 16 different locations across the visual field. The center of the stimulus was located 1.6 or 2.6 degrees lateral from fixation and 1.6 or 2.0 degrees vertical from fixation. For each subject, the location was chosen from which the best C1–P1 visual evoked potentials (VEPs) were observed. We sought to optimize our sensitivity to the obligatory VEPs because cortical AV effects have been previously reported in the latency range and general scalp location of C1 and P1 (40 to 90 ms post-stimulus; [24]). There were four different locations used across our cohort (see Fig. 1).

2.2.3. Auditory and visual simultaneous

The auditory-alone and visual-alone conditions described above were presented simultaneously. The auditory and visual stimuli were in close proximity, with the speaker placed on top of the monitor in vertical alignment with the visual stimulus.

2.3. Procedure

Participants were seated in a comfortable chair in a dimly lit and electrically shielded (Braden Shielding...
Systems) room and asked to keep head and eye movements to a minimum, while maintaining central fixation. Eye position was monitored with horizontal and vertical electro-oculogram (EOG) recordings. Subjects were instructed to make a button-press response with their right index finger when a stimulus in either sensory modality was detected, as quickly as possible without making errors. The three stimulus conditions were presented with equal probability in random order. Stimulus onset asynchrony (SOA) varied randomly between 750 and 3000 ms. Stimuli were blocked into sequences of 150 trials, and each subject completed a minimum of 18 blocks. Breaks were encouraged between blocks to maintain high concentration and prevent fatigue.

2.4. Data acquisition and analysis

Continuous EEG was acquired from 128 scalp electrodes (impedances <5 kΩ), referenced to the nose, band-pass filtered from 0.05 to 100 Hz, and digitized at 500 Hz. The continuous EEG was divided into epochs (−100 ms pre- to 800 ms post-stimulus onset) and baseline corrected over the full 900 ms. Trials with blinks and eye movements were rejected off-line on the basis of the EOG. An artifact criterion of ±60 μV was used at all other scalp sites to reject trials with excessive EMG or other noise transients. The average number of accepted sweeps per stimulus condition was 681 (ranging from 661 to 702 sweeps over the three conditions).

EEG epochs were sorted according to stimulus condition and averaged from each subject to compute the event-related potential (ERP). Baseline was defined as the epoch from −50 ms to 10 ms post-stimulus onset. In addition, ERPs from the auditory-alone and visual-alone conditions were summed for statistical comparison with the ERP response to the simultaneous auditory–visual condition. Group-averaged ERPs for each of the three stimulus conditions and the summed auditory and visual alone ERPs were calculated for display purposes and for identification of the auditory N1 and P2 and the visual C1, P1, and N1.

Button press responses to the three stimulus conditions were acquired during the recording of the EEG and processed offline. Responses falling between 100 and 800 ms post stimulus onset were considered valid. This window was used to avoid the double categorization of a response.

2.5. Statistical analyses

2.5.1. Behavioral

For individual subjects, the percent hits and average response-time were calculated and their reaction-time distributions were recorded for each stimulus condition. To test for the presence of an RSE, planned comparisons between each of the alone stimulus conditions and the simultaneous stimulus condition were performed. To determine if a co-activation explanation of the RSE was necessary, Miller’s test of the race model [37] was implemented. The model places an upper limit on the cumulative probability (CP) of RT at a given latency for stimulus pairs. For any latency, t, the race model holds when this CP value is less than or equal to the sum of the CP from each of the single stimuli minus an expression of their joint probability (CP(t)simultaneous−(CP(t)alone 1 + CP(t)alone 2)−(CP(t)alone 1 × CP(t)alone 2)). For each subject the RT range within the valid RTs (100–800 ms) was calculated over the three stimulus conditions and divided into quantiles from the first to the hundredth percentile in 5% increments (1%, 5%, ..., 95%, 100%). T-tests comparing the actual (CP(t)simultaneous) and predicted ((CP(t)alone 1 +CP(t)alone 2)−(CP(t)alone 1 × CP(t)alone 2)) facilitation were performed to assess the reliability of violations of the race model.

