Auditory-Visual Integration during Multimodal Object Recognition in Humans: A Behavioral and Electrophysiological Study

M. H. Giard and F. Peronnet
Mental Processes and Brain Activation, INSERM-U280, Lyon, France

Abstract

The aim of this study was (1) to provide behavioral evidence for multimodal feature integration in an object recognition task in humans and (2) to characterize the processing stages and the neural structures where multisensory interactions take place. Event-related potentials (ERPs) were recorded from 30 scalp electrodes while subjects performed a forced-choice reaction-time categorization task: At each trial, the subjects had to indicate which of two objects was presented by pressing one of two keys. The two objects were defined by auditory features alone, visual features alone, or the combination of auditory and visual features. Subjects were more accurate and rapid at identifying multimodal than unimodal objects. Spatiotemporal analysis of ERPs and scalp current densities revealed several auditory-visual interaction components temporally, spatially, and functionally distinct before 200 msec poststimulus. The effects observed were (1) in visual areas, new neural activities (as early as 40 msec poststimulus) and modulation (amplitude decrease) of the N185 wave to unimodal visual stimulus, (2) in the auditory cortex, modulation (amplitude increase) of subcomponents of the unimodal auditory N1 wave around 90 to 110 msec, and (3) new neural activity over the right fronto-temporal area (140 to 165 msec). Furthermore, when the subjects were separated into two groups according to their dominant modality to perform the task in unimodal conditions (shortest reaction time criteria), the integration effects were found to be similar for the two groups over the nonspecific fronto-temporal areas, but they clearly differed in the sensory-specific cortices, affecting predominantly the sensory areas of the nondominant modality. Taken together, the results indicate that multisensory integration is mediated by flexible, highly adaptive physiological processes that can take place very early in the sensory processing chain and operate in both sensory-specific and nonspecific cortical structures in different ways.

INTRODUCTION

Since the early formal description of our sensory organs by Aristotle about 350 A.D., philosophers and scientists have kept on questioning how individuals achieve perceptual experience of the surrounding world events. Put in more topical terms, one fundamental question for psychologists and neuroscientists today is: How do the sensory systems integrate separate features of an “object” to form a unitary percept?

The question of perceptual integration has been widely investigated within separate sensory systems at both psychological and neurophysiological levels, especially in the visual modality (e.g., Oram & Perrett, 1996; Singer & Gray, 1995; Treisman, 1996) and to a less extent in the auditory modality (e.g., Loveless, Levénén, Jousmäki, Sams, & Hari, 1996). Yet our daily perceptions and behaviors are not based on the juxtaposition of separate, independent sensory experiences but on the global impression resulting from the combination of components from different modalities. For example, we orient more easily toward a cue defined by auditory and visual components, or we may feel sick when visual and vestibular informations conflict.

Although multisensory convergence and integration “are far more ancient schemes than is the segregation of different senses” (Stein, London, Wilkinson, & Price, 1996), research on the mental and brain processes that subserve them shows surprisingly important gaps. In humans, most research work on multisensory integration is still at the stage of demonstrating the phenomenon and understanding the operative factors at perceptual and behavioral levels (for review, see O’Hare, 1991; Stein & Meredith, 1993; Welch & Warren, 1986). Most frequently investigated are the aspects related to intersensory bias, that is, the capacity of a sensory system to modify—enhance, degrade—the perception from another system (e.g., Bernstein, 1970; Hershenson, 1962; Hubbard, 1966; McGurk & MacDonald, 1976; O’Leary & Rhodes, 1984; Sekuler, Sekuler, & Lau, 1997) and the relationships between this bias and multimodal spatial processing (how the sensory signals compete with one another according to the relative spatial position of the unimodal cues) (Bertelson & Radeau, 1981; Ehrenstein &
Reinhardt-Rutland, 1996; Fisher & Pylyshyn, 1994; Frens, Vanopstal, & Vanderwilligen, 1995; Held, 1955; Pick, Warren, & Hay, 1969; Shelton & Searle, 1980; Stein et al., 1996; Stein, Meredith, Huneycutt, & McDade, 1989; Warren, McCarthy, & Welch, 1983) or the relationships between intersensory bias and the subject’s attentional focus (Bernstein, 1970; Miller, 1982, 1986; Spence & Driver, 1997). In these investigations, the tasks have been most often to detect, localize, and/or react to stimuli containing cues from one or more sensory modalities. Perceptual integration, however, can be defined as a process of physiological input systems that not only facilitates detection by amplification of the sensory signals but can combine these signals to form a new, multimodal representation of the source object (O’Hare, 1991). Yet, very few studies on multisensory integration have focused on the genuine perceptual question of multimodal object recognition.

In behavioral studies, a general finding has been that the reaction time (RT) to stimuli containing redundant bimodal information is shorter than that to either unimodal stimuli. This effect, known as the “redundant-signal effect” (Miller, 1982, 1986), has given rise to two explicative models. The race model assumes that the shorter RT is due to triggering the response on the basis of the first detected cue. The coactivation model claims that the parallel processing of unimodal channels interacts somewhere in the processing system, the question being the level at which interactions occur—sensory processing, response selection, or motor execution. Psychophysic studies have provided arguments for all three levels (Fournier & Eriksen, 1990; Hughes, Reuter-Lorenz, Nozawa, & Fendrich, 1994; Miller, 1982; Schmidt, Giele, & van den Heuvel, 1984), but they could not give information on the neural sites of interaction at the sensory processing level.

To date, the investigations made using event-related potentials (ERPs) (Andrassi & Greco, 1975; Costin, Neville, Meredith, & Stein, 1991; Schröger & Widmann, 1997, 1998; Walter, 1964) or magnetoencephalography (MEG) (Sams et al., 1991) have still more often contributed to the demonstration of the phenomenon rather than to the knowledge of the underlying physiological processes. Recent neuroimaging data have, however, begun to shed light on the neural sites of multisensory convergence: a positron emission tomography (PET) study has identified several cortical areas of interaction in synaesthete subjects (subjects making color-word associations) in the parieto-occipital junction and in the right prefrontal cortex (Paulesu et al., 1995). MEG recordings have reported auditory-visual interactions in the “visual projection area” (Busch, Wilson, Orr, & Papanicolaou, 1989) and in the right parieto-temporal area (Sams & Imada, 1997).

Animal studies have provided significant insights into the physiological mechanisms of multisensory integration at the single neuron level. Probably the best-known site of multimodal convergence and integration is the superior colliculus, a midbrain structure that plays a fundamental role in attentive and orientation behaviors (for review, see Stein & Meredith, 1993). Multisensory neurons have been found in the deep layers of the superior colliculus in cats (e.g., Gordon, 1973; Meredith & Stein, 1983; Peck, 1987) monkeys (e.g. Jay & Sparks, 1984), and guinea pigs, mice, hamsters, and rats (for review, see Wallace, Wilkinson, & Stein, 1996). Multisensory neurons in the superior colliculus receive afferents from the three sensory modalities, each modality being represented in this structure by a map of sensory space. The different maps overlap each other so that stimuli of different sensory modalities originating from the same spatial source activate the same region of superior colliculus (King & Palmer, 1985; Meredith & Stein, 1986a). These sensory maps are also in register with motor representations, thereby allowing the different sensory inputs to produce identical behavior (Stein & Meredith, 1993), for example, saccadic eye movements (Hughes et al., 1994; Frens et al., 1995).

