Dynamics of Cortico-subcortical Cross-modal Operations Involved in Audio-visual Object Detection in Humans

Alexandra Fort, Claude Delpuech, Jacques Pernier and Marie-Hélène Giard

INSERM U280, Mental Processes and Brain Activation, 151 Cours Albert Thomas, 69003 Lyon, France

Very recently, a number of neuroimaging studies in humans have begun to investigate the question of how the brain integrates information from different sensory modalities to form unified percepts. Already, intermodal neural processing appears to depend on the modalities of inputs or the nature (speech/non-speech) of information to be combined. Yet, the variety of paradigms, stimuli and technics used make it difficult to understand the relationships between the factors operating at the perceptual level and the underlying physiological processes. In a previous experiment, we used event-related potentials to describe the spatio-temporal organization of audio-visual interactions during a bimodal object recognition task. Here we examined the network of cross-modal interactions involved in simple detection of the *same* objects. The objects were defined either by unimodal auditory or visual features alone, or by the combination of the two features. As expected, subjects detected bimodal stimuli more rapidly than either unimodal stimuli. Combined analysis of potentials, scalp current densities and dipole modeling revealed several interaction patterns within the first 200 ms post-stimulus: in occipito-parietal visual areas (45–85 ms), in deep brain structures, possibly the superior colliculus (105–140 ms), and in right temporo-frontal regions (170–185 ms). These interactions differed from those found during object identification in sensory-specific areas and possibly in the superior colliculus, indicating that the neural operations governing multisensory integration depend crucially on the nature of the perceptual processes involved.

Introduction

A general observation of behavioral studies in humans is that we react more rapidly to external events characterized by features from different sensory modalities than to the same events presented in unimodal conditions alone (Miller, 1982, 1986; Hughes *et al.*, 1994). Several theoretical models have been proposed to explain this cross-modal facilitation effect. According to the *race model*, the shorter reaction times to bimodal stimuli are due to triggering the responses on the basis of the first detected cue (Raab, 1962). However, as the reaction times to bimodal targets are generally shorter than those predicted by this model, a convincing alternative proposed by Miller is that the processing of target information in one modality is affected by the presence of information of the other modality (Miller, 1991). The questions then arise: where, when and how does the parallel processing of unimodal cues interact in the processing chain?

Electrophysiological studies in animals have identified multisensory neurons in the deep layers of the superior colliculus (SC), a midbrain structure involved in orientation to external stimuli. These neurons display a much higher firing rate when two or more cues of different sensory modalities are presented in spatial and temporal coincidence [reviewed in (Stein and Meredith, 1993)]. Multisensory cells with similar properties have also been found in a number of cortical sites in cats (Toldi *et al.*, 1986; Stein *et al.*, 1993) and in monkeys (Benevento *et al.*, 1977; Ettlinger and Garcha, 1980; Bruce *et al.*, 1981; Hikosaka *et al.*, 1988). Interestingly, however, recent cat studies have shown that the multisensory neurons in the superior colliculus receive projections from *unimodal* neurons located in heteromodal regions of the cortex (anterior ectosylvian sulcus and lateral suprasylvian cortex) (Wallace *et al.*, 1993; Jiang *et al.*, 2001); this would suggest that the multisensory neurons in the superior colliculus and in cortex belong to separate integrative neural circuits possibly involved in different aspects of integration (Hughes *et al.*, 1994; Stein and Wallace, 1996).

In humans, recent neuroimaging studies using electromagnetic (ERPs, MEG) or hemodynamic (fMRI, PET) measures have begun to shed light on the neural operations mediating multisensory integration in various experimental conditions. These experiments have already shown that different cross-modal networks may be recruited according to the modality of the sensory inputs or the nature (speech/non-speech) of information to be bound. For example, a combination of two stimuli of different modalities presented simultaneously at the same location can modulate the brain responses to the corresponding separately presented unimodal stimuli in sensory-specific cortices (Sams *et al.*, 1991; Calvert *et al.*, 1999, 2000; Giard and Peronnet, 1999; Foxe *et al.*, 2000), in addition to having effects in a number of heteromodal structures such as the parietal cortex, the superior temporal sulcus, the right insula and/or the right prefrontal region (Calvert *et al.*, 2000; Downar *et al.*, 2000; Gonzalo *et al.*, 2000; Bremmer *et al.*, 2001; Bushara *et al.*, 2001; Sakowitz *et al.*, 2001). Integration of bimodal speech information appeared to activate mainly the left superior temporal sulcus (Calvert *et al.*, 2000) while audio-visual non-speech stimuli (passively perceived) were found to have strong effects in the superior colliculus (Calvert *et al.*, 2001). Furthermore, the brain areas involved in audio-visual integration of verbal material, whatever the form of visual inputs (lip movements or written letters), were shown to partly differ according to the congruence or non-congruence of unimodal components (Calvert *et al.*, 2000; Raij *et al.*, 2000).

Although these results indicate that the neural mechanisms of multisensory integration in perception are complex and depend on experimental context, the variety of paradigms, stimuli and technics used make it difficult to understand the relationships between the operative factors at perceptual and behavioral levels and the underlying neurophysiological processes. Preliminary evidence for the necessity to carefully control these factors comes from two previous event-related potential (ERP) experiments in our laboratory that have shown that cross-modal operations in perception not only depend on the nature of the sensory inputs, but, for a given bimodal stimulus, they vary with the perceptual processes involved and the sensory expertise of the individual for the task required. Indeed, using the same physical stimuli (deformations of a circle associated to a sound) in two object recognition tasks, we observed partly different cross-modal interaction patterns within the first 200 ms of stimulus analysis, according to whether the informative content of the unimodal cues was redundant (Giard and Peronnet, 1999) or non-redundant (Fort *et al.*, 2002) to identify the bimodal stimulus. The differences appeared mainly in the number, amplitudes and latencies of the interactions in sensory-specific cortices. Yet, in both studies, the effects were stronger in the cortex of the non-dominant modality of the subject to perform the task (auditory cortex for subjects having better performances for visual than for auditory object identification, and vice versa: see Discussion). In contrast, significant cross-modal interactions were found in the right temporo-frontal region in all the subjects in the two tasks.

Both these tasks required object identification. One may therefore expect that simple detection of the same objects — demanding more superficial stimulus analysis — will activate at least partly different cross-modal networks. This study aimed to test this hypothesis.

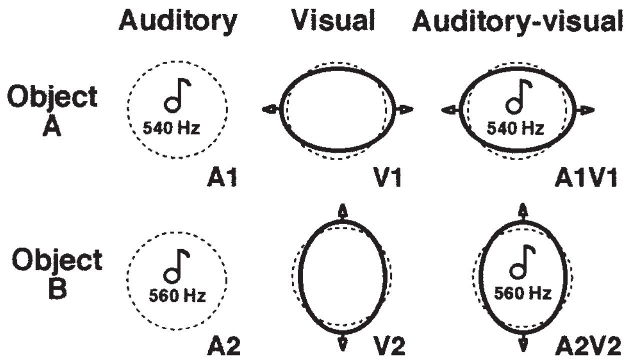

Figure 1. Two objects, A and B, were used. Each of them could be presented either in auditory condition alone (object A: 540 Hz tone burst, object B: 560 Hz), or in visual condition alone [transient deformation of a circle into an horizontal (object A) or vertical (object B) ellipse)], or in bimodal condition by the combination of their respective auditory and visual features.

