



## When audition alters vision: an event-related potential study of the cross-modal interactions between faces and voices

F. Joassin<sup>a,\*</sup>, P. Maurage<sup>a</sup>, R. Bruyer<sup>a</sup>, M. Crommelinck<sup>b</sup>, S. Campanella<sup>a</sup>

<sup>a</sup> *Unité de Neurosciences Cognitives (NESC), Faculté de Psychologie et des Sciences de l'Éducation, Université Catholique de Louvain, Place du Cardinal Mercier, 101348 Louvain-la-Neuve, Brussels, Belgium*

<sup>b</sup> *Unité de Neurophysiologie Clinique (NEFY), Faculté de Médecine, Université Catholique de Louvain, Louvain-la-Neuve, Brussels, Belgium*

Received 21 May 2004; received in revised form 13 July 2004; accepted 16 July 2004

### Abstract

Ten healthy volunteers took part in this event-related potential (ERP) study aimed at examining the electrophysiological correlates of the cross-modal audio–visual interactions in an identification task. Participants were confronted either to the simultaneous presentation of previously learned faces and voices (audio–visual condition; AV), either to the separate presentation of faces (visual, V) or voices (auditive, A). As expected, an interference effect of audition on vision was observed at a behavioral level, as the bimodal condition was performed more slowly than the visual condition. At the electrophysiological level, the subtraction ( $AV - (A + V)$ ) gave prominence to three distinct cerebral activities: (1) a central positive/posterior negative wave around 110 ms, (2) a central negative/posterior positive wave around 170 ms, AND (3) a central positive wave around 270 ms. These data suggest that cross-modal cerebral interactions could be independent of behavioral facilitation or interference effects. Moreover, the implication of unimodal and multisensory convergence regions in these results, as suggested by a source localization analysis, is discussed.

© 2004 Elsevier Ireland Ltd. All rights reserved.

**Keywords:** Cross-modal interactions; Faces; Voices; ERPs

Cross-modal interactions between different cortical areas are one of the main cerebral processes contributing to our daily adapted behaviors. If the perception of stimuli within single sensory modalities begins to be well known, very little is known about the neural mechanisms by which the brain is able to establish relationships between sensory events and how it integrates them into a unified representation in order to interact properly with the objects of our environment.

Moreover, a great part of the studies examining cross-modal interactions has focused on facilitation processes: responses are facilitated (reduced reaction times and/or enhanced performances) when participants are confronted with two stimuli from two different sensory modalities relative to one unimodal stimulus [10].

However, cross-modal interactions can also take the form of an *interference effect* in which bimodal stimulations alter the performances of the participants rather than facilitating them. The most studied interference effects are the McGurk and the ventriloquist effects in which vision biases the auditory perception [1,12]. But conversely, audition can also alter vision. Sekuler et al. have shown that sounds can bias the perception of moving visual targets [14]. Moreover, the alteration of vision by audition is not limited to the cases in which the visual information is ambiguous or moving. Shams et al. have recently put in light that the synchronous presentation of multiple auditory beeps and a single visual flash induces the subjective perception of multiple flashes [15]. This sound-induced illusory flash phenomenon was also studied with neurophysiological techniques and revealed that auditory stimulations modulated the visual evoked potentials at short latencies [16] and produced significant induced gamma band responses in visual cortical regions [3].

\* Corresponding author. Tel.: +32 10 47 90 62; fax: +32 10 47 37 74.  
E-mail address: [frederic.joassin@psp.ucl.ac.be](mailto:frederic.joassin@psp.ucl.ac.be) (F. Joassin).

Therefore, the present study aimed at examining the electrophysiological correlates of the cross-modal interactions elicited by the interference of audition on vision with non-ambiguous complex stimuli. For this purpose, we used an identification task based either on the simultaneous perception of faces and voices, or on the isolated perception of one of the two kinds of information. We used faces and voices for their high ecological level. If many event-related potentials (ERPs) experiments were devoted to face perception [2,5] and voice perception [9,11], very little is known about the combination of the two. But actually, we are confronted everyday with social situations in which we have to integrate visual and auditory information, at least to know who is speaking. We speculated that voices could alter identification of faces for several reasons. First, auditory stimulations have been shown to be more slowly processed than visual ones [19]. Second, the amount of available information at the onset of the stimulations is not identical for faces and voices: for faces, all the visual information is available immediately whereas, for voices, the auditory information spreads in time. This temporal asynchrony has been argued to be an important feature leading to an interference effect [4]. In the same way, Shimojo and Shams showed that the direction of cross-modal interactions depends on the structure of the stimuli, that is the more discontinuous (or salient) signal becomes the influential signal [17]. In this case, voices can be considered as more discontinuous than faces and could thus alter the processing of faces.

