Neuropsychological Trends

1
April 2007

Cross-modal perception (face and voice) in emotions

ERPs and behavioural measures

Michela Balconi - Alba Carrera

Department of Psychology - Catholic University, Milan

michela.balconi@unicatt.it

ABSTRACT

Emotion decoding constitutes a case of multimodal processing of cues from multiple channels. Previous behavioural and neuropsychological studies indicated that, when we have to decode emotions on the basis of multiple perceptive information, a cross-modal integration has place. The present study investigates the simultaneous processing of emotional tone of voice and emotional facial expression by event-related potentials (ERPs), through an ample range of different emotions (happiness, sadness, fear, anger, surprise, and disgust). Auditory emotional stimuli (a neutral word pronounced in an affective tone) and visual patterns (emotional facial expressions) were matched in congruous (the same emotion in face and voice) and incongruous (different emotions) pairs. Subjects (N=30) were required to process the stimuli and to indicate their comprehension (by stimpad). ERPs variations and behavioural data (response time, RTs) were submitted to repeated measures analysis of variance (ANOVA). We considered two time intervals (150-250; 250-350 ms post-stimulus), in order to explore the ERP variations. ANOVA showed two different ERP effects, a negative deflection (N2), more anterior-distributed (Fz), and a positive deflection (P2), more posterior-distributed, with different cognitive functions. In the first case N2 may be considered a marker of the emotional content (sensitive to type of emotion), whereas P2 may represent a cross-modal integration marker, it being varied as a function of the congruous/incongruous condition, showing a higher peak for congruous stimuli than incongruous stimuli. Finally, a RT reduction was found for some emotion types for congruous condition (i.e. sadness) and an inverted effect for other emotions (i.e. fear, anger, and surprise).

Keywords: ERPs; emotion; voice; facial expression; cross-modal perception

1. INTRODUCTION

Normally, our sense systems are simultaneously achieved by multiple stimuli which belong to different sensory modalities. We could ask which is the function of this multisensory integration. The ability to integrate the stimulus from different modalities in order to create a whole perception is a main phenomenon for human cognition and behaviour. This ability allows to enhance stimulus elaboration, and to produce a more rapid and accurate behavioural response (Ghazanfar *et al.*, 2005). Indeed, multisensory redundant information integration improves signal identification, detection, and decoding. Moreover, recent studies indicate that it produces a better performance in attention orientation and recognition tasks (Schroeder *et al.*, 2003).

Even if cross-modal integration is a main perceptual phenomenon, our knowledge is still incomplete, and we don't know exactly how an object from different sensory modalities is perceived as an unitary event. Recent empirical researches demonstrated the complexity of cross-modal integration, due to the fact that this process includes multiple and heterogeneous cognitive and physiological mechanisms with endogenous and exogenous origins. As regard to endogenous mechanisms, Fort *et al.* (2002) have found the effect of the subjective expertise level on the task execution. Subjects with visual dominance and subjects with auditory dominance displayed different activation patterns in response to audio-visual stimuli. On the other hand, an important role is attributed to exogenous factors, as the typology of the stimuli, the experimental conditions and the specific task. Particularly, stimulation has a clear effect on integration modalities (Callan *et al.*, 2001; Calvert, 2001).

The present study investigates audio-visual emotion decoding employing emotional faces and voices. When persons decode the emotions, they have to integrate multiple information belonging to different sensory systems. Emotional facial and vocal indexes are complex stimuli, charged of social meaning. In this perspective, they greatly differ from stimuli that are traditionally used to study cross-modal perception (generally light flashes and inarticulate simple sounds). Some recent studies tried to understand this process, which have a key role in emotion communication. Using behavioural research paradigm, de Gelder and colleagues (de Gelder & Vroomen, 2000) tried to determine if, in a bimodal situation with emotional audio-visual information, both auditory and visual modalities contribute to emotion identification. Results indicate that emotion is more rapidly decoded when the face and the voice are congruous (i. e. expressing the same emotion) then in the unimodal condition (auditory or visual only). This seems to indicate that integration process is an usual and efficient mechanism. On the contrary, longest times are detected in the incongruous bimodal condition (voice and face expressing different emotions). De Gelder experiment indicates an effect of the face on the voice and, vice versa, an effect of the voice on the face, as previously found by the pioneer study by Massaro and Egan (1996). Cross-modal bias is noticed also when subject is explicitly required to pay attention to an unique sensory modality (visual or auditory). It suggests that integration has place at an early perceptual stage, in an automatic and obligatory manner. This hypothesis is confirmed by a study by de Gelder (de Gelder, Vroomen & Bertelson, 1998), which introduced a procedural variant (inverted faces). Face inversion implies a drastic decrease of the possibility to define identity and expression. Indeed face identification differs from object identification, because it involves a holistic elaboration of the whole configuration (gestalt) and not only an analytic elaboration of single components (Balconi, 2006). The study indicates that facial expression decoding influences tone of voice judgments only for rightly-oriented faces and not for inverted face. This effect is probably due to the automatic nature of cross-modal integration.