2.5.2. Event-related potentials: AV effects

To date only Giard and Peronnet [24] have measured AV multisensory processing by comparing ‘simultaneous’ and ‘sum’ ERPs. In the first phase of our analysis we used their findings to predefine the latencies and scalp regions at which we tested for AV interactions. As such, we tested for an early AV interaction over parieto-occipital scalp within the latency range of the C1, AV interactions within the latency range of the auditory N1 and P2 and the visual P1 and N1 components, over corresponding scalp areas, and an AV effect over right fronto-temporal scalp at about 150 ms. For each of the six tests the mean amplitude of the ‘simultaneous’ and ‘sum’ ERPs over a 10 ms window was submitted to a three-way repeated measures analysis of variance (ANOVA) with factors of stimulus type (‘simultaneous’ versus ‘sum’), hemisphere, and electrode (3 representing each hemisphere). The center of each of the latency windows was chosen on the basis of the maximum difference between the ‘simultaneous’ and ‘sum’ ERPs within the prescribed latency range, and is indicated in parentheses in the Results section. Geisser–Greenhouse corrections were used in reporting P values when appropriate. An alpha level of less than 0.05 was used for all statistical tests. For the reliable AV interactions, the topography and distribution of the effect was inspected in the difference voltage map; when the focus of the effect differed from the scalp region tested, a follow-up test was conducted.

The second phase of analysis was conducted to render a full description of the spatio–temporal properties of the AV interactions, and was considered exploratory. The ‘simultaneous’ and ‘sum’ ERPs were compared using point-wise running paired t-tests (two-tailed), with an AV interaction defined as at least 10 consecutive data points meeting a 0.05 alpha criterion (10 data points = 20 ms at a 500 Hz digitization rate) (for use of running t-tests see Refs. [26,40]). This criterion meets stringent standards for
assessing reliable effects, as posed by Guthrie and Buchwald [26], when a large number of \( t \)-tests are calculated across the electrode montage and epoch. This is a suitable alternative to Bonferroni correction for multiple comparisons, which would increase the likelihood of Type II errors through overcompensation for Type I errors. This ‘post-hoc’ phase of the analysis was critical because (1) it provided a description of the time course of the AV effects, and (2) the full description of the spatio–temporal properties of the AV interactions allows these data to be used as a hypothesis generation tool for future multisensory studies.

### 2.5.3. Event-related potentials: C1 onset

C1 onset was measured by comparing the amplitude of the ERP to the visual alone condition to baseline using point-wise running paired \( t \)-tests (two-tailed); onset was defined as the first point where a 0.05 alpha criterion was met and followed by at least 10 consecutive significant data points.

### 3. Results

#### 3.1. Behavioral

##### 3.1.1. Redundant signal effect

Mean reaction-times to the simultaneous condition (255 ms) were faster than mean reaction-times to either the visual or the auditory alone conditions (305 and 297, respectively). An RSE was confirmed with planned comparisons of each of the alone conditions to the simultaneous condition (for auditory alone condition compared to the simultaneous condition: \( t_{11} = 7.972, P < 0.001 \); for the visual alone condition compared to the simultaneous condition: \( t_{11} = 11.057, P < 0.001 \)).

##### 3.1.2. Test of the race model

There was violation of the race model for each of the 12 subjects. The CP at each quantile were group averaged separately for each stimulus condition to form an aggregate distribution that preserved the shape of the individuals’ data (also termed Vincent averaging; [60]). For the group averaged data, the race model was violated over the first six quantiles (see Fig. 2). These violations were reliable across subjects, with statistically significant differences between the actual (CP \(_{(\text{simultaneous})}\)) and predicted (\((\text{CP}_\text{alone}_1 + \text{CP}_\text{alone}_2 - \text{CP}_\text{alone}_1 \times \text{CP}_\text{alone}_2)\)) facilitation for three of these quantiles, and approaching significance for two (see Table 1). Violation of the race model strongly supports the contention that the interaction of