The multisensory neuron in the superior colliculus not only responds to inputs from several modalities but also integrates information from different modalities by increasing the number of impulses in a multiplicative ratio when presented with multimodal inputs (Meredith & Stein, 1986b; Wallace et al., 1996). Detailed studies on the characteristics and properties of multisensory neurons in the cat superior colliculus led Stein and colleagues to define four “integration rules” (review in Stein & Meredith, 1993): two rules on temporal and spatial coincidence, according to which the largest integration effects are obtained when inputs from different modalities are in close temporal and spatial relationships, respectively; the “magnitude or inverse effectiveness rule,” stipulating that the less effective the unimodal stimuli are, the larger the magnitude of the enhancement they are capable of generating by combination (Stein, Meredith, & Wallace, 1994); and the receptive field preservation rule. These integration rules, devised at the single neuron level in the superior colliculus, were found to apply also to the attentive and orientation behaviors mediated by this structure (Stein, Huneycutt, & Meredith, 1988; Stein et al., 1989).

Sites of multisensory convergence have also been evident at cortical level in cats (e.g., Roda & Reinoso-Suarez, 1983; Toldi, Feher, & Feuer, 1984; Kimura & Tamai, 1992; Wallace, Meredith, & Stein, 1992), monkeys (Hikosaka, Iwai, Saito, & Tanaka, 1988; Mistlin & Perrett, 1990; Duhamel, Colby, & Goldberg, 1991) and rats (Barth, Goldberg, Brett, & Di, 1995). Although the same rules appear to govern multisensory integration in the superior colliculus and cortex, differences in spatial factors sensitivity were observed between the two neural systems (Stein & Wallace, 1996) and between the behaviors (probably mediated by these two systems (orientation vs. perceptual judgments: Stein et al., 1996). Furthermore, cat stud-
ies have shown that the multisensory neurons in the cortex and in the superior colliculus belong to separate neural circuits, in the sense that they do not project to, or receive inputs from, each other (Wallace, Meredith, & Stein, 1993). All these observations suggest that the multisensory systems in the cortex and in superior colliculus may underlie different integrative functions (Stein et al., 1996; Wallace et al., 1993; Hughes et al., 1994), although very little is still known about the cortical multisensory interactions and the integrative processes mediated by the cortex.

We report here a behavioral and electrophysiological study in humans, aimed at shedding light on the neural (cortical) substrates and mechanisms underlying multimodal feature integration in an object recognition task. Two objects, A and B, were devised, following the rules for multisensory integration (Stein & Meredith, 1993), and were presented randomly to subjects. The objects were defined either by the combination of visual and auditory features or by unimodal features alone and consisted of the deformation of a circle to an ellipse producing a sound, the circle deformation alone, or the sound alone (Figure 1). The subject’s task was to identify, at each trial, object A or B by pressing a key with the index or middle finger, respectively. In a pilot experiment (not reported here), we ensured that this recognition task presented the behavioral characteristics of multisensory integration—that is, that identification of objects A and B was improved (reaction time and accuracy) when the objects were presented with multimodal features compared to unimodal features alone. In the present study, the behavioral performances and the spatiotemporal distributions of ERPs and scalp current densities (SCDs) from 30 electrodes were analyzed in 24 subjects to characterize the processing stages and central loci of auditory-visual interactions. Particular emphasis was given to early processing analysis (below 200 msec poststimulus) for several reasons: (1) given the number of neurophysiological evidences for subcortical interactions in animal studies, one may expect to observe integration effects during the early stages of sensory processing, (2) in humans, it has been shown that, in some experimental situations, visual object categorization may be achieved within 150 msec (Thorpe, Fize, & Marlot, 1996), and (3) below about 200 msec, auditory and visual ERPs are known to reflect modality-specific activities; the interaction effects may therefore be estimated using a simple model, in the difference between the ERPs to multimodal objects and the sum of ERPs to unimodal objects, a model already used by Barth et al. (1995).

**BEHAVIORAL RESULTS**

Subjects (n=24) identified objects A and B more rapidly and with less errors when the objects combined auditory and visual (AV) features than when they were presented with only auditory (Au) or visual (Vi) features (Figures 1 and 2). Two-way analyses of variance (ANOVAs) with Type of object (A,B) and Stimulus features (Au, Vi, AV) as within-subject factors showed significant effects of stimulus features both on the reaction time (F(2, 46) = 16.80, p < 0.001, Greenhouse-Geisser ε = 0.65) and on accuracy (F(2, 46) = 9.42, p < 0.0004, ε = 0.66). There was no effect of the object type or interaction between the two factors. The reaction times were on average 64 msec shorter for identifying multimodal objects (562 msec) than unimodal auditory or visual objects (621 and 631 msec, respectively).

Although the visual modality is generally considered as dominant, some subjects (those generally having some elementary music or singing training) found it easier and were faster at identifying the objects on the basis of their auditory attributes than visual attributes and found it easier to do so. This led us to consider “VIS-subjects” (subjects having shorter RTs for visual object recognition in the unimodal conditions; n=14) and “AUD-subjects” (better at auditory object recognition; n=10) (Table 1) separately for some aspects of ERP analysis to

**Figure 1.** Two objects A and B were devised. Each of them was defined by either visual (Vi) features alone, auditory (Au) features alone, or the conjunction of auditory and visual (AV) features. Object A consisted in the deformation of a circle to an horizontal ellipse and/or a tone of 540 Hz and object B, in the deformation of a circle to a vertical ellipse and/or a tone of 560 Hz. The six stimuli were presented randomly to the subjects (n = 24) with equal probability. The task was to identify, at each trial, which object A or B was presented by pressing one of two keys.
compare the neural correlates of auditory-visual integration in subjects with different sensory modality dominance.

**ERP RESULTS**

Because objects A and B were physically close and identified with equivalent performance, they were grouped according to their modal features (Au, Vi, AV) in ERP averaging to increase the signal-to-noise ratio of the responses.

**ERPs to Unimodal (Au and Vi) Stimuli**

Figure 3 presents the ERPs elicited by unimodal (auditory and visual) and multimodal objects over 1000 msec at a subset of electrodes. In their early phase (below about 200 msec latency), the unimodal auditory and visual responses display “classical” morphologies known to reflect activities in the respective sensory-specific areas. In auditory ERPs, the N1 wave was maximum around 95-msec latency at frontal sites (~8.06 μV at Fz) and reversed polarity at mastoids (4.15 and 4.08 μV at M1 and M2, respectively). N1 was followed by the P2 wave emerging around 150 msec at centro-parietal sites and reaching its maximum amplitude around 180 msec (10.54 μV at Cz). In visual ERPs, the first salient deflection was a prominent wave peaking around 185 msec (N185) at occipito-parietal electrodes (~6.58 μV and ~7 μV at PO3 and PO4, respectively).