In order to explain these results, we can suppose that emotional information redundancy allows a more efficient behavioural response. This enhancement is due not to the complementarity of the auditory and visual systems but to the fact that the system is able to integrate information in an early processing stages, and it could produce a rapid response. On the contrary, the process would be slower, if the system elaborates stimuli separately and integrates them only in a final step. Indeed, several studies suggested that cross-modal convergence has already place during the first information processing phase, through the activation of sensory integration areas, as parietal, temporal (Schroeder & Foxe, 2002) and frontal cortex (Graziano *et al.*, 1997) in addition to the specific unimodal sensory areas (Teder-Sälejärvi *et al.*, 2002). Among the current approaches, the *interactive coactivation model* receives broad confirmation. This model suggests the presence of a reciprocal effect between stimuli belonging to different sensory modalities, supposing that the integration has place at multiple processing levels, that is at *sensory*, *cognitive* and *response choice levels* (Calvert *et al.*, 2000; Hadjikhani & Roland, 1998). Primarily, researchers agree that integration begins at an early sensory elaboration stage (Giard & Peronnet, 1999).

At the neuropsychological level, many studies aim to reveal cerebral processes implied in emotion multimodal decoding (Johnstone *et al.*, 2007; Pourtois *et al.*, 2005). Specifically, most of the studies focused attention on early perceptual stages of emotional stimuli elaboration. Nevertheless, the qualitative nature and the time course of these phenomena constitute a controversial matter, and, in this perspective, event-related potentials (ERPs) are very useful research method (Rugg & Coles, 1997). Indeed, more frequently ERPs, such as the N1 and the MMN, are bounded to sensory processes. First researches exploited MMN (*mismatch negativity*) index, elicited by auditory stimuli when, after a series of repetitive stimuli, a deviant stimulus appears (Näätänen, 1992). This ERP is not under attentive control, but it has an automatic nature and it signals the detection of a stimulus which disregards the expectations (Levänen & Sams, 1997). De Gelder and colleagues (de Gelder *et al.*, 1999) used MMN in order to investigate emotional facial effect on auditory emotional signal processing. Congruous and incongruous couples of auditory (words pronounced in a sad or angry tone) and visual (facial expressions of sadness and anger) stimuli were presented to subjects, who were required to pay attention to the face and to ignore the auditory component. Results indicate that when, after a series of congruous (or incongruous) couples, an incongruous (or congruous) one is presented, an early negative cerebral response (latency 178 ms) has place in the anterior areas. This response is like to be a MMN variation. Similarly, Surakka *et al.* (1998) observed the effect of emotional images on auditory stimuli, finding that MMN amplitude was significantly influenced by hedonic valence and arousal of the emotional stimulus.

Moreover, other research on emotional cross-modal processing examined the amplitude of ERP components, finding an increasing or decreasing of unimodal early indexes, such as the auditory peak N1, which has place about 100 ms post stimulus in the sensory specific areas (Calvert, Brammer & Iversen, 1998; Giard & Peronnet, 1999; Raij, Uutela & Hari, 2000). Indeed, the increasing activity in modality-specific cortex is a main electrophysiological correlate of cross-modal processes (de Gelder, 2000; Driver & Spence, 2000). For example, an increased activation was found in the auditory cortex as a consequence of facial stimuli presentation (Calvert *et al.*, 1997), in the fusiform gyrus and in the amygdala during emotion bimodal stimuli presentation (Dolan, Morris & de Gelder, 2001), and in the tactile areas during a visual-tactile stimulation (Macaluso, Frith & Driver, 2000). In his recent studies, Pourtois (Pourtois *et al.*, 2000; 2005) presented to subjects congruous and incongruous couples consisting of auditory (4-syllable fragments pronounced in a sad or angry manner) and visual (sad and angry facial expressions presented with a right and an inverted orientation) stimuli. Experimental task required to pay attention to faces ignoring voices. Results revealed that visual information influences auditory stimulus processing already after 110 ms. This effect consists of a N1 intensity increasing, as previously indicated by studies that used non emotional stimuli (Giard & Peronnet, 1999). Moreover, integration has place only when stimuli are congruous, confirming the corresponding behavioural data (de Gelder &