<table>
<thead>
<tr>
<th>Quantile</th>
<th>CP: simultaneous</th>
<th>CP: predicted</th>
<th>( t_{11} )</th>
<th>( P &lt; )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.001</td>
<td>0.000</td>
<td>1.49</td>
<td>0.08</td>
</tr>
<tr>
<td>2</td>
<td>0.002</td>
<td>0.001</td>
<td>1.23</td>
<td>0.12</td>
</tr>
<tr>
<td>3</td>
<td>0.017</td>
<td>0.007</td>
<td>2.76</td>
<td>0.009</td>
</tr>
<tr>
<td>4</td>
<td>0.118</td>
<td>0.066</td>
<td>3.23</td>
<td>0.004</td>
</tr>
<tr>
<td>5</td>
<td>0.328</td>
<td>0.253</td>
<td>2.61</td>
<td>0.012</td>
</tr>
</tbody>
</table>

\( t \)-tests comparing the CP for the ‘simultaneous’ stimulus condition to the CP predicted by the race model over the first six quantiles, in which the ‘simultaneous’ CP exceeded the predicted CP.

---

Fig. 2. (a) Mean \((N=12)\) reaction times (S.E.M. indicated) for simultaneous (AV: red bar), and auditory (A: blue bar) and visual (V: green bar) alone stimulus conditions. (b) Cumulative probability (CP) distributions for the simultaneous (red trace) and visual (green trace) and auditory (blue trace) alone stimulus conditions, and the CP predicted by the race model (black trace). (c) Miller inequality: values greater than zero signify violation of the race model (seen here over the first six quantiles).
auditory and visual stimulus information during neural processing contributed to the RSE.

3.2. Event-related potentials

In Fig. 3 the group averaged ERPs to the alone conditions (auditory and visual), and the simultaneous condition at selected electrode sites are displayed. The ERP elicited by the auditory alone condition was characterized by the typical auditory components Pa peaking at 40 ms, P1 peaking at 56 ms, N1 peaking at 105 ms, and P2 peaking at about 200 ms [43,57,58]. Pa and N1 appeared maximal over fronto-central scalp and inverted at the mastoids; P1 appeared maximal at fronto-central sites but did not invert at the mastoids; and P2 appeared maximal over centro-parietal scalp. The ERP elicited by the visual alone condition was characterized by the visual C1 from about 40 to 65 ms, and a typical P1 and N1 peaking at 110 and 160 ms, respectively [10,21]; P1 appeared maximal over occipital scalp and N1 over occipito-temporal scalp. The ERP elicited by the ‘simultaneous’ condition displayed classic auditory and visual componentry, in general as described for the alone conditions.

3.2.1. AV interactions

Differences between the group averaged ‘simultaneous’ and ‘sum’ ERPs could be readily observed. For the electrode sites over parieto-occipital to occipital scalp, and in the posterior temporal regions, the ‘simultaneous’ ERP was more positive in amplitude from about 45 to 80 ms compared to the ‘sum’ ERP, and had a smaller peak amplitude for the positive going wave at about 100 ms and the negative going wave at about 160 ms. At the central and fronto-central electrode sites the ‘simultaneous’ ERP was of greater amplitude than the ‘sum’ ERP for the negative going wave at about 120 ms, and for the positive going wave at about 180 ms.

3.2.2. Posterior AV interactions

The early right parieto-occipital difference between the ‘simultaneous’ and ‘sum’ ERPs was confirmed with a significant stimulus type by hemisphere interaction (centered at 58 ms: $F(1,11) = 5.73, P = 0.036$); the ‘simultaneous’ ERP was more positive than the ‘sum’ ERP (see Fig. 4a). A follow up focused contrast over the right electrode sites, with factors of stimulus type and electrode, showed a main effect of stimulus type ($F(1,11) = 8.83, P = 0.013$).