Ten healthy undergraduate subjects (eight males) took part to this experiment (mean age: 21.5; S.D.: 1.7). All but one were right-handed and all had a normal-to-corrected vision. A pre-experimental learning phase served to familiarize them with the twelve face-name-voice associations used in the experiment. Each association was composed by an unknown face picked from the Stirling face database (black/white photos, front view, neutral facial expression, six males/six females, see at <http://www.pics.psych.stir.ac.uk/>), a Belgian family name controlled for its frequency of appearance in the Belgian population, and a particular voice pronouncing the family name. At the end of the learning phase, all participants were able to identify each voice and to recuperate the name associated with each face.

During the experiment, participants were confronted to three different conditions mixed in 20 blocks of 30 trials (200 trials per condition), that differed in the kind of stimuli presented: either a face alone, a voice alone or the simultaneous presentation of a face and a voice. Each trial began with the presentation of a grey fixation cross appearing for 300 ms on a black screen. It was replaced by a family name presented visually for 500 ms. After an empty interstimulus interval varying randomly from 1000 to 2000 ms, the stimulus(i) appeared for 1000 ms, followed by an empty intertrial interval of 1000 ms. For each trial, participants had to decide if the presented stimulus(i) was (were) correctly primed by the preceding visual family name or not (50% of the trials were correct, the remaining were primed by another family name

from the set of the twelve proper names previously learned). This procedure ensured that (1) the participants were implied in a better controlled cognitive task than the simple passive perception of stimuli [4,12], (2) the task was identical across all conditions, allowing the comparison of the conditions and the subtraction of the unimodal conditions from the bimodal one. The participants answered by pressing a key on a paddle. In the unimodal auditory condition, the incorrect trials were always constituted by a wrong voice pronouncing the good name whereas in the unimodal visual condition, the incorrect trials consisted in one of the five other faces from the same gender. In the bi-modal condition, the incorrect trials were half constituted by the good face but a wrong voice (CV/IA), half by the good voice but a wrong face (IV/CA). All incorrect visuo-auditive trials were regrouped in the ERP analyses. Furthermore, it is worth noting that in both auditory conditions (auditive and visuo-auditive), the voice was always pronouncing the primed name so that participants had to identify the voice itself.

EEG was recorded by 32 electrodes mounted in an electrode Quick-Cap. Electrode positions included the standard 10–20 system locations and intermediate positions. Recordings were made with a linked mastoid physical reference. The EEG was amplified by battery-operated A.N.T.<sup>®</sup> amplifiers with a gain of 30,000 and a band-pass of 0.01–30 Hz. The impedance of all electrodes was kept below 20 k $\Omega$ . EEG was continuously recorded (sampling rate 512 Hz, Eeprobe, A.N.T.<sup>®</sup>) and trials contaminated by EOG artifacts were manually eliminated off-line. Epochs beginning 200 ms prior to stimulus onset and continuing for 1500 ms were created. Codes synchronized with stimulus delivery were used to average selectively the epochs associated with different stimulus types. Two parameters were coded for every stimulus: (1) the type of the stimulus (voices alone, faces alone, simultaneous faces and voices), (2) the type of trials (correct or incorrect). This coding allowed us to compute different averages of ERP target stimuli. The averages were computed for each subject individually. Then, grand-averaging were computed for the six conditions (correct visual (CV), incorrect visual (IV), correct auditive (CA), incorrect auditive (IA), correct visuo-auditive (CV/CA), incorrect visuo-auditive (mean of CV/IA and IV/CA)).

At a behavioral level, an ANOVA with type of stimulus (visual, auditory, visuo-auditive) and type of trial (correct, incorrect) as within-participants factors was performed on percentages of correct responses (Table 1, lower part). Please note that incorrect visuo-auditive trials were grouped in this first and in the following analysis. It showed a significant main effect of the type of stimulus ( $F(2,18) = 14.2, P < .01$ ), but no significant main effect of the type of trial ( $F(1,9) = 1.65, n.s.$ ). The interaction between the type of stimulus and the type of trial failed to reach a significant value ( $F(2,18) = 3.58, n.s.$ ). The main effect of the type of stimulus is due to the fact that auditory condition was less well performed than the two others, that did not differ significantly between each other.