Vroomen, 2000). On the whole, these studies suggest that cross-modal integration is signalled by the amplification effect in the sensory specific areas and at the level of cortical networks that have a multimodal integration function, such as posterior parietal cortex and medial temporal gyrus (Johnstone *et al.*, 2007; Mesulam, 1998).

As regard to the *cognitive level*, Pourtois and colleagues (2002) empirically demonstrated that audiovisual emotional stimuli processing implies specific ERP variations that signal even the cognitive integration at a successive temporal stage. In particular, a positive phenomenon about 240 ms post-stimulus with a posterior localization was found (P2b). The authors suggested that already at about 100 ms an increasing of unimodal component intensity occurs (de Gelder *et al.*, 1999; Giard & Peronnet, 1999; Pourtois *et al.*, 2000), followed by a second component, the P2b, which is sensible to audiovisual stimulus semantic content. Moreover, Pourtois study demonstrated that emotional audiovisual stimuli integration at cognitive level probably has place during pattern elaboration at about 220 ms, in posterior cerebral cortex. A specific analysis on the localization of P2b showed an involvement of the anterior cingulate cortex, subserving in congruence/incongruence processing (McLeold & McDonald, 2000), as supported by recent fMRI (Calvert, Campbell & Brammer, 2000) and magnetoencephalographic studies (Raij *et al.*, 2000).

Nonetheless, at present several limits exist in neuropsychological research applied to emotional cross-modal perception. Particularly, since most of the studies focused on early perceptive phenomena, it is necessary to investigate in a broad manner the later cognitive components (after the 100 ms post-stimulus). Beside, previous studies employed a small number of emotions. The possible effect of different emotional patterns (for example as a function of hedonic valence and arousal of the emotion) requires further analysis (Balconi & Carrera, 2006; Balconi & Pozzoli, 2003; Johnstone *et al.*, 2007).

This study aims to investigate more deeply these processes. In particular the experimental analysis includes two phases:

Phase 1

1. First of all, we aim to verify the presence of medium- and long-term latency *cognitive effects* in emotional cross-modal decoding, following the first perceptive stage. Particularly, we expect to detect the presence of ERP variations that account for cognitive integration as a result of the synthesis between multimodal emotional information successively to the perception of the stimulus.
2. Secondly, we want to investigate the presence of different ERP profiles

for *congruous* and *incongruous conditions*. Specifically, we suppose that also at the cognitive level ERP differences in terms of latency and amplitude of the peak could be detected for congruous and incongruous emotional matches.

3. Third, we aim to define a possible *localization effect* as a function of the experimental conditions. Indeed, a heterogeneous distribution of peak variations in case of congruous/incongruous patterns can mark the presence of multiple cortical modules.
4. Finally, we consider behavioural data (RTs). Since previously it was found a facilitation effect (RT reduction) for congruous stimuli as compared with incongruous ones, we expect an analogous reduction RTs as a function of the experimental conditions.

Phase 2

A successive emotion-by-emotion comparison allow us to explore the significance of cross-modal integration processes as a function of the specific emotional patterns.

1. The use of a wide range of emotions (six emotional patterns) allow us to analyze more deeply the effect of different emotional content on integration process. Moreover, it is possible that congruous or incongruous perception could affect in a different manner the intersensory integration process as a consequence of the emotion expressed.
2. Moreover, the effect of specific emotional patterns on RT variation was investigated.

2. METHOD

Participants

30 subjects took part in to the study (9 males and 21 females; age: *M* = 24.4; *SD* = 3.4), university students attending to Catholic University of Milan. The whole sample had normal visual and auditory abilities and was right-handed. Subjects provided their agreement to take part in a research about audio-visual stimuli perception.