At later latencies, both Au and Vi stimuli elicit broad positive wave forms over the parietal sites (P3 waves) and maximum at Pz. P3 amplitude was larger for Vi than Au stimuli (16.6 vs. 11.1 μV, p < 0.001). P3 latencies did not differ significantly (Vi: 418 msec, Au: 382 msec, p < 0.08).

**ERPs to Multimodal (AV) Stimuli**

The early phase (<200 msec) of the ERPs to multimodal objects will be examined in detail below. At later latencies, multimodal ERPs were characterized by positive (P3) waves peaking at parietal electrodes (Figure 3). There was no clear relationship between the latencies or amplitudes of P3 elicited by unimodal and multimodal stimuli; P3 latency for AV stimuli (364 msec) was significantly shorter than for Vi stimuli (p < 0.001) but was not significantly different from Au stimuli; P3 amplitude for AV stimuli (16.0 μV) was significantly larger than for Au stimuli (p < 0.001) but did not differ from P3 amplitude to Vi stimuli. In addition, neither P3 amplitude nor latency showed any straightforward relationships with the reaction times (see Figure 2). The multimodal interactions possibly occurring later than 200 msec were therefore not further investigated in this study.

**Table 1.** Mean reaction times ± standard error (msec) for VIS-subjects and AUD-subjects in the recognition of objects A and B defined by multimodal (AV) or unimodal (Vi or Au) features. The reaction times to multimodal (AV) objects did not significantly differ between AUD- and VIS-subjects.

<table>
<thead>
<tr>
<th>VIS-subjects (n = 14)</th>
<th>AUD-subjects (n = 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AV</td>
<td>Vi</td>
</tr>
<tr>
<td>Object A</td>
<td>565 ± 32</td>
</tr>
<tr>
<td>Object B</td>
<td>545 ± 26</td>
</tr>
<tr>
<td>Average</td>
<td>555 ± 29</td>
</tr>
</tbody>
</table>
Auditory-Visual Interactions (All Subjects)

Figure 4 displays the superimposition of the ERPs to multimodal AV stimuli and the algebraic sum of ERPs to unimodal Au and Vi stimuli within 0- to 200-msec latency. Although the morphology of the multimodal response strongly resembles the sum of unimodal ERPs, the differences between the two traces at some electrodes and latencies indicate that the multimodal response is not just the linear sum of separately evoked unimodal responses. The auditory-visual interactions were therefore measured in the difference wave [AV = (Au + Vi)] between the multimodal ERPs and the sum of unimodal ERPs (see Methods). Student $t$ maps displaying the scalp areas where [AV = (Au + Vi)] amplitudes differ from zero were built at each latency. Those scalp areas where [AV = (Au + Vi)] had significant amplitudes for at least 15 consecutive milliseconds were considered as interaction patterns (see Methods).

Several interaction patterns could be temporally and topographically dissociated before 200 msec: over the posterior visual areas around 40 to 90 msec, 110 to 145 msec, and 155 to 200 msec; over the temporo-central areas around 95 and 180 msec; and on the right fronto-temporal hemiscalp around 140 to 165 msec. Figures 5 and 6 depict these different interaction patterns. Figure 6 displays for each of them, at an illustrative latency, the scalp potential (SP) and SCD distributions of [AV = (Au + Vi)], the corresponding $t$ map (computed on the potential values), and the unimodal and multimodal (Au, Vi, Au + Vi, and AV) responses at the same latency, to better interpret the nature of the interaction (modulation—increase or decrease—of unimodal responses or additional specific activity). Table 2 (first line) summarizes the mean amplitude and statistical significance of each interaction pattern. These different patterns are analyzed in detail below.

Interactions on Posterior Visual Areas

Significant [AV = (Au + Vi)] amplitudes were found over the whole 40- to 200-msec period at posterior sites. Figure 5 details the statistical significance of [AV = (Au + Vi)] amplitude at the nine posterior electrodes within this period. This figure together with the spatio-temporal analysis of [AV = (Au + Vi)] distributions led to dissociate three distinct and stable interaction patterns over the occipito-temporo-parietal areas: one from 40- to 90-msec latency, a second from 90 to 145 msec, and a third one from 155 to 200 msec.

40 to 90 msec (Figure 6a) In this latency range, [AV = (Au + Vi)] distribution was slightly asymmetric with the largest amplitudes on the right hemiscalp around O2 and PO4 (Figure 5; mean amplitude over O2-PO4 and 40 to 90 msec: 0.90 μV). SCD distributions were characterized by a sharp current source around O2-PO4 (mean amplitude: 0.35 μA/m$^3$). Figure 6a (right part) illustrates these SP and SCD distributions at a specific latency (70 msec).

No similar activation pattern could be observed in the unimodal responses during this period (Figure 6a, left part): The ERP amplitude to unimodal visual stimuli did not reach statistical significance at any of the nine posterior electrodes. This suggests that the interaction pattern observed is due to new activities in the visual
cortex, elicited by the conjunction of auditory and visual features in the stimulus (see Discussion).

90 to 145 msec (Figure 6b) Between 90 and 110 msec, the interaction pattern extended to mastoids with somewhat weaker amplitudes at occipital sites but still had high statistical significance (see Figure 5). Over all the 90- to 145-msec period, \([AV - (Au + Vi)]\) presented a maximum centered over the occipital electrodes between 01, 02, and Pz (mean amplitude: 0.74 mV). Figure 6b (right part) illustrates this effect at 130-msec latency.

Corresponding SCD maps showed equally stable current patterns composed of a crescent-like source spreading between PO3 and PO4 over the entire period (mean amplitude: 0.37 μA/m²). From about 115 to 135 msec, this current source was associated with current sinks of smaller amplitude at centro-parietal sites.

Over this period, the ERP amplitude to unimodal visual stimuli still did not reach statistical significance at any of the nine posterior electrodes, suggesting that here too, conjunction of auditory and visual inputs may have activated neurons in the visual cortex not (or very few)
activated by visual stimuli alone (see Discussion, however).

155 to 200 msec (Figure 6c) From about 155- to 200-msec latency, [AV - (Au + Vi)] amplitudes were highly significant over all nine occipito-temporal electrodes (Figure 5). [AV - (Au + Vi)] distributions revealed bilateral potential patterns, arising earlier and with larger amplitudes on the right than on the left hemiscalp. They reached their maxima at 185 msec around O2-PO4-T6 (2.20 µV) and O1-PO3-T5 (1.88 µV) (Figure 6c, right part). These interhemispheric differences were confirmed by statistical analyses: [AV - (Au + Vi)] amplitude began being significant (p < 0.01) at 160 msec on the right hemiscalp and at 172 msec on the left (see also Figure 5); the peak amplitude of [AV - (Au + Vi)] was

**Figure 6.** Auditory-visual interactions below 200-msec latency (all subjects). Lines (a) to (f) illustrate the different interaction effects found in occipital, temporal/central, and right fronto-temporal areas. Each line displays (1) the time window of the interaction effect (and the illustrative latency at which the interaction pattern is depicted), (2) the topographies (potential distributions) of the unimodal responses (Au, Vi), of the sum of the unimodal responses (Au + Vi), and of the multimodal responses (AV) at this latency; (3), the topography (potential and SCD distributions) of the interaction pattern quantified in the difference [AV - (Au + Vi)] between the multimodal responses and the sum of the unimodal responses. Right column: Student t map estimated on potential values; the gray color indicates the scalp areas where [AV - (Au + Vi)] amplitude differs significantly from zero at the probability level indicated below the map. In potential and SCD maps, half range of the scale (µV or µA/m²) is given below each map.