Table 1

Mean reaction times and percentages of correct responses (standard deviations in parentheses) for correct and incorrect trials of each condition

	Visual		Auditive		Visuo-auditive		
	CV	IV	CA	IA	CV/CA	CV/IA	IV/CA
RTs (in ms)	853 (204)	826 (129)	1009 (222)	990 (176)	953 (211)	1043 (173)	907 (142)
Percent of correct responses	95.2 (1.8)	96.3 (2.7)	91 (5.8)	87.9 (7.6)	95.7 (3.5)	82.8 (15.4)	96.5 (2.7)
						Mean of CV/IA and IV/CA: 975 (156)	
						Mean of CV/IA and IV/CA: 89.3 (8.5)	

CV: correct visual; IV: incorrect visual; CA: correct auditive; IA: incorrect auditive; CV/CA: correct visuo-auditive; CV/IA: correct visual/incorrect auditive; IV/CA: incorrect visual/correct auditive.

The same ANOVA was performed on reaction times (Table 1, upper part). We did observe a significant main effect of the kind of stimulus ( $F(2,18) = 78.6, P < .0001$ ) but no significant main effect of the kind of trial ( $F(1,9) = 0.11, n.s.$ ) and no significant interaction between the type of stimulus and the type of trial ( $F(2,18) = 2.02, n.s.$ ). The auditory condition was thus less quickly performed than the two others. Moreover, the visual condition was performed faster than the visuo-auditive condition.

Two further analyses were performed to make sure of the existence of an interference effect. First, we compared the trials in which faces were presented in isolation (CV and IV) or with a voice (CV/IA and IV/CA). An ANOVA with type of stimulus (visual or visuo-auditive) and type of trial (correct or incorrect face) was computed. It showed a main effect of the type of stimulus ( $F(1,9) = 104.92, P < .001$ ) indicating that the visual conditions were performed more quickly than the bimodal ones, a main effect of the type of trial ( $F(1,9) = 17.94, P < .001$ ) and a significant interaction between the type of stimulus and the type of trial ( $F(1,9) = 12.14, P < .01$ ). This interaction is due to the fact that CV and IV did not differ significantly between each other ( $F(1,9) = 0.80, n.s.$ ) whereas the IV/CA trials, in which the visual information was sufficient to take a decision, were performed more quickly than the CV/IA trials in which the auditory information was necessary to answer correctly ( $F(1,9) = 63.20, P < .001$ ). Secondly, we compared the trials containing only

voices (CA and IA) or voices and faces (CV/IA and IV/CA). The same analysis was performed. It revealed a main effect of the type of trial ( $F(1,9) = 13.04, P < .01$ ) but the main effect of the type of stimulus failed to reach a significant statistical level ( $F(1,9) = 2.08, n.s.$ ). The interaction between the type of stimulus and the type of trial was significant ( $F(1,9) = 25.77, P < .001$ ) and due to the fact that CA and IA did not significantly differed between each other ( $F(1,9) = 0.42, n.s.$ ).

The behavioral results clearly show an interference effect of audition on vision as (1) the auditive and bimodal conditions were performed more slowly than the visual condition, (2) the addition of an auditive information interfered with the processing of the visual information, and (3) the addition of a visual information did not facilitate the processing of the auditive information.

At an electrophysiological level, we first checked the ERP complexes classically associated with the early processing of faces and voices (Fig. 1, left) [2,11]. In the visual condition, the P1 wave was maximally recorded at posterior sites (Oz, O1, O2) around 120 ms (mean amplitude of  $4 \mu V$ ) and the N1 wave was maximally recorded at central electrodes (Cz, C1, C2) around 130 ms (mean amplitude of  $-2.13 \mu V$ ). The N170/VPP complex was maximally recorded around 180 ms at Oz, O1 and O2 for the N170 component (mean amplitude of  $-7 \mu V$ ) and Cz, C1 and C2 for the VPP component (mean amplitude of  $4 \mu V$ ).