Stimuli

Vocal stimuli. We used a single short stimulus (composed by a single syllable) with an emotionally neutral content ("ora", "now"). We choose a single word in order to control more easily the vocal parameters without the interference of other components (such as sentence speech profile, temporal variations, i.e. pauses). Moreover, it allowed not to cause the subjects to be more attentive to the vocal component than the visual one. Finally, this choice allowed a perfect synchronization between facial and vocal patterns (temporal overlapping) (Pourtois *et al.*, 2002). An actor reproduced the word giving an emotional valence to the stimulus. He/she was explicitly required to express the different emotional valence using vocal components, carrying different emotional correlates thanks to intonational profile variations (respectively anger, fear, sadness, happiness, disgust surprise and a control condition represented by a neutral expression) (Portois *et al.*, 2002). In other words, pitch, intensity and vocal profile were modified in order to produce the six different emotions (Scherer, 1984). In order to facilitate the coherence of vocal modification with each emotional correlate, the actor was invited to read a short emotional script which favoured the self-engagement with the emotional situation. The recording was performed using the software Sonic Foudry Acid Music. As regard to temporal parameters, a substantial homogeneity between the vocal stimuli was maintained (duration *M* = 550 ms; *SD* = 223 ms; range 400-550 ms).

The semantic value and the pertinence of the stimuli were verified by pre-experimental subjects (11 university students) in a pre-experimental phase, through an apposite questionnaire. A first item verified the correct labelling of each emotion (disgust 80.3%; happiness 100.0%; fear 100.0%; anger 83.3%; surprise 100.0%; sadness 76.7%; neutral 83.3%). Afterwards, judgment of consistency in evaluating the vocal emotional stimuli was assessed (5 points Likert scale). We didn't found significant differences between the emotions, since all emotional patterns were considered adequate in expressing the emotions (disgust *M* = 4.5; *SD* = 1.6; happiness *M* = 3.7; *SD* = 1.4; fear *M* = 4.2; *SD* = 1.3; anger *M* = 4.7; *SD* = 0.5; surprise *M* = 4.0; *SD* = 1.1; sadness *M* = 3.8; *SD* = 1.2; neutral *M* = 3.5; *SD* = 1.4).

Facial stimuli

Visual stimuli were facial patterns black/white (11 x 15 cm) portraying the face of a young male actor (Ekman & Friesen 1976), for the same emotions vocally reproduced. Emotional evaluation and consistency was assessed in a pre-experimental phase by 10 subjects. All emotions were correctly labelled (disgust 78.3%; happiness 95.1%; fear 98.0%; anger 83.4%; surprise 98.0%; sadness 75.7%; neutral 78%). Moreover, subjects expressed a judgment of consistency for each emotion (disgust *M* = 4.3; *SD* = 1.4; happiness *M* = 3.8; *SD* = 1.5; fear *M* = 4.9; *SD* = 1.2; anger *M* = 4.4; *SD* = 0.9; surprise *M* = 4.1; *SD* = 1.5; sadness *M* = 4.2; *SD* = 1.1; neutral *M* = 3.46; *SD* = 1.2).

Cross-modal stimulation.
Vocal and visual stimuli were composed in congruous (the same emotion for both sensory modalities) and incongruous (face and voice expressing different emotions) pairs. In the incongruous condition, each emotion was combined with all the others (totally 5 incongruous pairs for each emotion). For each emotion, there were ten repetitions in the congruous condition and ten in the incongruous condition (totally 60 congruous couples and 60 incongruous couples). The neutral pair was reproduced 20 times.

Procedure
The experiment was conducted in a darkened, electrically shielded and sound-proof room, in order to warrant an optimal registration of the EEG signal. Subjects were in front of a monitor (90 cm of distance) connected to a PC. Visual stimuli were projected at the center of the monitor (visual horizontal angle 4°; visual vertical angle 6°) through the stimulation software STIM 4.2. A fixation point (white cross on black ground) appeared at the centre of the screen. Auditory stimuli sequence was reproduced through the stimulator, with sound emission by two loud-speakers incorporated in the PC, oriented towards the subject and located at about 90 cm of distance from him/her. Visual and auditory stimuli were synchronised, detecting the coincidence of the onset. The sequence of congruous and incongruous couples was randomized, avoiding consecutive identical pairs. The stimulation duration was about 10 minutes. Subjects were required to pay attention and to understand the stimuli, without any further indication about decoding modalities or stimulus emotional content. Moreover, they were required to maximally reduce body and eye movements, in order to improve EEG recording. The familiarization with the experimental procedure was guaranteed by a pre-experimental session, with the presentation of a sequence of ten stimulus couples (5 congruous and 5 incongruous) similar to that presented during the experimental phase. Finally, in order to test the correct comprehension of the emotional patterns, after the ERP registration subjects were required to evaluate:

