The time-course of intermodal binding between seeing and hearing affective information

Gilles Pourtois,[1,2] Beatrice de Gelder,[1,2,CA] Jean Vroomen,[1] Bruno Rossion[2] and Marc Crommelinck[2]

[1]Cognitive Neuroscience Laboratory, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands; [2]Laboratoire de Neurophysiologie, Université de Louvain, Bruxelles, Belgium

[CA,1]Corresponding Author and Address

Received 18 January 2000; accepted 16 February 2000

Acknowledgements: Thanks to S. Philippart for technical assistance during the process of the EEG and thanks to the participants for their patience and interest.

Intermodal binding between affective information that is seen as well as heard triggers a mandatory process of audiovisual integration. In order to track the time course of this audiovisual binding, event related brain potentials were recorded while subjects saw facial expression and concurrently heard auditory fragment. The results suggest that the combination of the two inputs is early in time (110 ms post-stimulus) and translates as a specific enhancement in amplitude of the auditory N1 component. These findings are compatible with previous functional neuroimaging results of audiovisual speech showing strong audiovisual interactions in auditory cortex in the form of magnetic response amplifications, as well as with electrophysiological studies demonstrating early audiovisual interactions (before 200 ms post-stimulus). Moreover, our results show that the informational content present in the two modalities plays a crucial role in triggering the intermodal binding process. *NeuroReport* 11:1329–1333 © 2000 Lippincott Williams & Wilkins.

Key words: Audiovisual interaction; ERP; Face expression; Intermodal binding; Inversion effect; Multimodal integration; Voice prosody

INTRODUCTION

In a natural habitat information is acquired continuously and simultaneously through the different sensory systems. As some of these inputs have the same distal source (such as the sight of a fire, but also the smell of smoke and the sensation of heat) it is reasonable to suppose that the organism should be able to bundle or bind information across sensory modalities and not only just within sensory modalities. For one such area where intermodal binding (IB) seems important, that of concurrently seeing and hearing affect, behavioural studies have shown that indeed intermodal binding takes place during perception [1–3]. In these experiments, audiovisual stimuli (i.e. facial expression combined with an affective voice fragment) are presented to subjects instructed to judge, dependent on the condition, the facial expression, the tone of the voice or both. Strong crossmodal biases are evidenced at the behavioural level by slower reaction times in incongruent situations between voice and face than in congruent situations. This merging of inputs does not await the outcome of separate modality specific decisions and is not under attentional control [4]. What could possibly be the neurophysiological correlates of this early binding between seeing and hearing affective information? The question has been raised for a case that is very similar, that of concurrently presented input from hearing and seeing speech. Recent neuroimaging studies [5,6] have shown a response enhancement of the magnetic signal in unimodal auditory cortex during combined auditory and visual stimulation. Nevertheless, no information is yet available regarding the moment in time this increase of activity in auditory cortex takes place.

In the present study, we used event related brain potentials (ERPs) to track the temporal course of audiovisual interaction and assess whether these interactions would translate as an enhancement of early auditory electrophysiological components. Two main hypotheses were addressed: (1) will IB manifest itself as an increase in amplitude of an early auditory component (like the auditory N1 component) when a facial expression is presented concurrently with a voice fragment providing two congruent expressions (i.e. angry voice and angry face); (2) if processing of facial expression is required for IB to occur, the effect should disappear in a control condition when presenting the same facial expression upside-down which substantially hinders face recognition [7,8].

MATERIALS AND METHODS

Subjects: Seven native Dutch-speaking right-handed subjects (four males; three females) with an average age of 27

years participated in the study. They were paid for their participation.

Stimuli: All stimuli consisted of combination of an auditory with a visual stimulus. Visual materials consisted of four faces from the Ekman-Friesen set [9] (male actor number 4 and female actor number 5 each presenting once an angry and once a sad expression). Mean size of the face was 8×12 cm. Mean luminance of the visual stimuli was $25\,cd/m^2$ and of the room and the background $< 1\,cd/m^2$. Construction of auditory materials started from three sentences spoken in an angry tone of voice by a male and female semi-professional actor. Only the last four syllables were used as test materials. The average sound level of the speech was 78 dB. Visual stimuli were then combined with auditory stimuli in order to construct six audiovisual trials (2 visual stimuli $\times$ 3 auditory stimuli) with either congruent or incongruent affective content. Moreover, congruous and incongruous inverted pairs (i.e. concurrent affective voice and inverted face stimulations) were constructed by rotating the orientation of the face 180° (upside-down). Gender between voice and face was always congruent. A trial started with the presentation of the face. After a variable delay (750–1250 ms) following the onset of the face, the voice fragment (duration 980 ± 216 ms) was presented via a loudspeaker. The face stayed on until the end of the voice fragment. The delay between voice and face onsets was introduced in order to reduce interference of the brain response elicited by the faces. Total duration of a trial was 2500 ms, inter-trial interval (measured from the offset of the visual stimulus) was randomly varied between 0.5 and 1 s.