Materials and Methods

Subjects

Fourteen right-handed subjects (mean age: 23.5; seven females) free of neurological illness and with normal hearing and normal or corrected-to-normal vision participated in this study. The protocol was approved by the regional Ethical Committee, and all subjects gave written informed consent of participation.

Stimuli

Two objects (A and B) were used. Object A consisted of a 540 Hz tone burst (stimulus A1) associated with the transient deformation of a circle into an horizontal ellipse (stimulus V1). Object B was designed similarly as a 560 Hz tone burst (A2) associated with a deformation of the circle into a vertical ellipse (V2). Each object could be presented either unimodally (stimulus A1, V1, A2, V2) or bimodally by the synchronous combination of their respective auditory and visual features (A1V1, A2V2) (Fig. 1). The basic circle had a diameter of 5 cm and was presented permanently in yellow on a dark monitor screen placed 1.3 m from the subject (visual angle: 2.2). The vertical and horizontal ellipses were formed by a ±10% deformation of the horizontal and vertical diameters of the circle. The duration of the deformation was 240 ms. The tone bursts had the same duration (including 10 ms of rise/fall times) and were delivered through a loudspeaker placed behind the video monitor with an intensity of ~50 dB HL.

Procedure

Subjects were seated in front of the video screen in a dark, sound-attenuating room. They were instructed to respond as quickly as possible upon the detection of any stimulus whatever its nature (auditory, visual or audio-visual), in pressing a key with the index finger of the right hand. Twelve blocks of stimuli were presented. A block started with the presentation of the circle on the video screen, the center of which served as a fixation point during the whole block. Each block included 72 trials composed of 12 repetitions of the six stimuli (A1, A2, V1, V2, A1V1 and A2V2) delivered randomly with an inter-stimulus interval (ISI) equal to the reaction time plus a random time varying between 1100 and 3000 ms, to avoid anticipatory effects. At the end of each block, the subjects were informed of their performances (mean reaction time, number of omissions and anticipations) in order to maintain attention during the whole period of recording (~45 min). Subjects could take breaks if necessary between blocks to minimize tiredness and eye movements.

EEG Recording

EEG was continuously recorded through DC coupled amplifiers (0.1–320 Hz analog bandwidth; sampling rate: 1 kHz) from 35 scalp electrodes referenced to the nose with placement based on the International 10–20 System: Fz, Cz, Pz, POz, Iz; Fp1, F7, F3, FT3, FC1, T3, C3, TP3, CP1, T5, P3, P13, O1, and their counterparts on the right hemiscalp; Ma1 and Ma2 (left and right mastoids, respectively); IMa and IMb (midway Iz-Ma1 and Iz-Ma2, respectively). Horizontal eye movements were recorded from the outer canthus of the right eye; eye blinks and vertical eye movements were measured in channels Fp1 and Fp2. Electrode impedances were kept below 5 kΩ. ERPs were computed separately for each stimulus type over a time period of 600 ms including 100 ms pre-stimulus, and digitally filtered (0–30 Hz). Trials with signal amplitudes exceeding 100 μV at any electrode were automatically rejected to discard the responses contaminated by eye movements or excessive muscular activities. Similarly, trials with reaction times below 200 ms or omissions were not taken into account in averaging. The mean numbers of averaged trials (by subject) were 230, 267 and 220 in auditory, visual and audio-visual conditions, respectively. The mean amplitude over the 100 ms pre-stimulus period was taken as the baseline for all amplitude measures.

Data Analysis

Estimation of Audio-visual Interactions

We assumed that, at an early stage of stimulus analysis, the responses to bimodal (AV) stimuli were composed of the sum of the responses evoked by the auditory (Au) and visual (Vi) stimuli presented separately, plus the putative neural activities related specifically to the bimodal nature of the stimulation (audio-visual interactions) (Barth *et al.*, 1995; Giard and Peronnet, 1999). This assumption is valid only while the period of analysis does not include non-specific activities that would be common to all three types (Au, Vi, AV) of stimuli, particularly late activities related to target processing (N2b, P3 waves), response selection or motor processes. These activities usually arise after 200 ms post-stimulus. We therefore restricted the analysis period to 0–200 ms and used the summative model to estimate the AV interactions:

$$\text{ERP (AV)} = \text{ERP (Au)} + \text{ERP (Vi)} + \text{ERP (Au} \times \text{Vi interactions)} \qquad (1)$$

This expression is valid whatever the nature, configuration or asynchrony of the neural generators and is based on the law of superposition of electric fields. AV interactions were therefore quantified as the differences between the responses to bimodal stimuli (AV) and the sum of the unimodal responses (Au + Vi). Significant effects were assessed by Student's *t*-tests comparing the amplitudes of the [AV – (Au + Vi)] difference waves to zero for each time sample at each electrode. Student's *t* maps could then be displayed at each latency. We considered as significant cross-modal interactions the spatio-temporal patterns having a stable topography with a significant amplitude ($P < 0.05$) at least one electrode during 15 consecutive samples (15 ms) (Rugg *et al.*, 1995; Thorpe *et al.*, 1996).

Topographic and Dipole Model Analysis

Scalp potential maps were generated using a two-dimensional spherical spline interpolation and a radial projection from Cz (top views), Oz (back

Table 1
Mean reaction times (in ms) ± standard errors to objects A and B presented in the auditory (Au) or visual (Vi) modality alone, or combining the two modalities (AV)

	Modality of presentation		
	Au	Vi	AV
Object A	275 ± 43	313 ± 26	248 ± 32
Object B	278 ± 43	307 ± 27	247 ± 34

views) or T3/T4 (lateral views), which respects the length of the meridian arcs. Scalp current densities (SCDs) were estimated by computing the second spatial derivative of the spline functions used in interpolation (Perrin *et al.*, 1987, 1989). SCDs do not depend on any assumption about the brain generators or the volume conductor. Compared to voltage maps, SCDs are reference-free and enhance the contribution of local intracranial sources. In addition, their amplitudes at the scalp decrease more rapidly with the depth of the intracerebral generators than those of the potentials: SCD maps therefore emphasize shallow, cortical activities and are blind to deeper sources (Perrin *et al.*, 1987; Pernier *et al.*, 1988).