a) For the vocal stimuli, the emotional content evaluation (disgust 76.7%; happiness 96.7%; fear 93.1%; anger 80.0%; surprise 90.0%; sadness 83.3 %; neutral 93.3%); the consistency of judgment (disgust M = 4.1; SD = 0.8; happiness M = 4.5; SD = 0.6; fear M = 4.7; SD = 0.7; anger M = 4.2; SD = 0.8; surprise M = 4.3; SD = 0.7; sadness M = 3.8; SD =0.9; neutral M = 4.1; SD = 1.0).
b) For the visual stimuli, the emotional content evaluation (disgust 78.0%; happiness 83.3%; fear 93.3%; anger 76.7%; surprise 73.3%; sadness 83.3%; neutral 80.0%); the consistency of judgment (disgust M = 3.9;

SD = 0.9; happiness *M* = 4.0; *SD* = 1.1; fear *M* = 4.1; *SD* = 0.8; anger *M* = 3.9; *SD* = 1.0; surprise *M* = 3.8; *SD* = 0.9; sadness *M* = 4.4; *SD* = 0.6; neutral *M* = 4.1; *SD* = 0.9).

ERP registration procedure

The EEG was recorded with a 62-channel DC amplifer (SYNAMPS system) and acquisition software (NEUROSCAN 4.2). An ElectroCap with Ag/AgCl electrodes were used to record EEG from 62 active scalp sites referred to earlobe (10/20 system of electrode placement, Jasper, 1958). Additionally two EOG electrodes were sited on the outer side of the eyes. The data were recorded using sampling rate of 256 HZ, with a frequency band of 0.1 to 60 Hz. The impedance of recording electrodes was monitored for each subject prior to data collection and it was always below 5 kΩ. After EOG correction and visual inspection only artefact-free trials were considered. An averaged waveform (off-line) was obtained from about forty artifact-free (trials exceeding 50 μV in amplitude were excluded from the averaging process) individual target stimuli for each type of sentence (*grandaverage*). Only fourteen electrodes were used for the successive statistical analysis (four central, Fz, Cz, Pz, Oz; ten lateral, F3, F4, C3, C4, T3, T4, P3, P4, O1, O2).

3. STATISTICAL ANALYSIS AND RESULTS

3.1. Phase 1: General ERP effect

Data analysis included two different steps: ERPs analysis; behavioural data (RT) analysis. As regard to the first analysis, two further phases need to be distinguished: (1) a qualitative analysis regarding ERP morphological profile; (2) a quantitative analysis regarding the two dependent variables of peak amplitude and latency.

ERPs data

1) Morphological analysis. Firstly we considered the sweep profile, that revealed the presence of negative and positive peak deflections. Between the others, it is possible to observe a positive deflection (P1) and a negative deflection (N1) related to emotional stimulus perceptive elaboration. Previous studies pointed out the meaning of these deflections for cross-modal perception (see the introduction). As regard to the successive ERP variations, sweep profile shows a negative (N2) and a positive (P2) deflection at about 200 ms post-stimulus. It is interesting to note that they are not homogenously distributed on the scalp, but they are localized in some specific areas. Taking into account the objectives of the present study, we focused the following quantitative analysis on a later latency time-window (150-250 ms post-stimulus).

2) Quantitative analysis. For each of the two ERP variations (N2 and P2 effects) a repeated measures ANOVA was applied to the dependent variables of the *peak amplitude* and *latency*, with the two within-subject factors of condition (congruous/incongruous) and cortical localization (Fz/Cz/Pz/Oz) (for effects with degree of freedom > 1 a Greenhouse-Geiser correction was applied).