Design and procedure: A total of 24 blocks (6 audiovisual trials $\times$ 2 congruencies $\times$ 2 orientations) of 70 audiovisual trials were randomly presented in an oddball paradigm. For each block, 60 trials served as standard (85%) and 10 trials as deviant (15%). Twelve blocks (six blocks with upright pairs and six blocks with inverted pairs) had congruous pairs as standard and incongruous pairs as deviant, and in the other 12 blocks, standard and deviant pairs were exchanged. The six congruous and six incongruous audiovisual pairs were each presented 140 times in random order. Subjects were tested in a dimly lit, electrically shielded room with the head restrained by a chin rest and 130 cm away from the screen fixating a central fixation point. Subjects were instructed to pay attention to the faces and ignore the auditory stimuli.

Electrophysiological recording and data processing: Visual event-related brain potentials (VEPs) and auditory event-related brain potentials (AEPs) were recorded and processed using a Neuroscan 64 channels. Horizontal EOG and vertical EOG were monitored using four facial bipolar electrodes placed on the outer canthi of the eyes and in the inferior and superior areas of the orbit. Scalp EEG was recorded from 58 electrodes mounted in an electrode cap (10-20 System) with a nose reference, and amplified with a gain of 30 K and bandpass filtered at 0.01–100 Hz. Impedance was kept below 5 kΩ. EEG and EOG were continuously acquired at a rate of 500 Hz. Epoching was made 100 ms prior to stimulus onset and continued for 924 ms after stimulus presentation. Data were low-pass filtered at 30 Hz. Maximum amplitudes and mean latencies of AEPs and VEPs were measured relative to a 100 ms pre-stimulus baseline and assessed using repeated measures analyses of variance (ANOVAs). Analyses were focused on early visual and auditory activities (250 ms post-stimulus).

RESULTS

In order to assess whether face orientation (upright *vs* inverted facial expressions) had been indeed processed and led to different early visual component [10–12], the brain responses time-locked to the presentation of the face were first analysed. Second, brain responses time-locked to the presentation of voice fragments concurrently presented with faces (upright *vs* inverted pairs) were analysed using several repeated measures ANOVAs both for amplitude and latency parameters of two early auditory components, the auditory N1 and P2 components (250 ms post-stimulus).

VEPs: When the analysis is time-locked to the presentation of the face, the visual N1 component (at C_z electrode) is first evaluated. This early visual component is conceptualized as the negative counterpart at the vertex of the occipital P1 component (P1-N1 complex) [13]. Recently, this component (PI) has been shown to be sensitive to visual affective processing (i.e. the valence of the stimulus) [14]. Following the N1 component, a specific brain response maximally evoked by facial stimuli [10,15] namely the vertex positive potential (VPP) is manifested by a positive deflection at the vertex (C_z) and occurring 180 ms post-stimulus. This component is sensitive to face orientation: inverted faces generally evoke a delayed and enhanced VPP [16]. The VPP can be considered as the positive counterpart of an occipito-temporal negativity (the N170 component), best recorded at electrodes T_5 and T_6 [11,16]. From the grand average waveforms comparing upright and inverted facial expressions (Fig. 1), no effect of orientation is evident in the N1 component but inverted faces evoked a delayed and higher VPP than normal faces

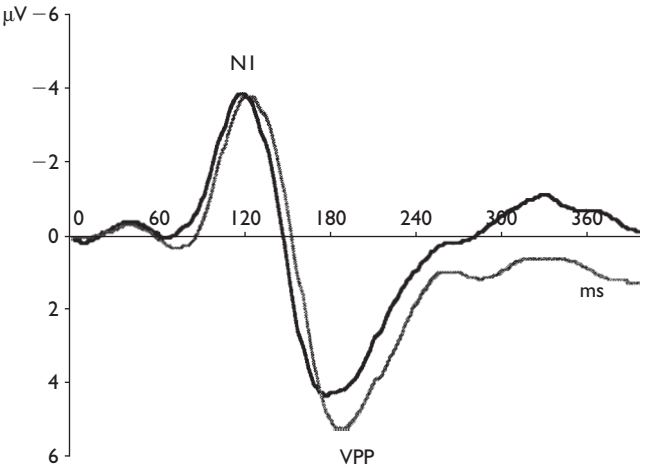

Fig. 1. Grand average waveforms (VEPs) at CZ electrode for upright facial expressions (black) and inverted facial expressions (grey).