Fig. 3. The ‘simultaneous’ (red trace), and visual (green trace) and auditory (blue trace) alone ERPs at left (a) and right (b) fronto-central electrode sites and at left (c) and right (d) occipital electrode sites. The placement of the electrode sites are indicated in magenta in top and back views of the electrode montage.
Fig. 4. Auditory–visual interactions are illustrated in ‘simultaneous’ versus ‘sum’ difference voltage maps of the focus of the AV interactions, at their peak latencies. The color scale is to the left of Fig. 4a, and the voltage scale used for each of the maps is indicated below it, in gray. These are accompanied by ERP traces to the ‘simultaneous’ (red), ‘sum’ (blue), and ‘simultaneous’ versus ‘sum’ difference (green) from corresponding electrode sites. The electrode location is indicated with a black dot on the voltage map. A vertical black line is drawn through the traces at the latency of the corresponding voltage map. For each AV interaction voltage map, a unisensory voltage map (visual or auditory) at the same latency is illustrated, to the far right. Comparison of the AV interactions and corresponding unisensory voltage maps reveals the similarity or difference of the interaction to unisensory processing.

In the latency range and scalp region of the occipitally focused visual P1, the ‘simultaneous’ ERP was significantly less positive going than the ‘sum’ ERP (centered at 120 ms: $F(1,11)=5.89, P=0.034$). However, inspection of the voltage map suggested that the AV interaction in this latency range was largely focused over left centro-parietal scalp (see Fig. 4c). A post-hoc two-way repeated measures ANOVA on left centro-parietally placed electrodes supported this contention ($F(1,11)=15.69, P=0.002$) and we concluded that the AV effects in this latency range mainly reflected interactions over left centro-parietal scalp.

In the latency range and scalp area of the occipito-temporally focused visual N1 the ‘simultaneous’ ERP was less negative going than the ‘sum’ ERP (see Fig. 4f), with a two-way interaction of stimulus type by electrode that approached significance (centered at 165 ms: $F(1,11)=4.22, P=0.061$). The ‘simultaneous’ versus ‘sum’ difference was focused over the right scalp. Given the strong
Fig. 4. (continued)
trend toward significance we did a post hoc contrast on the right hemisphere electrode sites, with factors of stimulus type and electrode; this showed a main effect of stimulus type \( (F(1,11)=6.31, P=0.029) \).

### 3.2.3. Fronto-central/central AV interactions

In the latency range and scalp region of the fronto-centrally focused auditory N1, the amplitude of the ‘simultaneous’ ERP was significantly more negative going than the amplitude of the ‘sum’ ERP (centered at 120 ms: \( F(1,11)=10.20, P=0.009 \)). However, inspection of the topography of the effect suggested that it was largely due to the left centro-parietally focused AV interaction (see Fig. 4c, and the results from the post-hoc ANOVA on left centro-parietal sites, centered at 120 ms, reported above).

In the latency range and scalp area of the centrally focused auditory P2 the ‘simultaneous’ ERP was significantly more positive going than the ‘sum’ ERP (see Fig. 4g) (centered at 180 ms: \( F(1,11)=5.23, P=0.043 \)). Inspection of the simultaneous and sum ERPs across centro-temporal scalp suggested that there were no amplitude differences focused over this area (centered at 150 ms: \( F(1,11)=1.03, P=0.331 \)).