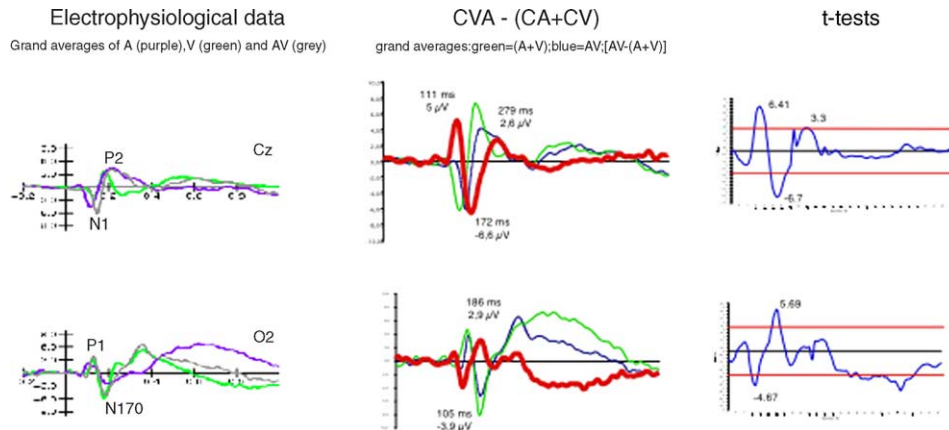


Fig. 1. (Left) ERP waves elicited by the three conditions (A in purple, V in green and AV in grey); (middle) ERPs elicited toy the subtraction [AV - (A + V)] (in red), the sum of A + V (in green) and AV (in blue); (right) *t*-values for the subtraction (significance levels are represented by red lines). (Upper part) ERPs recorded at Cz; (lower part) ERPs recorded at Oz.

The processing of voices also elicited two classical cortical ERP complexes. The P1/N1 complex peaked around 114 ms at posterior electrodes (Oz, O1, O2) for the P1 (mean amplitude of 3.12  $\mu\text{V}$ ) and at central electrodes (Cz, C1, C2) for the N1 (mean amplitude of  $-4.63 \mu\text{V}$ ). The second ERP complex, the P2/N2, was maximally recorded at occipital electrodes around 198 ms (mean amplitude of  $-3.81 \mu\text{V}$ ) for the N2 component, and at central electrodes around 212 ms (mean amplitude of 4.7  $\mu\text{V}$ ) for the P2 component.

These classical P1/N1 and N2/P2 complexes were also observed in the bi-modal condition. The P1 wave peaked at occipital electrodes around 132 ms (mean amplitude of 3.5  $\mu\text{V}$ ) and the N1 wave at central electrodes around 145 ms (mean amplitude of  $-6.13 \mu\text{V}$ ). The N2 wave was maximally recorded at posterior sites around 187 ms (mean amplitude of  $-6.32 \mu\text{V}$ ) and the P2 was maximal at central sites around 215 ms (mean amplitude of 4.43  $\mu\text{V}$ ).

To conclude these first observations, the classical ERP components associated with the perception and the early processing of stimulation have been elicited by our own faces and voices, which ensures us that our stimulations were correctly encoded by the participants.

The main analysis consisted in subtracting the unimodal averages from the bi-modal average [CV/CA – (CA + CV)], which is a classical method to put in light the electrophysiological correlates of the cross-modal interaction. Note that we performed this subtraction only on the correct trials, as it has been underlined that cross-modal interactions are more prone to appear in correct than incorrect trials because of their references towards representations in long-term memory [19]. We first computed the subtraction for each participant individually. We then performed the grand-average of the results of the subtraction.

It revealed three main complexes of electrical activities at distinct latencies (Fig. 1, middle): (1) around 100 ms, a positive-negative complex recorded respectively at central (Cz: peak latency at 111 ms, maximum amplitude of 5  $\mu\text{V}$ ) and posterior sites (Oz: peak latency at 105 ms, maximum amplitude of  $-3.9 \mu\text{V}$ ); (2) around 175 ms, a negative-positive complex, the negativity peaking at Cz around 172 ms (maximum amplitude of  $-6.6 \mu\text{V}$ ) and the positivity peaking at Oz around 186 ms (maximum amplitude of 2.9  $\mu\text{V}$ ); (3) a positive wave recorded at central electrodes and maximally peaking at Cz around 279 ms (maximum amplitude of 2.6  $\mu\text{V}$ ).