(Table 1). These observations were confirmed by several statistical analyses (ANOVAs) computed on amplitude and latency parameters at C_z of the N1 and VPP with the factors orientation (upright *vs* inverted face) and affect (angry *vs* sad).

Considering the maximum amplitudes at C_z in the interval 80–120 ms (N1) there was no significant main effect or significant interaction. The maximum amplitudes at C_z electrode in the interval 160–200 ms (VPP) were entered into the same repeated measures ANOVA and the analysis revealed a significant effect of orientation ($F(1,6) = 14.64$, $p = 0.009$) in the sense that inverted faces elicited a larger VPP (mean amplitude 6.15 μV) than normal faces (mean amplitude 4.6 μV). Analysis of latencies at the C_z electrode corresponding to the maximum amplitudes in the interval 160–200 ms (VPP) revealed a significant effect of orientation ($F(1,6) = 6.82$, $p = 0.04$) in the sense that inverted faces elicited a delayed VPP (mean latency 191.7 ms) compared with normal faces (mean latency 176 ms).

AEPs: In order to assess the interactions between facial expression and voice, early auditory components were assessed. The N1 and P2 components are late cortical components [17], each composed of multiple subcomponents. The N1 has been shown to be modulated by auditory selective attention (i.e. enlarged N1 elicited by attended stimuli). Analysis of the waveforms comparing congruent and incongruent trials when upright faces are presented (Fig. 2) shows a strong amplitude effect on the N1 component in the sense that congruent trials trigger a higher N1 component than incongruent trials (Table 2) suggesting an amplification of the early auditory processing when congruent audiovisual pairs are present. Furthermore, this amplitude effect seems to be absent when inverted faces are presented (Fig. 2). Considering the P2 component, the orientation factor seems to interact with the latency parameter of this component in the sense that inverted pairs are delayed in comparison with upright pairs whatever the congruency of the pair.

These observations were confirmed by four repeated measures ANOVAs with the factors orientation (upright *vs* inverted pair), congruency (congruent *vs* incongruent pairs), anterior–posterior electrode position (frontal, central or parietal) and laterality (left, midline or right): two ANOVAs were carried out on the maximum amplitudes of two early auditory components (N1 and P2) and two other ANOVAs on the corresponding latencies of these peaks.

Table 1. Mean latency and amplitude of VPP for each subject for upright and inverted facial expression.

Subject	Upright		Inverted	
	Latency (ms)	Amplitude (μV)	Latency (ms)	Amplitude (μV)
1	178	7.62	186	10.88
2	162	4.73	176	6.47
3	196	4.69	214	7.31
4	174	0.3	210	0.81
5	170	4.46	186	5.2
6	186	3.7	174	4.24
7	166	6.72	196	8.16

The maximum amplitudes in the interval 90–130 ms (auditory N1 component) were first analysed. The analysis revealed a significant main effect of electrode position ($F(2,12) = 7.71$, $p = 0.007$), and of laterality ($F(2,12) = 8.53$, $p = 0.005$), a significant congruency × electrode position interaction ($F(2,12) = 7.04$, $p = 0.009$], a significant orientation × congruency × electrode position × laterality interaction ($F(4,24) = 3.25$, $p = 0.029$]. In order to explore the interaction between congruency and other factors, separate 2 (congruency) × 3 (electrode position) × 3 (laterality) repeated measures ANOVAs were computed for upright pairs and inverted pairs. For upright pairs, the analysis revealed a significant interaction between congruency × electrode position × laterality ($F(4,24) = 3.25$,) $p = 0.029$] and a significant main effect of electrode position ($F(2,12) = 11.14$, $p = 0.002$]. *Post-hoc* tests revealed that congruent pairs elicited a higher N1 component (at C_3: −6.038 μV) than incongruent pairs (at C_3: −5.271 μV) significantly at electrode C_3 ($F(1,6) = 8.32$, $p = 0.028$) and almost significantly at electrode P_3 ($F(1,6) = 5.95$, $p = 0.05$].