Comparing the interaction patterns with the corresponding unimodal responses, it may be seen that: (a) and (b) within 40-90 and 90-145 msec, the patterns of interaction at occipital sites do not correspond to any activation in the unimodal visual response, suggesting that the interactions are due to new activities in the visual cortex elicited only by the conjunction of auditory and visual features in the stimulus. (c) Between 155 and 200 msec, the unimodal visual response (N185) and the interaction pattern display similar topographies (of opposite polarity), suggesting that the interactions are due to a decreased activity of the unimodal N185 generators in visual cortex. (d) Around 95 msec, both the unimodal auditory response (N1 wave) and the interaction pattern present mastoid positivity and fronto-central negativity, a typical pattern of auditory cortex activity; this suggests that the interactions are due to an increased activity of N1 generators in the auditory cortex. (e) Within 150 to 185 msec, the interactions may be due either to an increase in the unimodal auditory P2 amplitude or to new activities of unknown origin. (f) Around 140 to 165 msec, interactions are expressed by highly significant activities over the right fronto-temporal areas that do not exist in either unimodal responses, suggesting new neural activities in the anterior part of the right temporal lobe or in the right prefrontal cortex.
significantly higher over the right than the left hemisscalp ($p < 0.006$). SCD maps showed similar spatio-temporal evolution, with two distinct current sources around PO4- T6 and T5, respectively, maximum around 185 msec (peak amplitudes: 0.83 and 0.78 mA/m$^3$, respectively) (Figure 6c, right part).

Both the peak latency and topography of this interaction pattern corresponded to those of the visual N185 wave in unimodal response (Figure 6c, second and third columns). This strongly suggests that the interaction pattern in this latency range reflects a decrease in amplitude of the N185 wave elicited by unimodal visual stimuli.

**Interactions in Temporal and Central Areas**

To know whether auditory-visual interactions could occur in the auditory cortex, [AV − (Au + Vi)] amplitude and topography were examined at scalp areas usually responding to auditory stimulus (mastoid and fronto-central electrodes). Two interactions patterns were found.

**90 to 105 msec (Figure 6d)** In this latency range, the positive amplitudes of [AV − (Au + Vi)] observed at posterior electrodes (see above) extended to the mastoid electrodes. Significant positive amplitudes ($p < 0.05$) were found at the right mastoid from 79 to 114 msec (mean amplitude: 0.64 µV) and at the left mastoid from 94 to 116 msec (mean amplitude: 0.55 µV); statistical significance reached 0.001 between 92 and 106 msec at M2 and 0.01 between 98 and 112 msec at M1. This mastoid positivity was associated with negative amplitudes over the frontal areas, although the frontal negativity tended only to statistical significance ($p < 0.10$ between 83 and 98 msec, $p < 0.05$ between 86 and 92 msec at FT4 and F4). The corresponding SCD distribution of [AV − (Au + Vi)] similarly showed a polarity reversal between positive current sources peaking around the mastoids at 90-msec latency (0.44 and 0.30 µA/m$^3$ at M1 and M2, respectively) and negative current sinks at fronto-temporal electrodes (−0.39 and −0.37 µA/m$^3$ at FT3 and FT4, respectively) (Figure 6d, right part). (Note the additional current sources around the occipito-parietal electrodes, as described earlier.) Although the frontal negativity of [AV − (Au + Vi)] was only marginally significant, the polarity reversal of this interaction pattern in temporal regions was similar to that of the unimodal auditory N1 wave in the same latency range (Figure 6d, first column), known to be generated in the auditory cortex (the interaction pattern including additional activity over the posterior area). This suggests that the conjunction of auditory and visual features in the stimulus may enhance the activity of at least some generators of the N1 response in auditory cortex, parallel to the activity enhancement in visual areas.

**150 to 185 msec (Figure 6e)** A significant interaction pattern was found between 150- and 185-msec latency at the three left fronto-central electrodes Cz, C3, FC1 (mean amplitude over Cz-C3-FC1 and 150 to 185 msec: 1.04 µV; $p < 0.05$ at all three electrodes between 150 and 185 msec; significance reached 0.01 at C3 and FC1 from 155 to 185 msec). The pattern was maximum at 175 to 180 msec (peak amplitude: 1.4 µV at FC1). SCD distribu-

---

**Table 2.** Mean amplitude ± standard deviation (in µV) of [AV − (Au + Vi)] averaged in time and space over the periods and electrodes indicated at the top of each column and compared against zero (student $t$ test). The mean amplitudes and significance levels are given for all subjects and for AUD-subjects and VIS-subjects separately. Last line: Significance level in the comparison of AUD-subjects and VIS-subjects amplitudes (Student $t$-tests).

<table>
<thead>
<tr>
<th>Temporo-Parieto-Occipital</th>
<th>Temporal</th>
<th>Central</th>
<th>Right Frontal</th>
</tr>
</thead>
<tbody>
<tr>
<td>O2-PO4</td>
<td>O1-02-Pz</td>
<td>O2-PO4-T6</td>
<td>M2</td>
</tr>
<tr>
<td>All subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 24)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUD-subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VIS-subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 14)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUD-subjects vs VIS-subjects</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Mean Amplitudes

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>O2-PO4</td>
<td>0.90 ± 0.95</td>
<td>0.74 ± 1.24</td>
<td>1.49 ± 1.58</td>
<td>0.76 ± 0.87</td>
<td>−0.34 ± 1.14</td>
<td>1.04 ± 1.44</td>
</tr>
<tr>
<td>(n = 24)</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>***</td>
<td>**</td>
<td>***</td>
</tr>
<tr>
<td>AUD-subjects</td>
<td>1.34 ± 0.80</td>
<td>1.54 ± 1.07</td>
<td>1.63 ± 1.73</td>
<td>0.77 ± 0.81</td>
<td>0.13 ± 1.03</td>
<td>1.07 ± 1.36</td>
</tr>
<tr>
<td>(n = 10)</td>
<td>***</td>
<td>**(*)</td>
<td>(*)</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>VIS-subjects</td>
<td>0.57 ± 0.95</td>
<td>0.17 ± 1.04</td>
<td>1.39 ± 1.53</td>
<td>0.76 ± 0.94</td>
<td>−0.69 ± 1.13</td>
<td>1.03 ± 1.54</td>
</tr>
<tr>
<td>(n = 14)</td>
<td>*</td>
<td>n.s.</td>
<td>**</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>AUD-subjects vs VIS-subjects</td>
<td>*</td>
<td>**</td>
<td>n.s.</td>
<td>n.s.</td>
<td>(*)</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

### Significance Levels

- *** = $p < 0.001$; ** = $p < 0.01$; * = $p < 0.05$; (*) = $p < 0.10$; n.s. = nonsignificant.
tions similarly showed a current source around FC1 (Figure 6e, right part), stable over the whole period. Although the peak latency of this interaction pattern roughly corresponded to that of the auditory P2 wave in unimodal responses, its scalp distribution was slightly more anterior (Figure 6e, first and fifth columns). It is therefore uncertain whether this interaction pattern can be interpreted as an increased activity of auditory P2 generator(s) or as a specific neural activation.