Topographic analysis was complemented, for particular interaction effects, by spatio-temporal source modeling (Scherg and von Cramon, 1986; Scherg, 1990; Giard *et al.*, 1994). We used a classical three-concentric sphere head model for conductive volumes (brain, skull and scalp) and equivalent current dipoles (ECDs) for generators (local activity of brain regions). The procedure consists in identifying ECDs leading to the best fit between experimental and model distributions. The dipole parameters were determined by a non-linear iterative procedure for the spatial parameters (location and orientation), and with a linear least-mean square algorithm for the time-varying magnitude (Scherg, 1990). The model adequacy was assessed by a goodness-of-fit criterion based on the percentage of experimental variance explained by the model. The ECD solution was then projected onto 3-D magnetic resonance images (MRIs) of one subject (1.5 T Siemens system, 128 contiguous 1 mm thick horizontal sections parallel to the AC–PC line). The coordinate system used in the spherical head model was determined into the MRI: the sphere center was located near the crossing of T3–T4 and Fpz–Oz lines (Echallier *et al.*, 1992), and the radius was the distance between this center and Cz. The best fitting ECD was automatically displayed on the MRI whose section coordinates were the closest to the dipole coordinates.

Results

Behavioral Results

Two-way analysis of variance (ANOVA) on the reaction times with the type of object (A, B) and modality of presentation (Au, Vi, AV) as within-subject factors showed a significant effect of the stimulus modality [$F(2,78) = 23.47$, $P < 0.0001$)]. All the subjects had shorter reaction times to detect auditory (mean: 276 ms) than visual objects (mean: 310 ms; *post hoc* Fisher test: $P < 0.0022$), and shorter reaction times to detect bimodal (mean: 247 ms) than auditory objects ($P < 0.0004$). There was no effect of the type of object (Table 1).

Electrophysiological Results

Since the two objects A and B were physically similar and were detected with equivalent reaction times, they were grouped in ERP averaging according to their modality (Au, Vi, AV) to increase the signal-to-noise ratio of the responses.

Figure 2 presents the ERPs elicited by unimodal and bimodal stimuli from 100 ms before the stimulation to 200 ms after, at a subset of electrodes. The unimodal Au and Vi waveforms display morphologies typical of activities in sensory-specific areas. The auditory N1 is maximum ~100 ms at fronto-central site (–3.7 µV at Cz) with polarity reversal at mastoid electrodes (Ma1: 2.0 µV at 107 ms and Ma2: 1.7 µV at 106 ms), a pattern typical of neural activity in the supratemporal plane of the auditory cortex (Vaughan and Ritter, 1970). Auditory N1 was followed by the P2 wave peaking at Cz (5.0 µV) ~185 ms. In visual ERPs, the first salient deflection (N1) peaked at occipito-parietal electrodes (P24: –4.7 µV) ~170 ms (Fig. 2).

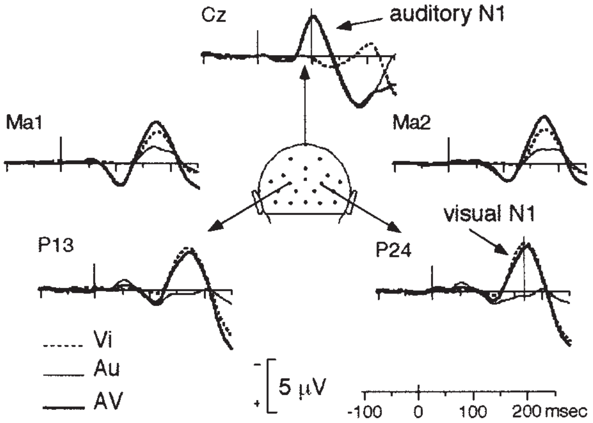

Figure 2. Grand-average ERPs across 14 subjects elicited by unimodal auditory (Au) and visual (Vi) stimuli, and by bimodal (AV) stimuli, at a subset of electrodes from 100ms before stimulus onset to 250ms after. Responses to objects A and B are grouped.

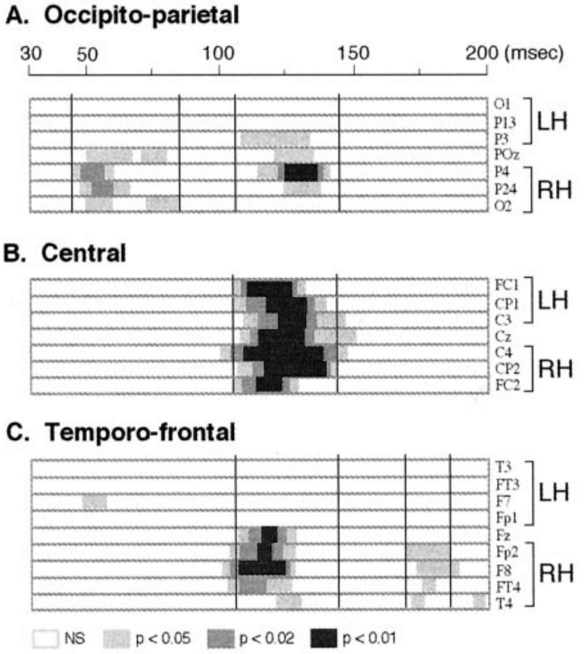

Figure 3. Statistical significance of the [AV – (Au+Vi)] difference waveform measuring the audio-visual interactions at a subset of (*A*) occipito-parietal, (*B*) central and (*C*) frontal and temporal electrodes on the right (RH) and left (LH) hemiscalps between 30 and 200 ms post-stimulus (Student's *t*-tests comparing the amplitude of [AV – (Au + Vi)] against zero at each latency). Three significant spatio-temporal patterns could be dissociated: at occipito-parietal sites from 45 to 85ms, at fronto-centro-parietal sites from 105 to 140 ms and at the right temporo-frontal electrode sites from 170 to 185 ms.

ERPs to bimodal stimuli had roughly the same morphology as the sum of ERPs for separately presented auditory and visual stimuli. Yet, using the criterion defined in the Materials and Methods section, several differences were found between the bimodal AV response and the sum of the unimodal (Au + Vi)