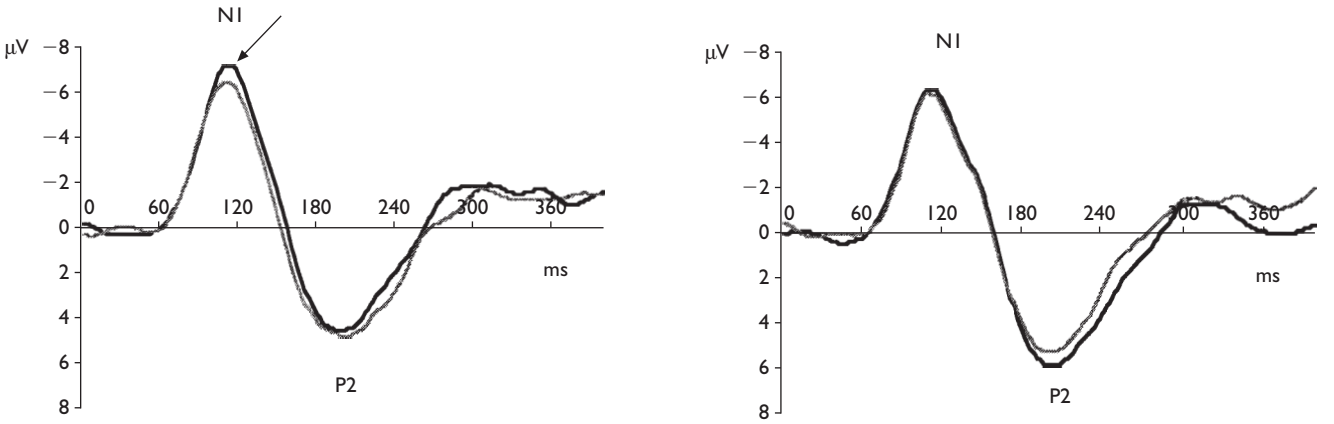

Fig. 2. (Left) grand average waveforms (AEPs) at Cz in the upright condition for congruent pairs (black) and incongruent pairs (grey), (Rigt) grand average waveforms (AEPs) at Cz in the inverted condition for congruent pairs (black) and incongruent pairs (grey).

Table 2. Mean amplitude (μV) of the auditory N1 component for the different conditions and locations.

		Upright			Inverted		
		Frontal	Central	Parietal	Frontal	Central	Parietal
Left	Congruent	−4.32	−6.04	−5.35	−3.83	−5.67	−4.88
	Incongruent	−4.0	−5.27	−4.29	−4.18	−5.35	−4.55
Midline	Congruent	−4.58	−7.27	−6.03	−4.85	−6.94	−5.47
	Incongruent	−4.91	−6.51	−4.59	−4.89	−6.67	−4.95
Right	Congruent	−4.38	−6.41	−5.44	−4.49	−6.21	−5.12
	Incongruent	−4.44	−6.44	−4.76	−4.68	−6.22	−4.56

For inverted pairs, the analysis revealed a significant main effect of electrode position ($F(2,12) = 5.36$, $p = 0.022$) in the sense that amplitudes are maximum at central leads, and a significant effect of laterality ($F(2,12) = 10.99$, $p = 0.002$) in the sense that amplitudes are maximum at mid-line electrodes.

The maximum amplitudes in the interval 180–220 ms (auditory P2 component) were then analysed. The analysis revealed a significant effect of electrode position ($F(2,12 = 7.53$, $p = 0.008$) and of laterality ($F(12 = 34.28$, $p < 0.001$), a significant congruency × electrode position interaction ($F(2,12) = 11.33$, $p = 0.002$) and a significant electrode position × laterality interaction [$F(4,24) = 4.128$, $p = 0.009$]. In order to explore the interaction between congruency and electrode position, separate 2 (congruency) × 3 (electrode position) × 3 (laterality) repeated measures ANOVAs were computed for upright pairs and inverted pairs. For upright pairs, the analysis revealed a significant congruency × electrode position interaction ($F(2,12) = 8.85$, $p = 0.004$), a significant electrode position × laterality interaction ($F(4,24) = 3.09$, $p = 0.035$), a significant main effect of electrode position ($F(2,12) = 8.78$, $p = 0.004$) and a significant main effect of laterality ($F(2,12) = 16.83$, $p < 0.001$). *Post-hoc* tests revealed that congruent pairs elicited a reduced P2 component (at $P_3 = 1.69\,\mu V$) compared with incongruent pairs (at $P_3 = 2.42\,\mu V$), significantly at electrode P_3 ($F(1,6) = 6.33$, $p = 0.046$). For inverted pairs, the analysis revealed a significant electrode position × laterality interaction [$F(4,24) = 3.68$, $p = 0.018$], a significant main effect of electrode position ($F(2,12) = 6.15$, $p = 0.015$) and a significant main effect of laterality ($F(2,12) = 25.64$, $p < 0.001$). *Post-hoc* tests revealed higher P2 amplitude at the left central electrode.