Interactions on Right Fronto-Temporal Area (140 to 165 msec)

A highly significant interaction pattern was found at the right fronto-temporal site around electrodes T4 and FT4 (Figure 6f, right part). There was no corresponding pattern on the left hemisphere. The effect began around 125 msec, reached a statistical significance of 0.01 between 140 and 165 msec and 0.001 between 145 and 160 msec at both T4 and FT4 (mean amplitude over 145 to 160 msec: −0.97 μV). It could be observed in almost all individual subjects. Neither the latency nor the topography of this interaction pattern corresponded to any ERP in unimodal responses, suggesting that this effect was due to new neural activities in the anterior part of the right temporal lobe or in the right prefrontal cortex.

Analysis of Auditory-Visual Interactions Separately in VIS-Subjects and in AUD-Subjects

If now we examine the multisensory interactions separately for subjects better at recognizing the objects on the basis of their auditory than visual features in unimodal presentation (AUD-subjects, n = 10) and for subjects better at visual than auditory object recognition (VIS-subjects, n = 14), the results are somewhat different. To easily clarify the differences, (1) Table 2 summarizes the results obtained over all subjects and presents the corresponding data separately for AUD-subjects and VIS-subjects, and (2) Figures 7 and 8 present the data corresponding to Figures 5 and 6 separately for AUD-subjects and VIS-subjects.

Effects on Posterior Visual Areas

40 to 90 msec (Figure 8a) Comparison of Figures 5 and 7 shows that most of the significant effects found at occipito-temporal sites between 40 and 90 msec when considering all subjects were due to AUD-subjects. In VIS-subjects, effects were not significant on the left hemisscalp and only scarcely significant on the right. Instead, AUD-subjects presented significant effects at occipito-temporal electrodes of both hemispheres, with the strongest effects on the right hemisscalp (p < 0.01 at O2
and PO4 over the entire period, 40 to 90 msec). This result is also illustrated in Figure 8a and in Table 2, which shows that the mean amplitude of the interaction pattern is significantly larger in AUD-subjects than in VIS-subjects (1.34 μV vs. 0.57 μV, \( p < 0.05 \)). Examination of the individual patterns suggests that AUD-subjects globally presented effects of more homogeneous amplitude than VIS-subjects: Out of the ten AUD-subjects, one had no interaction effect in visual areas and the remaining nine had pattern amplitudes ranging from 0.97 to 2.07 μV, whereas in VIS-subjects, three had no effect and eleven presented amplitudes from 0.1 to 1.82 μV.

90 to 145 msec (Figure 8b) The same differences between AUD- and VIS-subjects were found within 90 to 145 msec: VIS-subjects presented only little significant effects (Figure 7), whereas AUD-subjects still showed highly significant interaction patterns (\( p < 0.01 \) or \( p < 0.001 \) at most occipito-temporal electrodes; see Figure 7). Accordingly, Table 2 shows that the mean amplitude of [AV - (Au + Vi)] over 90 to 145 msec is highly significant for AUD-subjects (1.54 μV, \( p < 0.005 \)) but not significant for VIS-subjects (0.17 μV), the difference between AUD- and VIS-subjects being statistically significant (\( p < 0.01 \)). The interaction pattern was present in all but one AUD-subject (amplitudes ranging from 0.35 to 3.05 μV), whereas it appeared in only seven out of fourteen VIS-subjects (with amplitudes from 0.52 to 1.41 μV). The topographic differences between AUD- and VIS-subjects are illustrated in Figure 8b.

155 to 200 msec (Figure 8c) In the latency range of the N185 wave elicited by unimodal visual stimulus, AUD- and VIS-subjects showed similar significant interaction patterns composed of bilateral positive potential fields with the largest amplitudes over the right hemisphere (Figure 8c). Although VIS-subjects presented more significant effects around 185 msec (Figure 7), the difference between AUD- and VIS-subjects over the whole 155- to 200-msec period was not significant (Table 2), indicating that in both AUD- and VIS-subjects, auditory-visual interaction decreases the amplitude of the unimodal visual N185 wave.
Effects on Temporal and Central Areas

90 to 105 msec (Figure 8d) Although around the latency range of the unimodal auditory N1 wave both AUD- and VIS-subjects presented significant positive [AV = (Au + Vi)] amplitudes around the mastoids, only in VIS-subjects was this mastoid positivity associated with significant negative potentials at fronto-central sites (Table 2). Accordingly, Figure 8d shows that only VIS-subjects displayed an interaction pattern with potential distributions typical of that of the auditory N1 (positive potentials around the mastoids with polarity reversal along a line roughly delineating the sylvian fissure), compatible with an origin in the supratemporal plane of the auditory cortex (Vaughan & Ritter, 1970). Instead, in AUD-subjects, the interaction pattern around these latencies was dominated by posterior activities at temporo-occipital sites, as described earlier (Figure 8d). These results strongly suggest that the conjunction of auditory and visual features in the stimulus enhances the activity of unimodal N1 generators in the supratemporal plane of the auditory cortex only in VIS-subjects.

150 to 185 msec (Figure 8e) Table 2 shows that the interaction effects around 150 to 185 msec and Cz-C3-FC1 have the same significant amplitudes in AUD-subjects (1.07 μV) and in VIS-subjects (1.03 μV). Similarly, comparison of Figures 6e and 8e shows that the positive potential fields centered around Cz-C3-FC1 in SP maps for all subjects are found for both AUD- and VIS-subjects.

Effects on Right Fronto-Temporal Areas (140 to 165 msec)

Here again, the strongly significant interaction patterns found for all subjects around T4-FT4 were still observed in both AUD- and VIS-subjects with equal mean amplitude (~0.91 and ~1.02 μV) and with similar scalp distributions (Table 2 and Figure 8f), suggesting the existence of a robust interaction pattern in the anterior part of the right temporal lobe or in the right inferior frontal gyrus for all the subjects.