We report the data distinctly for peak amplitude (a) and peak latency (b). Due to the amount of data, only the significant effects were reported.

a) Peak amplitude

N2 variation. Repeated ANOVA showed a significant effect for localization, (F (3, 30) = 11.46, p = .001), but not for condition. Moreover, a significant interaction effect was observed, (F(6, 30) = 8.10, p = .001). Particularly (Fig. 1) post-hoc analysis (contrast analysis) applied to the main localization effect revealed a significant difference between frontal and posterior areas (Fz vs Pz: p =.001; Fz vs Oz: p=.001).

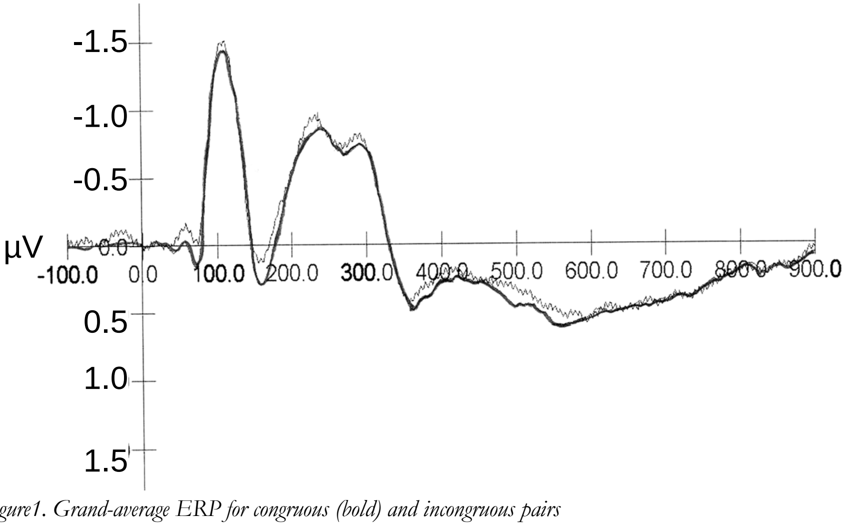

Figure1. Grand-average ERP for congruous (bold) and incongruous pairs

On the whole, N2 deflection was more localized on the anterior scalp area. Furthermore, post hoc comparisons for the interaction effect showed the presence of difference in peak amplitude for congruous and incongruous conditions as a function of localization. Indeed, N2 effect was more localized in the frontal areas for incongruous stimuli compared with congruous ones (p = .001), while in the other cortical areas the peak was homogeneously distributed.

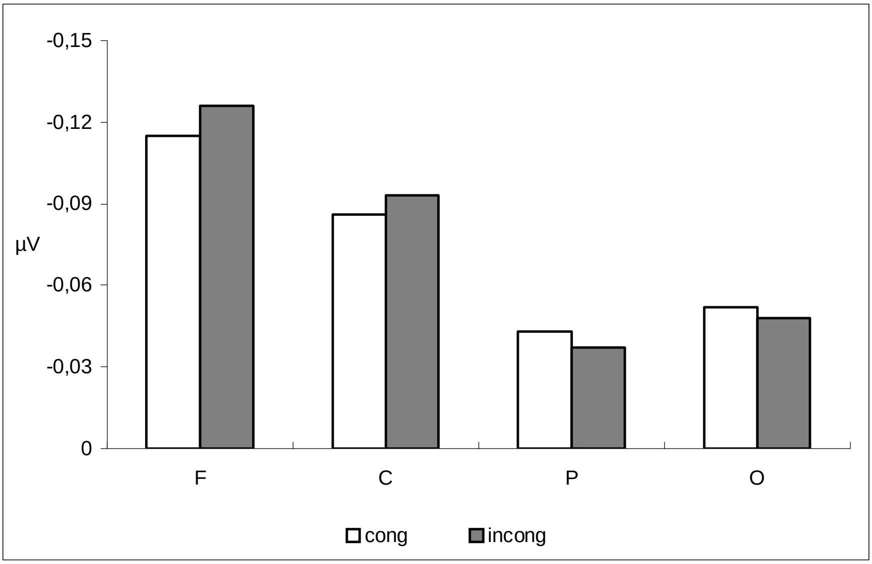

Figure 2. N2 peak amplitude in different cortical areas, as a response to congruous and incongruous condition