Analysis of latencies corresponding to the maximum amplitudes in the interval 90–130 ms revealed a significant main effect of orientation ($F(1,6) = 7.64$, $p = 0.033$) and a significant orientation × electrode position × laterality interaction [$F(4,24) = 3.37$, $p = 0.025$] in the sense that N1 latency was shortest at left central site. In order to explore the interaction between Orientation and other factors, separate 2 (congruency) × 3 (electrode position) × 3 (laterality) repeated measures ANOVAs were computed for upright pairs and inverted pairs. For upright and inverted pairs, the two analyses revealed no significant main effect nor interaction.

Finally, analysis of the latencies corresponding to the maximum amplitudes in the interval 180–220 ms revealed a significant effect of orientation ($F(1,6) = 11.52$, $p = 0.015$), indicating that inverted pairs (mean latency 200.29 ms) were delayed (upright pairs, mean latency 194.61 ms). Separate 2 (congruency) × 3 (electrode position) × 3 (laterality) repeated measures ANOVAs were computed for upright pairs and inverted pairs. For upright pairs, the analysis revealed no significant main effect nor interaction. There was a trend towards significance for electrode position × laterality ($F(4,24) = 2.58$, $p = 0.063$). For inverted pairs also, the analysis revealed no significant effect.

DISCUSSION

Our results clearly indicate that early auditory processing of a voice is modulated as early as 110 ms by the concurrent presentation of a facial expression. We have also provided evidence that IB occurs specifically when the facial expression is congruent and not when it is incongruent, or presented upside-down. Given the latter control conditions, an explanation in terms of acoustic differences cannot account for our results since exactly the same auditory fragments were used in the different conditions (congruent *vs* incongruent pairs; upright *vs* inverted faces).

When the analysis is time-locked to the presentation of the voice fragment in order to consider the auditory N1 component, the Congruency factor interacts mainly with the amplitude parameter of the AEPs only when upright faces are concurrently presented. This effect is manifested by an amplitude increase for congruent trials. This result supports the notion that auditory processing is enhanced when a congruous facial expression is concurrently presented. More precisely, the effect seems to be lateralized in the left hemisphere and is maximum at the central electrode position (C_3). The increase in the auditory N1 component found here points in the same direction to the fMRI results of Calvert *et al*. [5,6], showing significant magnetic signal enhancements in auditory cortex (BA 41/42) when audiovisual stimuli are presented. But the present results add crucial information regarding the moment in time these increases of activity indicative of IB take place by showing that increased activity in the auditory cortex is triggered as early as 110 ms post-stimulus. Moreover, this increase of activity is earlier than the magnetic wave (occurring 220 ms after the M100 wave) elicited in the auditory cortex when heard speech is combined with visible speech information, as evidenced by Sams *et al.* [18] using a different technique (magnetoencephalography) and a different methodology (an additive procedure). Increased activity only in the auditory cortex (auditory N1 component) when audiovisual processing is required has more

recently been reported by Giard and Peronnet [19] using the same technique (EEG) in a multimodal object recognition task. Finally, the orientation factor mainly plays a role on the latency parameter of the auditory processing (i.e. delayed auditory processing when inverted faces are concurrently associated), but there is no interaction with congruency, suggesting a global effect of inversion on the auditory processing.

When the analysis is time-locked to the presentation of the face, the results show that the brain responses elicited by the inverted faces evoked a delayed and larger VPP than normal faces, replicating previous observations [16]. Furthermore, there is no effect of facial expression (angry *vs* sad) before 180 ms post-stimulus at C_z, nor is there an interaction with face orientation. These results illustrate that subjects were indeed sensitive to face orientation. Our observations are compatible with previous electrophysiological studies that focused on the processing of face orientation [12] or on the processing of facial expression [20] suggesting that before 200 ms post-stimulus, some perceptual characteristics of the face (e.g. picture-plane orientation) are already processed, while others (e.g. facial expression) are not yet fully processed.