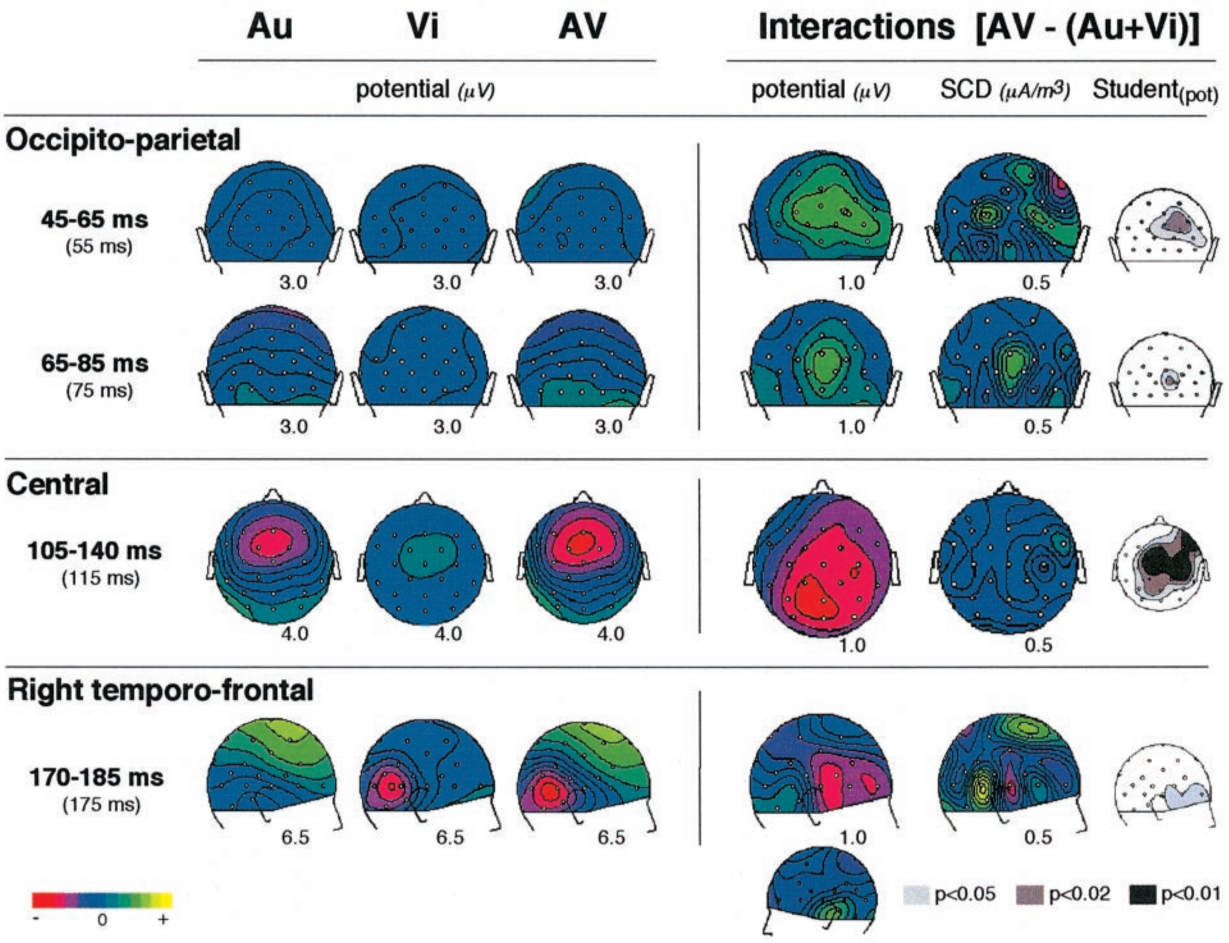

Figure 4. Topography of the different interaction patterns at occipito-parietal, central and right temporo-frontal scalp sites. Each line displays: the time window of the interaction pattern (and the illustrative latency at which the maps are depicted), the potential distributions of the unimodal (Au and Vi) and bimodal (AV) responses at this latency, the potential and SCD distributions of the cross-modal interactions quantified in the difference [AV – (Au+Vi)] between the bimodal responses and the sum of the unimodal responses, and the Student's *t* map estimated on the potential values. In Student maps, the gray colors indicate the scalp areas where [AV – (Au+Vi)] amplitudes differ significantly from zero with the probability coded in the gray level. In potential and SCD maps, half range of the scale (μV or μA/m^3) is given below each map. It may be seen that: (1) At occipito-parietal sites, [AV – (Au+Vi)] displays significant positive potential fields between 45 and 85 ms post-stimulus, typical of activities in the visual cortex. Over this period, the SCD maps present two different current patterns (45–65 and 65–85 ms) suggesting the existence of several interaction components. (2) Between 105 and 140 ms, the potential distribution of [AV – (Au+Vi)] extends to a wide central region while corresponding SCDs display only very weak currents, indicating the existence of cross-modal activities probably in deep brain structure(s) (superior colliculus?). (3) From ~170 to 185 ms, significant [AV – (Au+Vi)] amplitudes may be seen on the right temporo-frontal region (without counterpart on the left hemiscalp), suggesting cross-modal activities in the right prefrontal cortex and/or the right insula.

ERPs: from ~45 to 85 ms at occipito-parietal sites, from 105 to 140 ms over a wide central scalp region, and from 170 to 185 ms around the right temporo-frontal electrodes. Figure 3 details the statistical significance of the effects in these three spatio-temporal windows, and Figure 4 displays for each of them, at an illustrative latency: the potential maps of the unimodal (columns 1 and 2) and bimodal (column 3) responses, the potential and SCD distributions of the [AV – (Au + Vi)] interaction effects (columns 4 and 5, respectively), and the Student *t* map (computed on the potential values) showing the scalp areas with significant cross-modal effects (column 6).

Before 100 ms over Occipito-parietal Sites

Significant positive [AV – (Au + Vi)] amplitudes were found very early from 45 ms after stimulus onset around Pz and P4, extending to P24, POz, O2 and IMb electrode sites ~50 ms. The effect decreased in surface and remained stable around POz and O2 from 65 to 85 ms (Fig. 3*A* and Fig. 4: lines 1–2, columns 4 and 6). Over this period (45–85 ms), SCD distributions showed two distinct patterns including a pair of current sources around P13 and P24 from ~45 to 65 ms (Fig. 4: line 1, column 5) and a more centered neural source after 65 ms (Fig. 4: line 2, column 5), suggesting the existence of several components during this time range.

These interaction effects over the posterior visual regions did not correspond to clear pattern of activity in the unimodal responses in this time range. This suggests therefore activation of visually responsive neurons in posterior cortical areas little or not activated in unimodal conditions (see Discussion).

Between 100 and 150 ms over Central Scalp Sites

Within this time period, the topography of [AV – (Au + Vi)] was characterized by a wide negative potential field distributed over the frontal, central and parietal scalp regions between ~105 and 140 ms after stimulus onset (Fig. 3*B* and Fig. 4: line 3, column 4). This pattern did not appear in the corresponding unimodal responses (Fig. 4: line 3, columns 1–2). The effect began being significant at C4 electrode site (100 ms) and extended rapidly to