Significant effects were then calculated at each selected electrode using Student *t*-tests (amplitude of the difference wave compared to zero for successive 10-ms intervals from 0 to 800 ms). The spatio-temporal patterns that had a significant amplitude ( $P < .01$ ) at least on one electrode for two consecutive intervals were considered as significant (Fig. 1, right). For the central electrodes, the subtraction wave was significantly different from zero at Cz between 90 and 130 ms ( $t = 6.41$ ,  $P < .001$ ), 140 and 200 ms ( $t = -6.7$ ,  $P < .001$ ) and from 260 to 280 ms ( $t = 3.3$ ,  $P < .01$ ). For the posterior electrodes, the subtraction wave significantly differed from

zero at Oz between 90 and 110 ms ( $t = -4.67$ ,  $P < .01$ ) and between 170 and 200 ms ( $t = 5.69$ ,  $P < .001$ ).

Finally, a source localization analysis was performed on the grand-average ERP waves of the subtraction. We used the ASA 3.0 software (A.N.T.<sup>®</sup>) which determines the position and the orientation of intracranial dipoles and their time-varying strength by using a three-layer spherical head model. A fixed dipole method was applied in which the position of the dipoles was constrained but their orientation and strength remained free. Dipole modeling was performed within the time window of each wave (Table 2; Fig. 2). Between 90 and 130 ms, the electrophysiological results were best explained by two pairs of bilateral dipoles localized in the superior colliculi and the fusiform gyri (Brodmann Area 37), and more specifically in the Face Fusiform Area (FFA), a cerebral region highly responsive to human faces [8]. For the 140–200 ms time window, the lowest residual variance was obtained with three pairs of bilateral dipoles localized in the superior colliculi, the superior temporal gyri (associative auditory cortex, BA 22) and the inferior frontal gyri (BA 45). Finally, between 260 and 280 ms, the scalp cerebral activity was best explained by five dipoles, a pair of bilateral dipoles localized in the superior colliculi, one dipole in the superior frontal gyrus (BA 10), one in the inferior frontal gyrus (BA 45) and one in the fusiform gyrus (BA 37).

The present study aimed at examining the electrophysiological correlates of the interference effect of audition on vision with non-ambiguous complex stimuli in an identification task based either on the simultaneous perception of faces and voices, or on the isolated perception of one of the two kinds of information. At a behavioral level, we observed that audition actually altered vision, as the simultaneous presentation of a face and a voice slowed down the speed of response relative to the presentation of a face alone, confirming thus the hypothesized interference effect.

At the electrophysiological level, these visuo-auditive interactions were analyzed by subtracting the unimodal conditions from the bimodal condition. It revealed a positive-negative-positive ERP component recorded on the central electrodes, associated with a negative-positive wave recorded on the posterior electrodes. Furthermore, we observed that the first central positive/posterior negative wave was best explained by a pair of dipoles originating in the associative visual cortex, whereas the central negative/posterior positive wave was best explained by a pair of dipoles localized in the associative auditory cortex. Finally, the second central positive wave was best explained by a network of cortical areas including the superior colliculi, the fusiform gyrus, the associative auditory cortex and the superior frontal gyrus. Then we suggest that the present cross-modal interactions can be modeled by distinct temporal interactions: (1) the auditory information could influence the processing of the visual information in the fusiform gyrus around 90–130 ms, (2) the visual information could then influence the processing of the auditory information in the associative auditory areas (BA 22) around 140–200 ms, and (3) finally, around 260–280 ms,

Table 2  
Dipole modeling of the three significant ERP activities elicited by the subtraction  $AV - (A + V)$

Time window (ms)	Best fit latency (ms)	RV (%)	Cerebral area	Coordinates (Talairach coordinate system)			Magnitude (NAM)
				x	y	z	
90–130	113	2.6	Right SC	3	–28	–2	1297
			Left SC	–3	–28	–2	1183
			Right FG	40	–55	–10	100
			Left FG	–40	–55	–10	43
140–200	174	5.8	Left SC	–3	–28	–2	3326
			Right SC	3	–28	–2	3243
			Left MTG	–55	–20	1	168
			Right MTG	55	–20	1	132
			Left IFG	–41	26	13	105
			Right IFG	41	26	13	50
260–280	275	8.8	Left SC	–3	–28	–2	1085
			Right SC	3	–28	–2	1136
			Right SFG	11	72	19	164
			Right FG	–40	–55	–10	65
			Left IFG	–41	26	13	25

Cerebral areas: FG, fusiform gyrus (Brodmann Area 37); SC, superior colliculus; MTG, middle temporal gyrus (BA 22); IFG, inferior frontal gyrus (BA 45); SFG, superior frontal gyrus (BA 10).

the unimodal and cross-modal areas could interact with semantic areas, like the inferior and superior frontal gyri [13], in order to compare the created bimodal representation with one of the representations stored in long-term memory.