P2 variation. ANOVA showed the significance of the localization effect (*F* (3, 30) = 11.72, *p* = .001), with an ampler peak in the parietal (Pz) area, and of the condition effect (*F*(1, 30) = 8.90, *p* = .001). As showed by the contrast analysis, other scalp areas are less activated (Fz vs Pz: *p* =.001; Cz vs Pz: *p* =.001; Oz vs Pz: *p* =.001) (Fig. 2), and congruous condition produced a higher deflection than incongruous condition in Pz (p = .001).

b) Peak latency

N2 variation. As revealed by repeated measures ANOVA, there were not latency differences between congruous and incongruous conditions, as well as for condition x localization interaction. On the contrary, localization was significant (*F* (3, 30) = 5.60, *p* = .001). As indicated by the contrast analysis, N2 is delayed in the areas where its amplitude is decreased. Specifically, latency times progressively increased proceeding from the anterior (Fz) to the posterior sites (Fz vs Pz: *p* = .001; Fz vs Oz: *p* = .006).

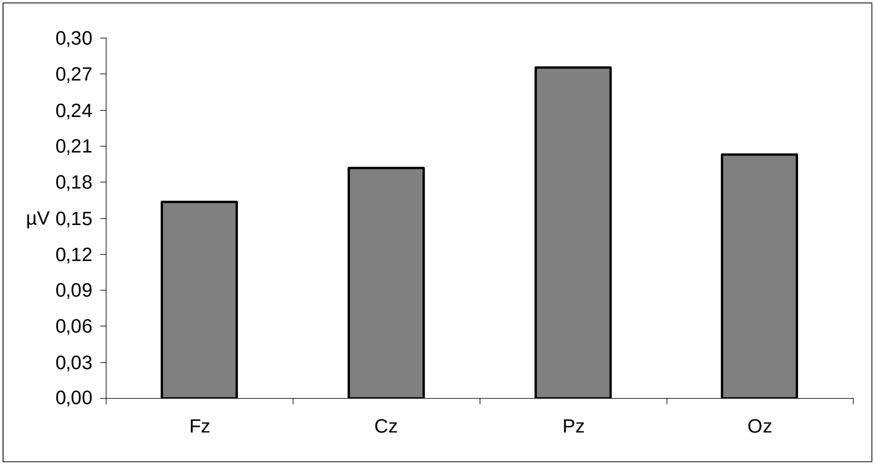

Figure 3. P2 peak amplitude in the frontal, central, parietal and occipital areas

P2 variation. As regard to latency, only localization effect was significant (*F* (3, 30) = 6.89, *p* = .001), with a peak delayed for Pz and Oz (in particular for the contrast Fz vs Oz *p* =.006). Thus, it is clear that the ampler peak in the posterior areas corresponds to an increasing of its latency.
Behavioural data (RTs). A repeated measures ANOVA was applied in order to explore the congruous/incongruous effect on behavioural response times. The analysis showed that the two experimental conditions did not differ as a function of the RT. Specifically we observed a quite similar RT for congruous (M = 1.32 ms; *SD* = 0.40), and incongruous (M = 1.31; *SD* = 0.40) stimuli.

3.2. Phase 2: Emotion-by-emotion analysis

ERPs data. In order to explore the effect of the emotional patterns on ERP variations, a further analysis set was applied to ERPs and behavioral data to each emotion separately.

a) Peak amplitude

N2 variation. A first repeated measures ANOVA (emotion 6, x localization 4) was applied to congruous emotional patterns, in order to detect possible differences between emotional correlates. The localization (*F* (3, 30) = 7.75, *p* = .001), and emotion x localization interaction (*F* (15, 30) = 2.23, *p* = .005), effects were significant. Specifically, post hoc (contrasts) analysis indicated that anger has a main frontal localization if compared with other localizations, as well as for happiness, fear and disgust. On the contrary, sadness and surprise showed a more homogeneous distribution on the different cortical areas and, in general, a less high peak than the other emotion (Fig. 4).