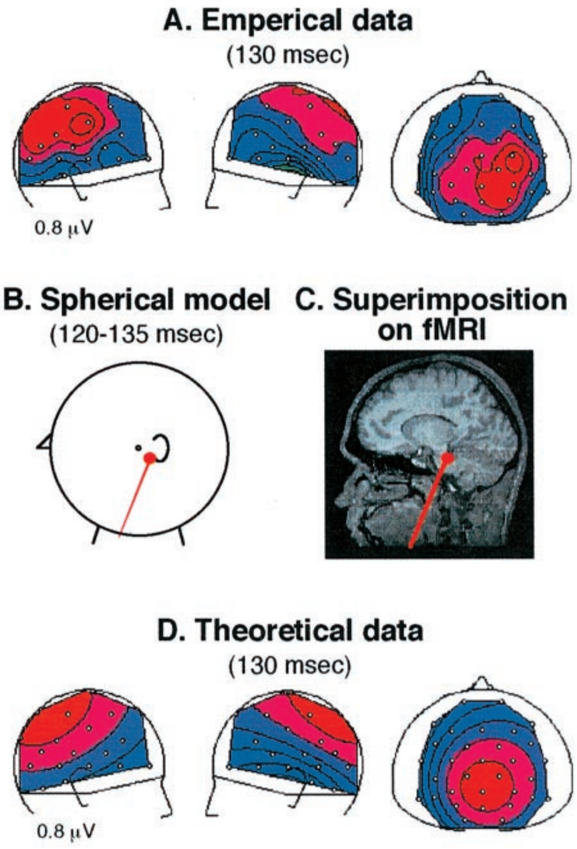

Figure 5. Results of spatio-temporal dipole modeling of [AV – (Au+Vi)] over the 120–135 ms period using the three-concentric sphere head model. (*A*) Experimental potential distributions of the mean amplitude of [AV – (Au+Vi)] between 120 and 135 ms after stimulus onset, on right, left and large-top views. (*B*) One ECD accounts for more than 80% of the total experimental variance. (*C*) Superimposition of the best fitting ECD onto the closest left sagittal MRI section from one subject. The dipolar source was projected within 15 mm from the superior colliculus. (*D*) Theoretical potential distributions reconstructed by the model.

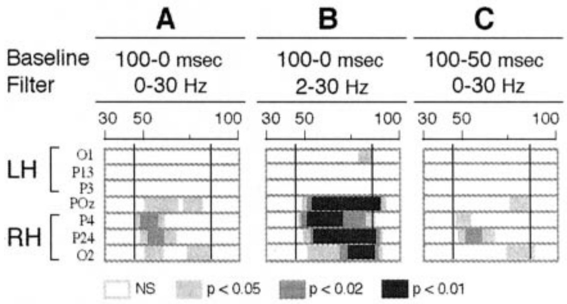

Figure 6. Statistical significance (Student's *t*-tests) of the [AV – (Au+Vi)] difference waveform at occipito-parietal electrodes between 30 and 100 ms post-stimulus in three conditions of data analysis differing according to the digital high-pass filter applied to the data (0 or 2 Hz), and according to the timing of the baseline reference used. (*A*) Standard conditions (similar to Fig. 3*A*). (*B*) Using a 2 Hz cut-off frequency did not decrease the significance of the interactions. (*C*) Changing the baseline reference period from 100 to 0 ms before stimulus onset to 100 to 50 ms before stimulus did not shorten the latency of the interaction effects (see text).

Fz, F4, F8, FC1, FC2, C3, Cz, CP1, CP2, Pz and P4. Statistical significance reached 0.01 from 110 to 135 ms at most of these electrodes (details in Fig. 3; mean amplitude over this spatio-temporal window: –0.68 μV; peak amplitude: –0.94 μV).

Interestingly, around the same latencies, SCD analysis revealed only very weak patterns of current (mean amplitude: –0.04 μA/m^3; Fig. 4: line 3, column 5). Given that SCDs decrease their amplitude more rapidly at the scalp surface with the deeper location of the intracerebral generators than the potentials do (Perrin *et al.*, 1987; Pernier *et al.*, 1988), the difference in morphology between the potential and the SCD distributions of [AV – (Au + Vi)] at the scalp strongly suggests that the underlying neural activities are localized in deep brain structures, possibly the superior colliculus known to be a major site of cross-modal integration.

To further test this hypothesis, we modeled these activities (in the grand-average waveforms) using one single ECD in a spherical head model (see Materials and Methods) in the 120–135 ms period, during which the topography of the effects was stable. The best fitting ECD was found at an eccentricity of 0.3 and explained the spatio-temporal distribution of the potentials between 120 and 135 ms with a goodness-of-fit of 81.7% (Fig. 5). The solution was stable whatever the different starting dipole parameters. The ECD projection onto the MRI of one subject using an automatic adjustment of the sphere model to the subject's head (see Materials and Methods) was found within 15 mm from the superior colliculus (Talairach coordinates of the projection: $x = 18$, $y = -22$, $z = -8$), a distance of the order of the intrinsic precision of spherical head dipole modeling in single subjects (Yvert *et al.*, 1997).

After 150 ms over Right Temporo-frontal Scalp Areas

From 170–185 ms, [AV – (Au + Vi)] displayed significant negative amplitudes ($P < 0.05$) at several temporo-frontal electrodes of the right hemiscalp. The effect began around Fp2 and T4, and reached F8 and FT4 sites at 175 ms (mean amplitude over 170–185 ms: –0.61 μV; peak amplitude: –0.72 μV; Fig. 3*C* and Fig. 4: line 4, column 4). No similar activity could be observed on the left hemiscalp (mean amplitude in the same spatio-temporal range: –0.16 μV). In addition, this pattern corresponded neither in latency nor in topography to any activity elicited by the unimodal stimuli (Fig. 4: line 4, columns 1–2), suggesting that this effect was due to neuronal populations in the anterior part of the right hemisphere responding only to bimodal inputs.

Discussion

As expected from behavioral studies, we found that subjects detected auditory stimuli more rapidly than visual stimuli (Hershenson, 1962; Welch and Warren, 1986), and that they reacted faster to bimodal than to either unimodal stimuli alone (Miller, 1986; Giray and Ulrich, 1993; Hughes *et al.*, 1994). This behavioral facilitation in bimodal processing was associated with multiple cross-modal neural activities, the nature and timing of which are discussed below.

Early ERP Effects: Genuine Cross-modal Interactions?

The earliest significant differences between the responses to bimodal objects and the sum of the unimodal responses appeared ~45 ms after stimulus onset. As specified in the Materials and Methods, the additive model is valid only while the analysis period is not contaminated by non-specific activities that would be common to all three types (Au, Vi, AV) of stimuli, and therefore not eliminated in the [AV – (Au + Vi)] difference wave. Among those non-specific activities, Teder-Sälejärvi *et al.* (Teder-Sälejärvi *et al.*, 2002) showed that anticipatory effects, manifested as slow potentials arising before each stimulus and continuing for a time after, can contribute to the bimodal minus unimodal difference waveform. When present, these activities usually appear as negative-going shifts in the pre-stimulus period. To dissociate these anticipatory effects from genuine cross-modal interactions, it was proposed either to high-pass filter the data with a 2 Hz cut-off frequency (which greatly decreases the amplitudes of the slow anticipatory potentials), or to change the reference baseline period (Teder-Sälejärvi *et al.*, 2002). Although our data did not reveal slow negative shifts in the pre-stimulus period (Fig. 2), they were submitted to these two control analyses. First, a 2 Hz high-pass filter did not decrease the amplitude or significance of the early interactions (Fig. 6*A*,*B*); second, unlike the results expected if anticipatory activities were present, changing the reference baseline period from [–100 to 0] ms to [–100 to –50] ms did not shorten the latency of the earliest effects (Fig. 6*A*,*C*).