### 3.2.4. Exploratory analysis of the spatio–temporal properties of the AV interactions

The \( P \)-values from the running \( t \)-test analysis comparing the amplitudes of the ‘simultaneous’ and ‘sum’ ERPs across the electrode montage and epoch provide a comprehensive picture of the spatio–temporal properties of the AV interactions (Fig. 5). The earliest AV interaction onset at 46 ms, was evident only over the right parieto-occipital region of the scalp, and lasted for 20 ms (also see Fig. 4a). This was followed by an AV interaction, not identified in our first phase of analysis, focused over frontal scalp that onset at 86 ms and lasted for up to 26 ms (also see Fig. 4b). Next, there were simultaneous AV interactions focused over left centro-parietal and right posterior temporal regions of the scalp. Left centro-parietally it onset at 102 ms, and right posterior temporally it onset at 100 ms; these evolved into AV interactions focused over left centro-parietal scalp and right superior temporal scalp that peaked at 128 ms (also see Fig. 4c, d, and e). These effects did not appear to be a modulation of the auditory N1. Starting at 160 ms there was an AV interaction focused over the right occipito-temporal scalp that lasted until at least 180 ms, and this difference had a topography and time course.

![Fig. 5. Significant \( P \) values over time for 119 electrodes from running \( t \)-tests comparing the ‘simultaneous’ and ‘sum’ ERPs (this excluded four mastoid electrodes, four non-scalp electrodes that recorded ocular activity, and one bad electrode). \( P \) values are differentiated with a color scale. Grey indicates an absence of significant \( P \) values. Time is plotted on the x axis from 0 to 200 ms. Electrodes are plotted on the y axis. Starting from the bottom of the graph, the electrodes are divided into sections from posterior to anterior scalp (O, occipital; P-O, parieto-occipital; P, parietal; C, central; F-C, fronto-central; F, frontal). Within a section the electrodes are arranged from the right most lateral to the left most lateral sites (indicated to the right of the graph for each section). The sections are labeled based on the mid line of the scalp (e.g. ‘central’ includes electrodes over temporal scalp).](image-url)
consistent with the visual N1 (also see Fig. 4f). Finally, there was a central AV interaction focused over the left scalp that onset at 174 ms and extended out to 200 ms post stimulus onset (also see Fig. 4g). We should note that the two latter AV interactions were present at only 3 to 4 electrode sites, while the other AV interactions were present at many electrode sites clustered within a scalp region.

3.2.5. The present multisensory effects are not an artifact of slow wave anticipatory potentials

The performance of a task in our experimental paradigm introduced the possibility of slow wave anticipatory potentials (e.g. the contingent negative variation; see Ref. [62]). These slow potentials begin several hundred milliseconds prior to stimulus onset and continue for a short time thereafter. If present, anticipatory potentials would be represented twice in the ‘sum’ ERP and only once in the ‘simultaneous’ ERP, resulting in an artificial difference between the two. Problematically, this difference would appear to begin post-stimulus onset, due to baseline correction on the pre-stimulus portion of the ERPs, and would be indistinguishable from a genuine AV interaction [56].

The current experiment was designed to minimize the incidence of anticipatory potentials. SOA was varied randomly between 750 and 3000 ms such that subjects could not accurately predict the arrival of the next stimulus, and hence would be discouraged from anticipating its onset (see Ref. [13]). To confirm that anticipatory potentials did not contribute to the reported AV effects we compared the onset of our earliest effect for two different baseline criteria. If the effect was a consequence of anticipatory potentials, then its onset latency should shift for different baseline criteria [56]. The early right parieto-occipital effect onset at the same latency of 46 ms for the two baseline criteria: 100 to 50 ms pre-stimulus onset; and our original 50 ms pre-stimulus to 10 ms post-stimulus onset. This analysis strongly suggests that anticipatory potentials did not contribute significantly to our AV effects.

To further rule out the contribution of slow wave anticipatory potentials to our AV interactions, we inspected the grand mean ‘simultaneous’ and ‘sum’ ERPs for differences in the pre-stimulus portion of the waves. The data were reprocessed to look at a longer pre-stimulus epoch (250 ms pre-stimulus to 200 ms post-stimulus onset). No pre-stimulus differences were observed at anterior electrode sites, and only minimal pre-stimulus differences were observed at posterior electrode sites, again suggesting that anticipatory potentials did not contribute to our AV effects.