CONCLUSION

Seeing a facial expression while hearing an affective tone of voice leads to a mandatory process of audiovisual integration, as shown in previous behavioural studies [1]. Here, we used ERPs in order to clarify the neural correlates of this phenomenon. Our observations illustrate that an early process of IB is triggered when subjects see and hear affective information simultaneously. The results clearly suggest that the temporal course of audiovisual interactions is early (i.e. at the perceptual level) rather than late (i.e. at a decisional level). These results are compatible with previous functional neuroimaging results of audiovisual speech [6], showing strong audiovisual interactions in auditory cortex in the form of magnetic response amplifications, with electrophysiological studies demonstrating early audiovisual interactions before 200 ms post-stimulus [19,21] as well as with previous behavioural studies showing audiovisual bias characterized as automatic, perceptual and mandatory [1,2,22]. Moreover, while Stein and Meredith [23] have pointed out that audiovisual integration required spatial and temporal coincidence, our results clearly show that in order for IB to be triggered the informational content between the two modalities (i.e. accessible affective content) is equally important. Further studies should assess whether the present phenomenon of IB with affective information is equally compelling for different, more or less basic emotions (happiness, fear, disgust) and explore the basic properties of the temporal window needed to evidence early audiovisual interactions.

REFERENCES

1. de Gelder B and Vroomen J. *Cogn Emotion* (in press).
2. de Gelder B, Vroomen J and Bertelson P. *Cur Psychol Cog* **17**, 1021–1031 (1998).
3. Massaro DW and Egan PB. *Psycho Bul Rev* **3**, 215–221 (1996).
4. de Gelder B (1999). Recognizing emotions by ear and by eye. In: Lane R and Nadel L, *eds. Cognitive Neuroscience of Emotions*. Oxford: Oxford University Press, 1999: 84–105.
5. Calvert GA, Brammer MJ and Iversen SD. *Trends Cogn Sci* **2**, 247–260 (1998).
6. Calvert GA, Brammer MJ, Bullmore ET *et al. Neuroreport* **10**, 2619–2623 (1999).
7. de Gelder B, Teunisse JP and Benson PJ. *Cogn Emot* **11**, 1–23 (1997).
8. Searcy JH and Bartlett JC. *J Exp Psychol Hum Percept Perform* **22**, 904–915 (1996).
9. Ekman P and Friesen WV. *J Enviro Psychol Non-verbal Behav* **1**, 56–75 (1976).
10. Jeffreys DA. *Vis Cogn* **3**, 1–38 (1996).
11. Bentin S, Allison T, Puce A *et al. J Cogn Neurosci* **8**, 551–565 (1996).
12. Rossion B, Gauthier I, Tarr MJ *et al. Neuroreport* **11**, 1–6 (2000).
13. Clark VP, Fan S and Hillyard SA. *Hum Brain Map* **2**, 170–187 (1995).
14. Pizzagalli D, Regard M and Lehmann D. *Neuroreport* **10**, 2691–2698 (1999).
15. Jeffreys DA. *Exp Brain Res* **78**, 193–202 (1989).
16. Rossion B, Delvenne JF, Debatisse D *et al. Biol Psychol* **50**, 173–189 (1999).
17. Hillyard SA, Mangun GR, Woldorff MG and Luck SJ. Neural systems mediating selective attention. In: MS Gazzaniga, ed. *The Cognitive Neurosciences* Cambridge MA: MIT, 1995: 665–681.
18. Sams M, Aulanko R, Hamalainen M *et al. Neurosci Lett* **127**, 141–145 (1991).
19. Giard MH and Peronnet F. *J Cogn Neurosci* **11**, 473–490 (1999).
20. Carretie L, Iglesias J and Bardo C. *J Psychophysiol* **12**, 376–383 (1998).
21. de Gelder B, Böcker KBE, Tuomainen J *et al. Neurosci Lett* **260**, 133–136 (1999).
22. Bertelson P. Starting from the ventriloquist: The perception of multimodal events. In: Sabourin M, Craik FIM and Robert M, eds. *Advances in Psychological Science, Vol.1: Biological and Cognitive Aspects*. Hove: Psychology Press, 1998: 419–439.
23. Stein BE and Meredith MA. *The Merging of the Senses*. Cambridge: Bradford Books, 1993.