We may therefore conclude that most of the early [AV – (Au + Vi)] effects reflect genuine cross-modal interactions. [The differences between Teder-Sälejärvi *et al.*'s observations and our data may be explained by the less 'comfortable' timing and attentional conditions in the former experimental design: in that study, subjects had to perform rapid and difficult discrimination between stimuli presented for a very short time (33 ms) and at fast rate (ISI of 600–800 ms), whereas in our experiment the stimuli were presented for 240 ms with an ISI varying on average from 1350 to 3250 ms. In the former case, anticipation can have been beneficial to subjects' efficiency while it was not necessary with our design.] While such early effects were also found in previous studies (Giard and Peronnet, 1999; Foxe *et al.*, 2000), they remain difficult to explain in the light of our current knowledge on sensory transmission. Of particular interest, however, are very recent findings in monkeys showing the existence of direct projections from the primary auditory cortex (usually already activated before 20–25 ms post-stimulus) to low-level areas (V1 and V2) of the visual cortex (Falchier *et al.*, 2001; Rockland and Ojima, 2001). [Note that similar projections from visual to auditory cortex (Schroeder *et al.*, 1995) and from somatosensory to auditory cortex (Fu *et al.*, 2001; Schroeder *et al.*, 2001) have also been described in monkeys.] While the functional role of such connections is not known, they might participate in increasing the efficiency of bimodal processing with a timing consistent with the early latencies observed here.

Multiple Audio-visual Interactions Below 200 ms Post-stimulus

Sensory-specific Cortical Areas (45–85 ms)

As discussed above, the significant [AV – (Au + Vi)] amplitudes on posterior visual scalp sites around 45–85 ms would indicate that adding an auditory cue to the visual input influenced the sensory analysis in visual cortex. The latency and topography of the effects correspond to those of the C1 component of visual ERPs, thought to be generated in striate and/or extrastriate cortices (Clark *et al.*, 1995; Foxe and Simpson, 2002). These interactions might therefore reflect also increased activities of the unimodal C1 generators (that would not be observable in the unimodal visual condition because of the weak saliency of the circle deformations). Whatever the exact mechanisms (modulation of C1 component or recruitment of neurons in visual cortex not activated by the sole visual inputs), they add to the growing number of findings in non-human primates (Watanabe and Iwai, 1991; Rockland and Ojima, 2001; Schroeder *et al.*, 2001; Schroeder *et al.*, 2002) and in humans (Sams *et al.*, 1991; Calvert *et al.*, 1999; Giard and Peronnet, 1999; Foxe *et al.*, 2000; Teder-Sälejärvi *et al.*, 2002), showing that cross-modal effects may occur in brain areas usually considered as sensory-specific.

Superior Colliculus (105–140 ms)?

Between ~105 and 140 ms, combined analysis of scalp potential and current density distributions, as well as the results of spatio-temporal dipole modeling, strongly suggest the existence of cross-modal interactions in deep structures of the brain, possibly the superior colliculus. While the non-unicity of dipole model solutions and the spatial precision of ERPs do not allow us to localize the effects precisely and unequivocally, the superior colliculus appears as a likely candidate in this brain region. Multiple electrophysiological studies in non-human primates and other mammals have shown that the deep layers of the superior colliculus contain multisensory cells that multiply their firing rate when two stimuli of different modalities are presented in close spatial and temporal proximity [reviewed in (Stein and Meredith, 1993)]. In addition, a recent fMRI study in humans has shown a major activation of this structure when subjects passively perceive non-speech audio-visual stimuli (Calvert *et al.*, 2001). Given the small size of this structure and its depth in the brain, it is worth noting that significant ERP amplitudes at the scalp surface would imply a vigorous response within the midbrain, a conclusion similar to that drawn from the findings of superadditive BOLD response enhancement in this structure (Calvert *et al.*, 2001). Interestingly, animal studies have shown that the multisensory integrative cells in the superior colliculus receive projections from cortical neurons (Wallace *et al.*, 1993; Jiang *et al.*, 2001): the timing of such cortico-tectal connections would be consistent with the relatively long latency (105–140 ms) of our effects.

We cannot rule out, however, that the large extent of the 105–140 ms interaction pattern in voltage maps could be, partially or wholly, explained by a more complex combination of activities distributed in deep cortical sulci and/or in widely distributed cortical areas. Indeed, monkeys studies have provided evidence for multisensory integration or convergence in the intraparietal sulcus (Lewis and van Essen, 2000), the posterior parietal cortex (Andersen, 1997), the superior temporal sulcus (Benevento *et al.*, 1977; Bruce *et al.*, 1981; Watanabe and Iwai, 1991; Lewis and van Essen, 2000) and the prefrontal cortex (Gaffan and Harrison, 1991), and neuroimaging studies have shown activation in the homologues of these regions in humans (Paulesu *et al.*, 1995; Calvert *et al.*, 2000, 2001; Downar *et al.*, 2000; Raij *et al.*, 2000; Bushara *et al.*, 2001; Callan *et al.*, 2001). While these hypotheses cannot be ruled out, again, the weak SCD amplitudes associated with the voltage maps together with the results of dipole modeling rather support a deep source hypothesis.

Right Temporo-frontal Scalp Areas (170–185 ms)

The cross-modal interactions found over the right temporo-frontal scalp sites between 170 and 185 ms did not correspond to any unimodal ERP activity. This brain region therefore would be specifically activated by the multimodal nature of the stimulus. While, again, ERPs do not allow for precise localization, the topography of the effect is consistent with sources in the right insula, the right prefrontal cortex and/or the temporo-polar cortex. Both animal studies (Jones and Powell, 1970; Benevento *et al.*, 1977; Füster *et al.*, 2000) and functional neuroimaging experiments in humans have shown that these sites could be involved in multisensory convergence or integration. Particularly the right insula has been repeatedly found active when two stimuli from different modalities were presented in temporal synchrony (Paulesu *et al.*, 1995; Hadjikhani and Roland, 1998; Downar *et al.*, 2000; Bushara *et al.*, 2001; Calvert *et al.*, 2001). Both the experimental design used and the topography of our effects fit with implication of this structure.

Influence of the Perceptual Task on Multisensory Integration

The results discussed above indicate that the detection of bimodal audio-visual targets induces multiple cross-modal operations in different cerebral structures within the first 200 ms of stimulus analysis. In the same way, we had shown in a previous experiment (referred to as the *recognition* study), that the identification of objects defined by auditory and visual components was facilitated (compared to the same objects presented unimodally), and also generated complex neural interaction patterns in the same latency range (Giard and Peronnet, 1999). The use of strictly identical stimuli in both studies allows us to compare the brain responses in the two experiments and to evaluate the influence of the task on the neural mechanisms of multisensory integration.