It is worth noting that many studies focused on the neural correlates of cross-modal interactions producing a facilitation of the responses, i.e. shorter RTs and/or higher performances when participants are confronted to bimodal stimulations relative to unimodal ones. Nevertheless, to our knowledge the present study is the first one to have examined the cerebral markers of the inhibitory processes of cross-modal interactions by using complex stimulations such as faces and voices. Interestingly, the present results are very similar to those obtained by Teder-Sälejärvi et al., who showed, in an ERP experiment using light flashes and noise bursts, that bimodal stimulations *facilitated* the task (detection of infrequent tar-

get stimuli) and produced a positive–negative–positive wave on frontal, central and parietal electrodes [19]. Moreover, their dipole modeling gave results very similar to our present data. The similarity between the results of Teder-Sälejärvi et al. [19] and the present results allow us to hypothesize that (1) these electrophysiological activities reflect general interactions between neural populations processing different sensorial stimulations, (2) these cross-modal interactions could be independent of facilitatory or inhibitory processes of one type of stimulation on another one.

It is also interesting to note that the results of our dipole modeling always included the superior colliculus, a mid-brain structure considered as a multimodal convergence region [18] and that has already been observed in an ERP study investigating the audio–visual object recognition processes [7]. When we excluded these regions from the source localization

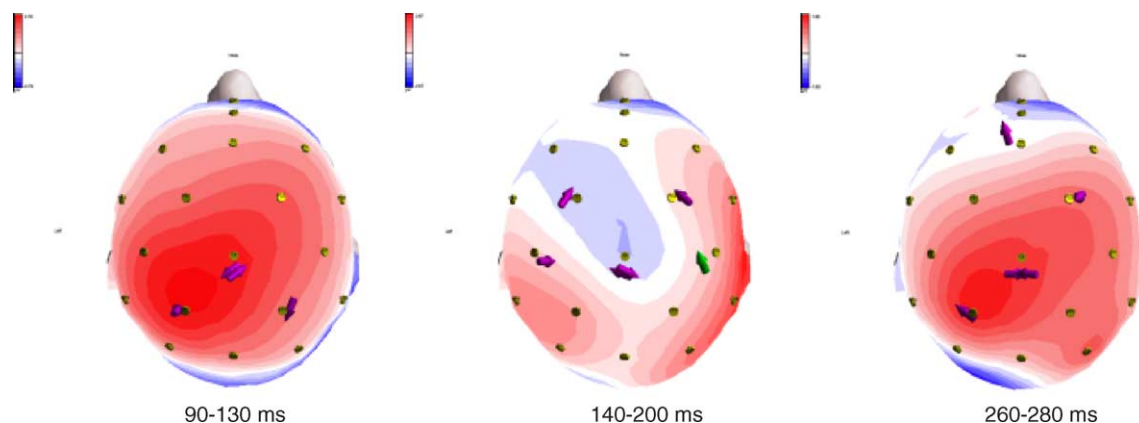


Fig. 2. Top views of the dipole modeling and topographies of the ERPs waves elicited by the subtraction  $[AV - (A + V)]$ . (Left) Between 90 and 130 ms (two pairs of bilateral dipoles localized in the superior colliculi and the fusiform gyri); (middle) between 140 and 200 ms (three pairs of bilateral dipoles localized in the superior colliculi, the medial temporal gyri and the inferior frontal gyri); (right) between 260 and 280 ms (a pair of dipole localized in the superior colliculi, one dipole in the left fusiform gyrus, one dipole in the right inferior frontal gyrus, one dipole in the left superior frontal gyrus).

analyzes, the residual variance dramatically increased, which was not the case for visual or auditive stimulations presented in isolation. These results could give us new leads to better understand the mechanisms by which the different unimodal regions interact. It could be possible that the unimodal regions, processing simultaneously different sensory stimulations, send inputs to multimodal integration regions, such as the superior colliculi, where the representations of each stimulation are integrated. In turn, the multimodal convergence regions could send back projections to the unimodal brain areas, each “primary” sensory region being then modulated by the stimulations in the other modality [6].