DISCUSSION

This experiment was designed to provide evidence for multisensory integration in a task of multimodal object recognition in humans and to characterize the dynamics and neural correlates of the integrative processes induced. Three main effects were found: (1) the recognition of objects defined by the combination of auditory and visual features was improved compared to that of the same objects defined by unimodal features alone, indicating that auditory-visual information did interact during multimodal object identification. (2) These interactions began at early stage of sensory analysis (40 msec) and were expressed by several neural events of different natures within the first 200 msec poststimulus, involving several sites in both sensory-specific and nonspecific cortices; this indicates that multisensory integration is not mediated by a unitary physiological mechanism but results from multiple neural operations. (3) These operations partly depended on the dominant modality of the subject to perform the task in unimodal conditions; this shows that multisensory integrative processes are not rigid but rather form a highly adaptive system tuned to individual sensory/cognitive skill.

Multisensory Integration in Object Recognition

As emphasized earlier, most of the behavioral studies in humans and in animals have focused on the spatial aspects of multisensory integration and on the facilitation effects for detecting, localizing, and/or reacting to stimuli that contain redundant information. The present experiment provides behavioral evidence for multisensory integration in the genuine perceptual process of multimodal pattern identification. One may note that the visual cues were formed by a movement (deformation of a circle). If one excepts the phenomena related to the “McGurk effects” (misperception of a phoneme due to a conflict between the lips movement and the sound emitted (McGurk & MacDonald, 1976; Sams et al., 1991), or the family of the ventriloquism effects, the perceptual formation of a unitary “object” associating a movement and a sound has been very little documented in humans (see, however, Sekuler et al., 1997). Yet, because its natural frequency and ecological importance in everyday life, the ability to associate (and integrate) a movement/deformation and a sound must be strongly developed across all species.

Auditory-Visual Interactions in Multiple Cortical Sites within 200 msec Poststimulation

The ERP events specifically due to the conjunction of auditory and visual cues in the stimulus were characterized, before 200 msec, in the differences between the responses to multimodal stimuli and the sum of the unimodal responses. These difference waveforms revealed several multimodal interaction patterns, temporally, spatially, and functionally distinct. Each of these interaction components will be examined and discussed below.

Effects in Visual Cortical Areas

Most unexpected was the observation of an interaction pattern as early as 40 msec poststimulus at the right occipito-parietal site. The 40- to 90-msec pattern did not correspond to any ERP event in the unimodal responses. Yet, its latency and topography were approximately simi-
lar to those of a visual ERP component (C1) elicited by flashed checkerboards and thought to be generated in the striate cortex (Clark, Fan, & Hillyard, 1995). The C1 component, however, reflecting activity in the retinotopically organized calcarine cortex, would have displayed a negative polarity for stimuli presented along the horizontal meridian (as in the present experiment) (Clark et al., 1995). In addition, C1 appears to be insensitive to such factors as spatial attention (Clark & Hillyard, 1996) or attention to color or movement (Anllo-Vento & Hillyard, 1996). It is therefore unlikely that the 40- to 90-msec interaction pattern, of positive polarity, could reflect the modulation of a weak C1 component by a mechanism of “sensory gain” [similar to the effect of spatial attention on later P1 and N1 visual components (Luck, Heinze, Mangun, & Hillyard, 1990; Mangun, 1995)]. Rather, it could be due to the recruitment of neurons not activated by the sole visual inputs in or near the striate cortex. Such findings have been reported in animals: auditory inputs were found to activate bimodal neurons in areas 17, 18, and 19 of the cat visual cortex (Morrell, 1972; Fishman & Michael, 1973) and between areas 17 and 18a of the rat visual cortex (Barth et al., 1995).

A second interaction effect was observed in visual areas around 90 to 145 msec, with a neural pattern corresponding in latency, polarity, and topography to the P1 component of visual ERPs. P1 is thought to be generated in the extrastriate cortex (Clark et al., 1995; Heinze et al., 1994) and is sensitive to subjective factors as the spatial focus of attention (Luck et al., 1990; Clark & Hillyard, 1996). However, although P1 is usually elicited by a wide variety of visual stimuli, no P1-like activity could be observed here in the unimodal visual responses. This could be explained by the weak salience of the circle deformation, compared to the “stronger” stimuli generally used in ERP experiments. It would be therefore possible that the 90- to 145-msec interaction pattern represents an increased activity of the P1 generators in the extrastriate cortex, a neural effect of stimulus salience enhancement due to the addition of an auditory cue in the stimulus. Alternatively, however, the hypothesis of the recruitment of new neurons in the extrastriate cortex cannot be ruled out. In any case, the interaction effects in this latency range represent an increased activity in the visual (extrastriate) cortex and suggest, like the early 40- to 90-msec effects, that pairing a weak auditory cue to a subthreshold visual stimulus (in terms of visual evoked response) may lead to suprathreshold effects in visual cortex.

The ambiguity of the nature of the auditory-visual interactions in visual cortex below 150 msec is greatly reduced (if existing) for the later effects around 155 to 200 msec. Here the neural pattern of interaction has the same latency and topographic distribution as the N185 (N1) component of the ERPs to unimodal visual stimuli and may therefore represent a modulation of activity within the same neural generators in the extrastriate cortex (Mangun, 1995). Interestingly, the effect is a decrease of the unimodal visual response, a finding similar to MEG results (Busch et al., 1989) and consistent with the “depression” effect reported in early multimodal ERP studies (e.g., Davis, Osterhammel, Wier, & Gjerdingen, 1972). A possible interpretation is that the visual processing of a cue made more salient (by adding an auditory feature) requires less energy from the visual system (smaller N1 amplitude) than that necessary for processing a more difficult cue. Indeed, although the attentional level is probably similar during the processing of unimodal and multimodal objects, the energy required in both cases may be different. The apparent contradiction between an increased activity in the P1 latency range and a decrease in N1 amplitude may be resolved if one keeps in mind that P1 and N1 are functionally distinct and react differently to visual attention (Luck et al., 1990), and therefore probably also to energy demand. Although spatial attention enhances both P1 and N1 amplitudes, the enhancement of the two components has been interpreted differently: A P1 increase would reflect a facilitation of sensory processing of items at attended location (Harter & Aine, 1984), whereas an N1 enhancement would correlate with “the engagement or orienting of attention to the relevant location” (Luck et al., 1990), that is, with an energetic demand. A decrease in N1 amplitude during multimodal stimulus processing could therefore express, as the increased activity in P1 latency range, a facilitation of processing within the extrastriate cortex, thereby providing electrophysiological support to the “facilitation effect” reported in numerous behavioral studies.

**Effects in the Auditory Cortex**

The strongest evidence for multisensory interactions in the auditory cortex relies on the similarity between the topographic characteristics of the unimodal auditory N1 (polarity reversal between the mastoids and the fronto-central sites) and those of the \( [AV − (Au + Vi)] \) difference waveform around the same latency (95 msec), although the fronto-central negativity is of relatively weaker amplitude and significance in the interaction wave than in the auditory N1. N1 is known, however, to result from the activity of at least two dipolar neural sources in the auditory cortex, one “tangentially oriented” generated in the supratemporal plane and primarily responsible for the fronto-mastoid polarity reversal (Vaughan & Ritter, 1970) and another “radially oriented,” probably originating in the superior temporal gyrus (Scherg & von Cramon, 1986; Wolpaw & Penry, 1975). The stronger amplitude increase at mastoids than at the associated fronto-central sites could therefore be explained either by different enhancement of activity for the two N1 generators in the auditory cortex (with the largest effect on the radial generator) or by the recruitment of new neurons not activated by auditory inputs.
alone in the superior temporal gyrus (or by both mechanisms). Effects of multimodal integration in the auditory association cortex fit with PET findings in humans that have shown increased blood flow activation in the right superior temporal gyrus (area 22) in synaesthete subjects (Paulesu et al., 1995) and with animal studies that have provided evidence for multisensory zones in the extra-primary auditory cortex of monkeys (Watanabe & Iwai, 1991), cats (Schroeder, Mehta, Ulbert, Steinmetz, & Vaughan, 1995), and rats (Barth et al., 1995).