3.3. C1 onset

The early AV interaction fell within the time range of when the initial response in primary visual cortices, indexed by C1, generally occurs (e.g. Refs. [10,11,21]). We therefore set out to compare the onset of the early AV effect with the onset of C1. However, the C1 in the group averaged ERP to the visual alone stimulus condition was of relatively small amplitude, and activation did not reliably onset until the P1 portion of the ERP at 76 ms (at occipital sites). The relatively small amplitude C1 was likely due to our use of multiple visual stimulus locations across subjects. That is, because of the anatomy of the retinotopically arranged human visual cortex, the scalp topography of C1 changes as a function of stimulus position [10]. Thus, C1s to the same visual stimulus presented to different locations would have different scalp topographies, and when averaged together the C1 would appear diminished.

To get a more accurate measure of the onset of C1 to our visual stimulus, we presented it to a single location for an additional seven subjects (mean age 25±7.9 S.D.; five female; five right-handed; one also participated in the main experiment). Only differences in procedure between this and the main experiment are noted. The stimulus was always presented to the lower right quadrant of the visual field (2.6 degrees lateral and 2.6 degrees vertical from fixation). To maintain arousal a simple oddball paradigm was used in which subjects made a right-handed button-press response to an infrequently occurring change in color (5% of trials). SOA varied randomly between 750 and 1000 ms. Stimuli were blocked into sequences of 100 trials, and the first three subjects completed four blocks of trials while the remaining subjects each completed eight blocks of trials. The continuous EEG was divided into 100 ms pre-stimulus to 200 ms post-stimulus onset epochs. These were averaged for each subject to compute the ERP. Baseline was defined as the pre-stimulus epoch. The average number of accepted sweeps was 329±29 for the first three subjects and 721±21 for the remaining subjects.

The group averaged ERP exhibited a C1 that was largest over occipital scalp and appeared to diverge from baseline at ~40 ms. The t-test analysis showed that C1 onset at 44 ms over left occipital scalp and at 46 ms over central and right occipital scalp. Thus, with a temporal resolution of essentially 2 ms, the onset latencies of C1 and the early AV interaction were the same.

4. Discussion

The present study examined the spatial and temporal properties of cortical AV interactions to basic stimuli, while subjects performed a simple reaction-time task. In the behavioral data there was a significant reaction-time advantage when the visual and auditory stimuli were
presented simultaneously compared to when they were presented alone—the so-called redundant sensory effect (RSE). There was substantial violation of the race model. Hence the RSE could not be accounted for by simple probability summation, providing strong evidence that multisensory neural interactions are responsible for this sizeable RSE.

4.1. The early AV effect

The electrophysiological data exhibited surprisingly early auditory–visual multisensory interactions, over right parieto-occipital scalp. The combination of the early timing of this effect and its scalp topography over early visual areas is consistent with modification of very early visual sensory processing by auditory inputs. As unisensory transmission to auditory cortex is considerably faster than it is to visual cortex, this suggested the following possible interpretation to us. Firstly, auditory input activates primary auditory cortex within 15 ms of stimulus presentation and is then transmitted up the auditory processing stream. This input is then projected to ‘unisensory’ visual areas through one of two possible pathways.

The first possibility is a direct, feedforward projection from primary auditory or auditory association cortices to early visual cortices. Until recently, it was largely believed that such connectivity did not exist. However, very recently, anatomic tracer studies have provided evidence for direct projections from both primary auditory cortex (A1) and auditory association areas to both V1 and V2 in macaques [18,44]. The second possibility is a more indirect pathway, in which feedforward auditory input reaches areas of AV multisensory convergence (e.g. the superior temporal polysensory region—STS), and is transmitted via feedback connections to earlier ‘unisensory’ visual areas. The critical issue is one of timing. The question is whether there is sufficient time for auditory input to reach early visual areas to result in modulation of the later arriving visual input? Given the standard timing offset between the initial auditory and visual inputs to their respective primary cortices, there is a more than sufficient window of ~25–30 ms in which this process could occur (see e.g. Refs. [46,48,49]).