First, in both studies, early interaction patterns (from 45 ms) were found at posterior scalp sites. Although these cross-modal patterns were of weaker amplitudes and shorter duration in the *detection* than in the *recognition* task, they shared the same occipital topography typical of activities in visual cortical areas. In the *recognition* study, however, the magnitude of these early interactions partly depended on the dominant sensory modality of the subject for the task. Indeed, in that experiment, the subjects were divided into two groups according to a reaction time criterion in unimodal processing: subjects who were faster to identify the auditory than the visual objects were called 'auditory-dominant', while those who were more rapid for visual than auditory objects were 'visually-dominant' subjects. Interestingly, we found that the early interactions in visual cortex (40–150 ms) were much larger in the auditory-dominant group, while the visually-dominant subjects showed significant interactions in the auditory cortex (~100 ms). In other words, at an early stage of sensory analysis, the cross-modal interactions were predominant in the cortex of the non-dominant modality (Giard and Peronnet, 1999). In the present *detection* experiment, all the subjects were faster to respond to auditory than to visual stimuli and may therefore be considered, according to our criteria, as 'auditory-dominant' for that particular task. Thus, the significant interactions in visual areas (without effects at typical auditory scalp sites) in this experiment would suggest that multisensory integration operates, at an early processing stage, similarly in the two perceptual tasks by influencing predominantly the sensory cortex of the non-dominant modality.

In the *recognition* study, however, unlike the *detection* experiment, another form of neural facilitation was manifested in the visual cortex as an amplitude decrease of the unimodal visual N1 component (185 ms latency). It has been shown recently that the visual N1 deflection has a larger amplitude during discrimination than in simple reaction time tasks, thereby suggesting that it would be partly related to 'a discrimination process within the focus of attention' (Vogel and Luck, 2000). In accordance with these findings, the unimodal N1 amplitude was found to be larger in the *recognition* (–7.0 μV) than in the *detection* (–4.7 μV) task. More importantly — and as expected since facilitation for a discrimination process is not relevant for a detection task — no significant cross-modal effect was found on the visual N1 amplitude in the present experiment. These differential effects on N1 together with the earlier interactions discussed above emphasize the complexity and flexibility of the integrative processes already within the brain regions traditionally held for unisensory.

Another major difference between the integrative mechanisms in object recognition and in simple detection is the probable activation of deep sources, possibly the superior colliculus, in this last task ~105–140 ms. Multisensory integration in this midbrain structure is usually associated with facilitation for orientation and localization. While the detection task used here did not include any spatial or orienting aspect, it required only low-level sensory analysis, similarly to that needed to orient to a target whatever the stimulus content. This interpretation could also explain the predominant activation of the superior colliculus when subjects passively perceived non-speech audio-visual stimuli, in Calvert *et al.*'s fMRI study (Calvert *et al.*, 2001). Whatever the precise structures involved, the fact that this interaction pattern was not observed during the *recognition* task (Giard and Peronnet, 1999) may indicate that there was no cross-modal neural facilitation in these structures when deeper stimulus analysis was necessary. This again underlines the exquisite flexibility of the integrative processes, that probably adapt for the most efficient result at the lowest energy cost.

Lastly, in both *detection* and *recognition* studies, and in the latter in all (auditory- and visually-dominant) subjects, we found significant cross-modal interactions over the right temporo-frontal scalp regions with latencies varying between 140 and 185 ms post-stimulus. A similar pattern was also observed when subjects had to identify the same audio-visual objects containing non-redundant unimodal information (Fort *et al.*, 2002). To explain these last results, we proposed that interactions in those non-specific areas were related to a facilitation process only for detection of bimodal inputs (indeed, object recognition was not facilitated by the bimodal content of the stimulus since both unimodal cues had to be identified to achieve correctly the task). This interpretation is not refuted by the present experiment and fits, as discussed above, with implication of the right insula (Bushara *et al.*, 2001). In any case, activation of this brain region seems crucial after bimodal inputs since it has been observed in all our experimental designs. These non task-related activities therefore appear to add to the cross-modal operations differentially induced according to the nature of the perceptual task, possibly to increase the functional efficiency of multisensory integration (facilitation) for that particular task.

Notes

We are grateful to J.F. Echallier and P.E. Aguera for technical assistance. This work was supported by a grant from the French Ministry of Education and Research.

Address correspondence to Alexandra Fort, INSERM U280, 151 Cours Albert Thomas, 69424 Lyon cedex 03, France. Email: fort@lyon.inserm.fr.

References

Andersen RA (1997) Multimodal integration for the representation of space in the posterior parietal cortex. Philos Trans R Soc Lond Biol 352:1421–1428.

Barth DS, Goldberg N, Brett B, Di S (1995) The spatiotemporal organization of auditory, visual and auditory–visual evoked potentials in rat cortex. Brain Res 678:177–190.

Benevento LA, Fallon J, Davis BJ, Rezak M (1977) Auditory–visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. Exp Neurol 57:849–872.

Bremmer F, Schlack A, Shah NJ, Zafiris O, Kubischik M, Hoffmann KP, Zilles K, Fink GR (2001) Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. Neuron 29:287–296.

Bruce C, Desimone R, Gross CG (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. J Neurophysiol 46:369–383.

Bushara KO, Grafman J, Hallett M (2001) Neural correlates of auditory–visual stimulus onset asynchrony detection. J Neurosci 21:300–304.

Callan DE, Callan AM, Kroos C, Vatikiotis-Bateson E (2001) Multimodal contribution to speech perception revealed by independent component analysis: a single-sweep EEG case study. Cogn Brain Res 10:349–353.

Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS (1999) Response amplification in sensory-specific cortices during crossmodal binding. Neuroreport 10:2619–2623.

Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 10:649–657.

Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001) Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. Neuroimage 14:427–438.

Clark VP, Fan S, Hillyard SA (1995) Identification of early visual evoked potential generators by retinotopic and topographic analyses. Hum Brain Mapp 2:170–187.

Downar J, Crawley AP, Mikulis DJ, Davis KD (2000) A multimodal cortical network for the detection of changes in the sensory environment. Nat Neurosci 3:277–283.

Echallier JF, Perrin F, Pernier J (1992) Computer-assisted placement of electrodes on the human head. Electroencephalogr Clin Neurophysiol 82:160–163.

Ettlinger G, Garcha HS (1980) Cross-modal recognition by the monkey: the effects of cortical removals. Neuropsychologia 18:685–692.

Falchier A, Renaud L, Barone P, Kennedy H (2001) Extensive projections from the primary auditory cortex and polysensory area STP to peripheral area V1 in the macaque. Soc Neurosci Abstr 27:511.21.

Fort A, Delpuech C, Pernier J, Giard MH (2002) Early auditory–visual interactions in human cortex during nonredundant target identification. Cogn Brain Res 14:20–30.

Foxe JJ, Morocz IA, Murray MM, Higgins BA, Javitt DC, Schroeder CE (2000) Multisensory auditory–somatosensory interactions in early cortical processing revealed by high-density electrical mapping. Cogn Brain Res 10:77–83.

Foxe JJ, Simpson GV (2002) Flow of activation from V1 to frontal cortex in humans: a framework for defining 'early' visual processing. Exp Brain Res 142:139–150.

Fu KG, Johnston TA, Shah AS, Arnold L, Smiley J, Hackett TA, Garraghty PE, Schroeder CE (2001) Characterization of somatosensory input to auditory association cortex in macaques. Soc Neurosci Abstr 27:681.3.