It should be stressed, however, that the mastoid electrodes, placed above the posterior temporal gyrus, may pick up neural activity from a variety of structures within the ventral visual system (Ungerleider & Mishkin, 1982), known to be a locus of visual integration (review in Logothetis & Sh猩enberg, 1996). It is therefore possible that part of the amplitude increase at mastoids is due to interactions in temporal visual areas that overlap genuine effects in the auditory cortex.

Whatever the mechanisms, the interactions in the auditory cortex are, like those in the visual cortex, larger in the right than in the left hemisphere but are, across all subjects, of lower magnitude than the effects in visual cortex.

**Effect at Right Fronto-Temporal Sites**

One of the most robust interaction patterns was that observed unilaterally around 145 to 160 msec over the T4-FT4 site on the right hemiscape. This site was not activated by any of the unimodal stimuli and can therefore be seen as a locus of polysensory interactions. The spatial resolution of scalp ERPs does not allow us to know which structures are involved. Likely candidates are the anterior part of the right temporal lobe or the inferior lateral prefrontal cortex (or both). Both cortical regions have been described as integration sites: human PET data have shown the activation of several areas of the right prefrontal cortex in synaesthesia (Paulesu et al., 1995; Walsh, 1996); in monkeys, sites of multimodal convergence have been found both in the prefrontal cortex (Benevento, Fallon, Davis, & Rezak, 1977; Gaffan & Harrison, 1991; Ito, 1982; Vaadia, Benson, Heinz, & Golstein, 1986), and in the temporopolar cortex (Jones & Powell, 1970). The location of the right fronto-temporal interaction component in humans remains therefore to be determined.

**Right Hemisphere Dominance**

Our results emphasize the strong predominance of the right hemisphere in multisensory perceptual integration. The neural interactions were generally larger in the right hemisphere whatever the cortical sites involved (sensory and nonsensory-specific cortices). The hypothesis of a cortical network operating predominantly in the right hemisphere in multimodal perceptual integration is further supported by PET findings (activation of the right prefrontal cortex in synaesthesia: Paulesu et al., 1995) and by MEG data (activation of the right parieto-temporal regions: Sams & Imada, 1997).

**Different Mechanisms of Integration in Auditory-Dominant and in Visually Dominant Subjects**

A most striking aspect of our data concerns the differences in the cortical operations induced by multisensory integration in VIS- and AUD-subjects at early stage of sensory processing (below 150 msec). In VIS-subjects (subjects better in visual than auditory modality for unimodal object recognition), adding an auditory cue to the visual stimulus did not enhance (or enhanced very little) the level of neural activity in the visual cortex, but adding a visual cue to the auditory stimulus increased the activity in the auditory cortex. Conversely, in AUD-subjects (better in auditory than in visual task), adding a visual cue to the auditory stimulus did not change the activity in the auditory cortex, but adding an auditory cue to the visual stimulus clearly increased the level of activity in the visual cortex. More simply, visually dominant subjects showed only little early integration effects in the visual cortex, but they displayed clear effects in auditory cortex; auditory-dominant subjects, however, showed early integration effects in the visual cortex but no effect in the auditory cortex. In other words, at early stage of sensory analysis (below about 150 msec), multisensory integration induced enhanced neural activity predominantly—if not exclusively—in the cortex of the nondominant sensory modality.

These observations may be directly related to the “inverse effectiveness principle” of multisensory integration defined by Stein’s group (review in Stein & Meredith, 1993). This principle postulates that “the more effective the unimodal stimuli [are], the lower the magnitude of the enhancement they [are] capable of generating by combination” (Stein et al., 1994, p. 97). This rule has been shown to apply both at behavioral level in humans and in animals (Stein et al., 1989, 1996; Welch & Warren, 1986), and at single cell level in animals [e.g., in the superior colliculus of cats and monkeys (Meredith & Stein, 1986b; Wallace et al., 1996) and in the cat cortex (Wallace et al., 1992)]. Our results show that this “inverse effectiveness rule” also applies at the global, integrated level of sensory cortex activity in humans. Furthermore, they show that multisensory enhancement depends on the subjective effectiveness of the unimodal cues, because the same physical stimuli were used for all the subjects.

It is interesting to note that the effects of multisensory integration, different for AUD- and VIS-subjects in sensory cortices, did not differ between the two groups at right fronto-temporal sites. This strongly suggests that the different cortical loci of multisensory convergence do not
subserve the same integrative functions. Although the differential neural enhancements in sensory cortices may reflect exquisitely fine and flexible mechanisms that facilitate the processing of the least-efficient cues, the fronto-temporal sites could carry out more general integration functions.

In any case, these results show that the multisensory integrative processes are not rigid but rather form a highly adaptive system that is dynamically tuned to individual sensory/cognitive skill.

Concluding Remarks

First, the experimental paradigm (multimodal object recognition task) and the ERP technique used (sensitive to cortical activities) have allowed us to characterize cortical operations involved in multisensory perceptual integration. We ignore at present the relationships between this cortical network in humans and the neural systems involving the superior colliculus described in animal studies. Although the existence of independent integration systems has been suggested (Hughes et al., 1994; Wallace et al., 1993), further studies should bridge the gap between these neural integrative systems.

Second, the results above provide definite electrophysiological evidence for the coactivation model in redundant information processing (Miller 1982, 1986) in showing that bimodal information may interact from very early, and at multiple points, in the sensory processing chain. It should be interesting to compare these integration effects with those possibly generated in conditions of nonredundant information, that is, in situations when both auditory and visual informations are necessary to perform the recognition task.

METHOD

Subjects

Twenty-four right-handed subjects (13 female, mean age: 22) participated. All were free of neurological disease, had normal hearing, and normal or corrected to normal vision.

Stimuli

Two "objects" A and B were devised, each of them defined by either visual features alone, auditory features alone, or the conjunction of visual and auditory features. Object A consisted of either the deformation of a circle to an horizontal ellipse (stimulus V1) a tone burst of 540 Hz (stimulus A1), or the conjunction of the two features (stimulus AV1). Object B was designed similarly, as the deformation of a circle to a vertical ellipse (stimulus V2), a tone burst of 560 Hz (stimulus A2), or the conjunction of the two features (stimulus AV2).