4.1.1. Timing and what it tells us

The onset of an AV interaction at 46 ms in our study is much earlier in information processing than multisensory processing is generally assumed to occur, and reinforces the view that visual processing is modified by auditory inputs well before visual sensory analysis is complete. The C1 component of the VEP, our measure of initial input to visual cortices, and the AV effect, were found to onset essentially simultaneously. This timing of AV interactions relative to C1 onset suggests that the visual input to this multisensory process arrives through feedforward connect-

ions. Further, this earliest AV interaction has a different scalp topography to that of the C1, with a focus that is more dorsal over the right parieto-occipital scalp. Thus, on the basis of both timing and topography, we suggest that an early stage of the dorsal visual stream is the likely site of this AV interaction. Visual activation of such areas is essentially simultaneous with activation of V1 [46]. The auditory input, on the other hand, conceivably arrived through either feedforward or feedback connections, as auditory cortices are activated by tones of moderate intensity by at most 15 ms [9].

This finding of an early AV effect contributes to recent evidence [20,24,47,49] of multisensory interactions during early sensory processing. Models of sensory integration need to account for such early multisensory operations in addition to later multisensory processing that occurs following the extensive sensory analysis of the individual sensory inputs [33,39]. For example, based on functional imaging data, Calvert et al. [7] and Driver and Spence [16] suggest feedback accounts of multisensory processing in unisensory cortex, although they do not specify the point at which this occurs in information processing.

4.1.2. Candidate areas for initial integration of multisensory inputs

The right parieto-occipital focus of the earliest AV interaction is consistent with sources originating from one of several possible underlying brain areas: regions of the visual motion processing system7 (referred to here as MT+, which includes V5 and related areas), and known multisensory regions of the posterior parietal cortex and the superior temporal sulcus (STS). Of these three, STS is the most parsimonious with the detailed neurophysiology of visual processing in the macaque monkey. In the monkey, STS receives visual inputs only milliseconds after V1 (~3 ms; [46,49], and there are cells in STS that receive converging input from auditory cortices [3,5,25]. Further, human fMRI studies implicate STS in auditory–visual multisensory processing (e.g. Refs. [8,41]). However, while its auditory–visual response characteristics make STS a strong candidate for the source of the early AV interaction, the topography of the interaction observed here and in Giard and Perronet [24] seems posterior to what one might predict for a generator in STS.

MT+ is consistent with the topography of the early interaction, and there is strong evidence from monkey intracranial data that the arrival of the fastest inputs to V5 occurs at the same time as, and sometimes even earlier than, initial inputs to V1 [45,46,48]; and human data have found similar timing [6,22]. Further supporting the possibility that MT+ contributed to the early AV interaction, multisensory processing in human MT+ for auditory–

7 Transient stationary stimuli, such as ours, activate the visual motion processing system (e.g. Schroeder et al. [46]).
visual speech has been shown with fMRI (Ref. [7]; also see Ref. [4]). Although monkey studies have failed to convincingly show that MT+ responds to auditory stimuli (e.g. unpublished data from our lab), it is possible that such multisensory cells in the macaque have not been sampled.

Also consistent with the topography of the early interaction is posterior parietal cortex, which is activated by a visual stimulus just a few milliseconds following V1 activation [46,49]. Human and monkey studies have shown that posterior parietal areas are involved in the multisensory representation of location and motion [2,30]. Further, monkey studies indicate that these areas include neurons that respond to both auditory and visual inputs early in information processing (e.g. Ref. [34]; and see Ref. [2] for review). A caveat is that such convergence of auditory and visual inputs has only been found when monkeys have been extensively trained to make saccades to the auditory stimuli [25].