Nevertheless, further experiments are needed, notably to compare the present data with the neural correlates of cross-modal interactions between complex and non-ambiguous stimulations producing a behavioral facilitation, or to investigate the possible existence of synchronous cerebral activity between the unimodal regions within the range of the gamma band [3].

### Acknowledgements

This study was made possible by FRSM No. 3458303F. The first, second and last authors are supported by the Belgian Found for Scientific Research (F.N.R.S.).

### References

- [1] D. Alais, D. Burr, The ventriloquist effect results from near-optical bimodal integration, *Curr. Biol.* 14 (2004) 257–262.
- [2] S. Bentin, T. Allison, A. Puce, E. Perez, G. McCarthy, Electrophysiological studies of face perception in humans, *J. Cogn. Neurosci.* 8 (1996) 551–565.
- [3] J. Bhattacharya, L. Shams, S. Shimojo, Sound-induced illusory flash perception: role of gamma band responses, *Neuroreport* 13 (14) (2002) 1727–1730.
- [4] G.A. Calvert, P.C. Hansen, S.D. Iversen, M.J. Brammer, Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect, *Neuroimage* 14 (2001) 427–438.
- [5] S. Campanella, P. Quinet, R. Bruyer, M. Crommelinck, J.M. Guérit, Categorical perception of happiness and fear facial expressions: an ERP study, *J. Cogn. Neurosci.* 14 (2) (2002) 1–18.
- [6] J. Driver, C. Spence, Multisensory perception: beyond modularity and convergence, *Curr. Biol.* 10 (2000) 731–735.
- [7] A. Fort, C. Delpuech, J. Pernier, M.H. Giard, Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans, *Cereb. Cortex* 12 (10) (2002) 1031–1039.
- [8] N. Kanwisher, J. McDermott, M.M. Chun, The fusiform face area: a module in human extrastriate cortex specialized for face perception, *J. Neurosci.* 9 (1997) 462–475.
- [9] D.A. Levy, R. Granot, S. Bentin, Processing specificity for human voice stimuli: electrophysiological evidence, *Neuroreport* 12 (2) (2001) 2653–2657.
- [10] S. Molholm, W. Ritter, M.M. Murray, D.C. Javitt, C.E. Schroeder, J.J. Foxe, Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study, *Cogn. Brain Res.* 14 (2002) 115–128.
- [11] H. Müller, M. Kutas, What’s in a name? Electrophysiological differences between spoken nouns proper names and one’s own name, *Neuroreport* 8 (1996) 221–225.
- [12] I.R. Olson, J.C. Gatenby, J.C. Gore, A comparison of bound and unbound audio-visual information processing in the human cerebral cortex, *Brain Res.* 14 (2002) 129–138.
- [13] R.A. Poldrack, A.D. Wagner, M.W. Prull, J.E. Desmond, G.H. Glover, J.D.E. Gabrieli, Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex, *Neuroimage* 10 (1) (1999) 15–35.
- [14] R. Sekuler, A.B. Sekuler, R. Lau, Sound alters visual motion perception, *Nature* 385 (1997) 308.
- [15] L. Shams, Y. Kamitani, S. Shimojo, Visual illusion induced by sound, *Cogn. Brain Res.* 14 (2002) 147–152.
- [16] L. Shams, Y. Kamitani, S. Thompson, S. Shimojo, Sound alters visual evoked potentials in humans, *Neuroreport* 12 (17) (2001) 3849–3852.
- [17] S. Shimojo, L. Shams, Sensory modalities are not separate modalities: plasticity and interactions, *Curr. Opin. Neurobiol.* 11 (2001) 505–509.
- [18] B.E. Stein, W. Jiang, M.T. Wallace, T.R. Stanford, Nonvisual influences on visual-information processing in the superior colliculus, *Prog. Brain Res.* 134 (2001) 143–156.
- [19] W.A. Teder-Sälejärvi, J.J. McDonald, F. Di Russo, S.A. Hillyard, An analysis of audio-visual crossmodal integration by means of event-related potentials (ERP) recordings, *Cogn. Brain Res.* 14 (2002) 106–114.