Füster JM, Bodner M, Kroger JK (2000) Cross-modal and cross-temporal association in neurons of frontal cortex. Nature 405:347–351.

Gaffan D, Harrison S (1991) Auditory–visual associations, hemispheric specialization and temporo-frontal interaction in the rhesus monkey. Brain 114:2133–2144.

Giard MH, Peronnet F (1999) Auditory–visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci 11:473–490.

Giard MH, Perrin F, Echallier JF, Thevenet M, Froment JC, Pernier J (1994) Dissociation of temporal and frontal components in the human auditory N1 wave: a scalp current density and dipole model analysis. Electroencephalogr Clin Neurophysiol 92:238–252.

Giray M, Ulrich R (1993) Motor coactivation revealed by response force in divided and focused attention. J Exp Psychol Hum Percept Perform 19:1278–1291.

Gonzalo D, Shallice T, Dolan R (2000) Time-dependent changes in learning audiovisual associations: a single-trial fMRI study. Neuroimage 11:243–255.

Hadjikhani N, Roland PE (1998) Cross-modal transfer of information between the tactile and the visual representations in the human brain: a positron emission tomographic study. J Neurosci 18:1072–1084.

Hershenson M (1962) Reaction time as a measure of intersensory facilitation. J Exp Psychol 63:289–293.

Hikosaka K, Iwai E, Saito H, Tanaka K (1988) Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. J Neurophysiol 60:1615–1637.

Hughes HC, Reuter-Lorenz PA, Nozawa G, Fendrich R (1994) Visual–auditory interactions in sensorimotor processing — saccades versus manual responses. J Exp Psychol Hum Percept Perf 20:131–153.

Jiang W, Wallace MT, Jiang H, Vaughan W, Stein BE (2001) Two cortical areas mediate multisensory integration in superior colliculus neurons. J Neurophysiol 85:506–522.

Jones EG, Powell TPS (1970) An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. Brain 93:793–820.

Lewis JW, van Essen DC (2000) Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. J Comp Neurol 428:112–137.

Miller JO (1982) Divided attention: evidence for coactivation with redundant signals. Cognit Psychol 14:247–279.

Miller JO (1986) Time course of coactivation in bimodal divided attention. Percept Psychophys 40:331–343.

Miller JO (1991) Channel interaction and the redundant targets effect in bimodal divided attention. J Exp Psychol Hum Percept Perform 17:160–169.

Paulesu E, Harrison J, Baron Cohen S, Watson JDG, Goldstein L, Heather J, Frackowiak RSJ, Frith CD (1995) The physiology of coloured hearing. A PET activation study of colour-word synaesthesia. Brain 118:661–676.

Pernier J, Perrin F, Bertrand O (1988) Scalp current density fields: concept and properties. Electroencephalogr Clin Neurophysiol 69:385–389.

Perrin F, Pernier J, Bertrand O, Giard MH (1987) Mapping of scalp potentials by surface spline interpolation. Electroencephalogr Clin Neurophysiol 66:75–81.

Perrin F, Pernier J, Bertrand O, Echallier JF (1989) Spherical splines for scalp potential and current density mapping. Electroencephalogr Clin Neurophysiol 72:184–187.

Raab DH (1962) Statistical facilitation of simple reaction times. Trans NY Acad Sci 24:574–590.

Raij T, Uutela K, Hari R (2000) Audiovisual integration of letters in the human brain. Neuron 28:617–625.

Rockland KS, Ojima H (2001) Calcarine area V1 as a multimodal convergence area. Soc Neurosci Abstr 27:511.20.

Rugg MD, Doyle MC, Wells T (1995) Word and nonword repetition within- and across-modality: an event-related potential study. J Cogn Neurosci 7:209–227.

Sakowitz OW, Quiroga RQ, Schürmann M, Basar E (2001) Bisensory stimulation increases gamma-responses over multiple cortical regions. Cogn Brain Res 11:267–279.

Sams M, Aulanko R, Hamalainen H, Hari R, Lounasmaa OV, Lu ST, Simola J (1991) Seeing speech: visual information from lip movements modify activity in the human auditory cortex. Neurosci Lett 127:141–145.

Scherg M (1990) Fundamentals of dipole source potential analysis. In: Auditory evoked magnetic fields and electric potentials. Advances in audiology, vol. 5 (Grandori F, Hoke M, Romani GL, eds), pp. 40–69. Karger: Basel.

Scherg M, von Cramon D (1986) Evoked dipole source potentials of the human auditory cortex. Electroencephalogr Clin Neurophysiol 65:344–360.

Schroeder CE, Foxe JJ (2002) Timing and laminar profile of converging inputs in multisensory areas of macaque neocortex. Cogn Brain Res 14:187–198.

Schroeder CE, Mehta AD, Ulbert I, Steinschneider M, Vaughan HG (1995) Visual responses in auditory cortex and their modulation by attention. Soc Neurosci Abstr 21:694.

Schroeder CE, Lindsley RW, Specht C, Marcovici A, Smiley JF, Javitt DC (2001) Somatosensory input to auditory association cortex in the macaque monkey. J Neurophysiol 8:1322–1327.

Stein BE, Meredith MA (1993) The merging of the senses. Cambridge, MA, USA: The MIT Press.

Stein BE, Wallace MT (1996) Comparison of cross-modality integration in midbrain and cortex. In: Progress in brain research, vol. 112 (Norita M, Bando T, Stein B, eds), pp. 289–299. Elsevier Science: Amsterdam.

Stein BE, Meredith MA, Wallace MT (1993) The visually responsive neuron and beyond: multisensory integration in cat and monkey. Prog Brain Res 95:79–90.

Teder-Sälejärvi WA, McDonald JJ, Di Russo F, Hillyard SA (2002) An analysis of audio-visual crossmodal integration by means of potential (ERP) recordings. Cogn Brain Res 14:106–114.

Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. Nature 381:520–522.

Toldi J, Fehér O, Wolff JR (1986) Sensory interactive zones in the rat cerebral cortex. Neuroscience 18:461–465.

Vaughan HG, Ritter W (1970) The sources of auditory evoked responses recorded from the human scalp. Electroencephalogr Clin Neurophysiol 28:360–367.

Vogel EK, Luck SJ (2000) The visual N1 component as an index of a discrimination process. Psychophysiology 37:190–203.

Wallace MT, Meredith MA, Stein BE (1993) Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus. J Neurophysiol 69:1797–1809.

Watanabe J, Iwai E (1991) Neuronal activity in visual, auditory and polysensory areas in the monkey temporal cortex during a visual fixation task. Brain Res Bull 26:583–592.

Welch, RB, Warren, DH (1986) Intersensory interactions. In: Handbook of perception and human performance, vol. I: Sensory processes and perception (Boff KR, Kaufman L, Thomas JP, eds), pp. 25/1–25/36. New York: Wiley.

Yvert B, Bertrand O, Thevenet M, Echallier JF, Pernier J (1997) A systematic evaluation of the spherical model accuracy in EEG dipole localization. Electroencephalogr Clin Neurophysiol 102:452–459.