4.1.3. Functional significance of early AV integration

MT+ and the posterior parietal cortices are involved in the processing of motion, and are implicated in exogenously (stimulus) driven attentional orienting to moving stimuli. Consistent with the right hemisphericity of the early AV interaction, the exogenous attentional orienting systems in posterior parietal cortices are generally localized to the right hemisphere (e.g. Refs. [36,42]). The rapid detection and localization of moving objects is advantageous to survival, for example allowing one to dodge a hurtling object whistling through the air, and would be enhanced by the early integration of temporally coincident auditory and visual information in motion sensitive areas. Behavioral studies have shown that detection and localization of visual objects are improved when there is a temporally coincident, location concordant, sound [52,53].

4.2. The later AV effects

A centro-parietal AV interaction peaked at 120 ms. This interaction overlapped with the time course of both the auditory N1 and the visual P1. However, its centro-parietal topography suggested a different configuration of generators. This evolved into more lateralized activity focused over right superior temporal and left centro-parietal scalp that peaked at about 128 ms; these AV interactions were distinct from the activity in response to the auditory or visual alone stimulus conditions within the same latency range (see Fig. 4d for comparisons of the interactions with the activity to the auditory alone condition at the same latency).

A later AV interaction, over left central scalp (~180 ms), was consistent with generators in motor cortex and may reflect sensory motor integration. The timing of the effect is consistent with such an interpretation, with the mean RT to the simultaneously presented stimuli occurring ~75 ms later, as is the scalp topography since the right handed response used in this study would be expected to elicit a left lateralized motor response.

In addition to the above, post hoc analyses suggested two further AV interactions. One was over frontal scalp (~100 ms), and was unique to the ‘simultaneous’ stimulus condition (i.e. did not correspond to activity in response to either the visual or auditory alone stimulus conditions). Multisensory processing in the frontal cortices has been shown previously in both human and monkey studies [3,8]. The other was focused over right occipito-temporal areas (~165 ms). The timing and scalp topography of this effect was consistent with those of the visual N1, elicited by the visual alone stimulus condition. Auditory-based modulation of the visual N1 would suggest that auditory inputs can affect the visual object recognition system, as the visual N1 has been implicated in visual object recognition processes by many studies (e.g. Refs. [1,14,15]).

5. Summary and conclusions

Our electrophysiological data are in general agreement with those reported by Giard and Peronnet [24], the only comparable ERP study to date. Both studies revealed AV interactions of the same polarity over right parieto-occipital scalp (~40–50 ms), over occipito-temporal scalp (~165 ms), and over fronto-central (in Giard and Peronnet [24]) and central (the present study) scalp (~180 ms). However, our AV effects were of considerably shorter duration than those reported in Giard and Peronnet. Additionally, while we also found multisensory effects in the latency range of the auditory N1 and visual P1 components, high-density topographic mapping showed that these effects had different topographies than the unisensory components, suggesting different underlying source configurations for these multisensory effects. Furthermore, our data did not evidence fronto-temporal activity at about 150 ms as was found in Giard and Peronnet. Of course, due to the considerable differences in the paradigms used, it is not surprising that some differences were seen. Rather, given the general similarities in findings across these two studies, we suggest that AV multisensory processing comprises a base set of cortical interactions that will be present for a variety of stimuli and tasks.

These data contribute to the growing literature on the spatio–temporal properties of cortical multisensory processing in humans. The extent to which multisensory activity is unique to particular combinations of the senses versus common to various sensory combinations, and the impact of task effects on the observed activities, remains to be seen. For example, it has been well-documented that there can be profound crossmodal influences on attentional orienting (e.g. Refs. [19,23,35,51]). Further, the current data highlight the importance of a detailed consideration of the time course of sensory processing within the individual sensory modalities. The inherent differences in transmis-
sion time to cortex of stimulation within the different sensory modalities will likely have considerable effects on the time course and potential neural areas at which multisensory interactions will occur.

Acknowledgements

Sincere appreciation to Ms. Beth Higgins and Ms. Deirdre Foxe for excellent technical assistance. Work supported in part by grants from the NIH—NS30029-23 (WR), MH63434 (JJF) and MH61989 (CES).

References