The circle had a diameter of 5 cm and was presented permanently on a dark video monitor placed 130 cm in front of the subject's eyes, subtending a visual angle of 2.2°. The ellipses were formed by a 10% modification of the length of the horizontal and vertical diameters of the circle. The duration of the deformation was 223 msec. The tone bursts had same duration (including 10 msec of rise/fall times) and were delivered through a loud-speaker placed behind the video monitor with an intensity of 50 dB HL. For stimuli AV1 and AV2, the circle deformation and the tone burst were presented synchronously.

Procedure

Subjects were seated in a dark, sound-attenuating room and were given instructions describing the task along with 2 practice blocks. Then subjects underwent 10 blocks of recording. A block started with the presentation of the circle on the video monitor, the center of which served as a fixation point during the whole block. Each block consisted in at least 72 trials composed of 12 repetitions of the six stimulus types (A1, V1, AV1, A2, V2, and AV2). Within a block the stimuli were delivered randomly with the constraint that two stimuli sharing the same visual or acoustic parameters could not occur in sequence. The subjects were required to press a key with their right index finger on identifying object A (stimulus A1, V1, or AV1) and with their right middle finger on identifying object B (stimulus A2, V2, or AV2). Speed and accuracy were equally stressed. Stimuli were presented 1350 msec after the response to the previous stimulus if the reaction time was shorter than 1500 msec or 2850 msec after the onset of the previous stimulus in the other cases. Responses with latencies of greater than 1500 msec were excluded from all analyses. At the end of each block, the percentage of correct responses was displayed on the video monitor.

ERP Recording

An electroencephalogram (EEG) continuously recorded at a sampling rate of 1 kHz (0.1 to 320 Hz analog bandwidth) from 30 Ag-AgCl scalp electrodes referred to the nose. Eighteen were placed according to the international 10-20 system (Fz, Cz, Pz, Iz, Fp1, F3, C3, P3, T3, T5, O1 and their counterparts on the right hemiscalp) and two at the left and right mastoids (M1 and M2). The remaining electrodes were placed midway between two positions of the 10-20 system: FT3 (between F3 and T3), FC1 (between F3 and Cz), TP3 (between T3 and P3), PO3 (between O1 and P3), IM1 (between Iz and M1), and their counterparts on the right hemiscalp. A computer system based on a three-dimensional digitization of the head was used to facilitate accurate electrode positioning (Echallier, Perrin, & Pernier, 1992). 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 3 kΩ. ERPs were averaged off-line separately for each stimulus type, over a time period of 1100 msec including a 100-msec prestimulus baseline. Signals including incorrect responses or exceeding 100 μV at any electrodes were excluded from averaging. In this way, the trials contaminated by eye movements were not taken into account in subsequent analyses. Finally, ERPs were digitally filtered with a zero-phase shift low-pass filter (0 to 30 Hz) to eliminate high-frequency residual noise.

Data Analysis

Scalp potential maps were generated using a two-dimensional spherical spline interpolation and a radial projection from Oz (back views), from T4 (lateral views) or from Cz (top views), which respects the length of the meridian arcs. SCDs were estimated by computing the second derivative of the interpolated potential distributions (Perrin, Pernier, Bertrand, & Echallier, 1989). SCD maps are reference free and have sharper peaks than those of the potential distributions, which may facilitate the interpretation in case of multiple overlapping sources. In addition, SCDs enhance the contribution on the scalp of shallow, cortical generators compared to deeper sources.

Measurement of the Auditory-Visual Interactions

We assumed that at an early stage of stimulus analysis, the neural activities induced by the multimodal (AV) stimulus were equal to the sum of the neural activities induced separately by the auditory (Au) and the visual (Vi) stimulus, plus the putative neuronal activities induced uniquely by multimodal stimulation (auditory-visual interactions). This assumption is valid only while the stimulus analysis is not “contaminated” by late activities related to target processing (P3 waves) or by activities related to the response selection or motor processes (all these activities being common to all three stimulus types Au, Vi, and AV). The present data, in accordance with those generally observed in ERP studies, show that below about 200 msec, the unimodal ERPs reflect sensory analysis in modality-specific areas. Over this period, we may therefore use the summative model to estimate the AV interactions:

\[
\text{ERP (AV)} = \text{ERP (Au)} + \text{ERP (Vi)} + \text{ERP (Au × Vi interactions)}
\]

This expression is valid whatever the nature or configuration of the intracerebral generators and is based on the law of superposition of electric fields. Auditory-visual interactions were therefore quantified in the difference wave \([\text{AV} - \text{(Au + Vi)}]\) obtained by subtracting the sum of the responses to the unimodal stimuli from the response to multimodal stimuli. Significant effects were assessed by Student t tests (amplitude of the difference wave compared to zero), computed for each time sample at each electrode. Student t maps could then be obtained at each latency. Where considered as significant interaction effects, those spatio-temporal patterns that had a stable topography with a significant amplitude \((p < 0.01)\) at at least one electrode for 15 consecutive samples (15 msec) (Rugg, Doyle, & Wells, 1995).

“Mean amplitude” represents the signal amplitude averaged in space over the electrodes specified and in time over the entire period of the interaction pattern considered. “Peak amplitude” gives the maximum value within the spatio-temporal window considered.

Acknowledgments

We are grateful to J. Pernier for helpful discussions and to J. F. Echallier and P. E. Aguera for technical assistance.

Reprint requests should be sent to Marie-Hélène Giard, INSERM-U280, 151 Cours Albert Thomas, F-69424 Lyon Cedex 03, France, or via e-mail: giard@lyon151.inserm.fr.

Notes

1. In a pilot behavioral study using the same paradigm, we had compared the subjects' performances in two conditions: one in which the visual stimulus consisted in the apparition of an ellipse on the computer screen, and the other where the ellipse was formed by the deformation of a pre-existent circle (like in the present study). It was found that the integration effects were larger (shorter RT) in the deformation condition than in the apparition condition.
2. At later latencies (> 180 msec), the amplitude decrease in the N1 visual response was observed for both AUD- and VIS-subjects. As already discussed, this would mean that the combination of auditory and visual cues decreases the energy demand in the visual system for all subjects, this energy reduction being a little more pronounced for VIS-subjects than for AUD-subjects (see Figure 7). This global reduction in visual energy makes sense when considering the general dominance of the visual modality in information processing (Posner, Nissen, & Klein, 1976).
3. An aspect of this question concerns the relationships between the polarity of the scalp ERP components and the inhibitory/excitatory processes at the single neuron level. Voltage changes recorded on the scalp are mainly due to synaptic currents in pyramidal cells. A positive potential at the scalp surface may be due either to active excitatory synapses located near the soma generating a current flow into the cells (current sinks) and inducing distributed current sources along the apical dendrites, or to active inhibitory synapses (current sources) on the apical dendrites. This last case is very unlikely, because inhibitory synapses on many neurons are found clustered on the soma. Conversely, a negative potential on the scalp surface may be due either to active excitatory synapses distributed on the apical dendrites or to inhibitory synapses near the soma.
REFERENCES


troencephalography and Clinical Neurophysiology, 28, 360–367.