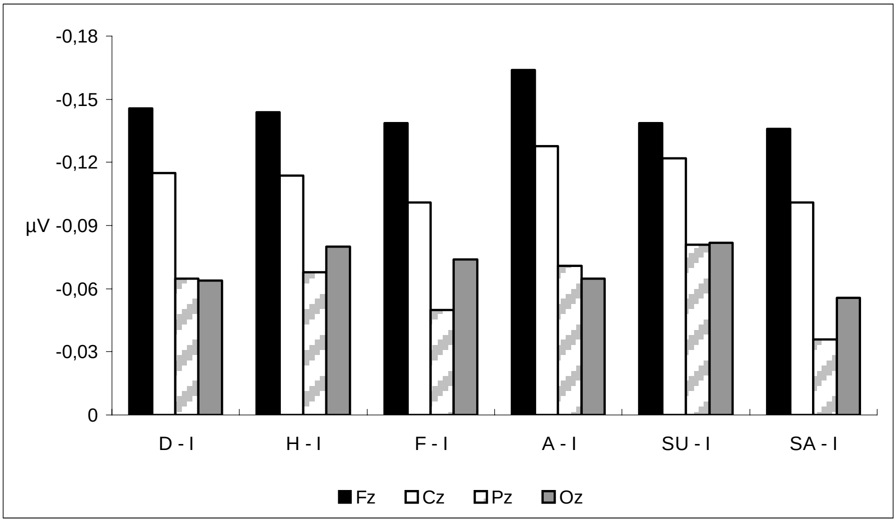

Figure 4. N2 peak amplitude as a function of emotion and localization variables

Successively, each emotion was submitted to repeated measures ANOVA, with independent factors condition (2) and localization (4). We reported each analysis.

- *Disgust.* As revealed by the significant effect of localization (*F* (3, 30) = 5.02, *p* = .033), and by the post-hoc comparisons, N2 peak amplitude was more accentuated in the frontal areas, compared to posterior areas. Nevertheless the condition effect was not significant.
- *Happiness.* Even for happiness, localization effect was significant (*F* (3,

30) = 6.25, *p* = .001), but not the condition, in the same direction as disgust, with a peak increasing for Fz more than the others sites.

- *Sadness.* None of the main effects showed the statistical significance. In particular sadness showed a homogeneous peak variation for the experimental conditions and the localization.
- *Surprise.* Main effects of condition (*F* (1, 30) = 23.55, *p* = .001), localization (*F* (3, 30) = 4.10, *p* = .009), as well as the interaction (*F* (3, 30) = 3.85, *p* = .012), were significant. Particularly, contrast analysis showed a N2 increasing for congruous then incongruous condition and for posterior area of the scalp.
- *Anger.* Localization revealed a significant effect (*F* (3, 30) = 10.80, *p* < .001). Particularly, N2 deflection had a maximum amplitude in the frontal areas, for both congruous and incongruous conditions.

P2 variation. The first screening (repeated measures ANOVA for congruous condition across the emotions) showed a substantial homogeneity between the six emotions and, on the contrary, significant differences for the localization (*F* (3, 30) = 9.57, *p* =.001). Contrast analysis revealed an increased peak amplitude for parietal and occipital areas then for anterior and central ones (Fig. 5).

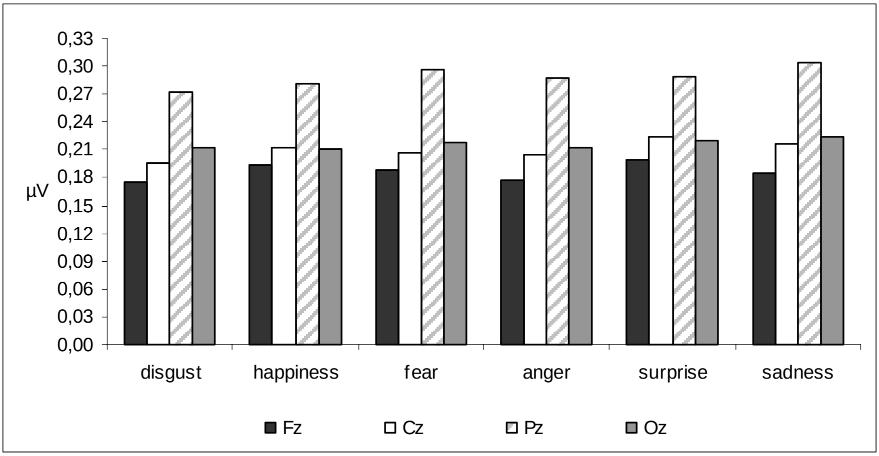

Figure 5. P2 peak amplitude as a function of emotion and localization variables

- *Disgust.* Only the main effect of localization was significant (*F* (3, 30) = 8.46, *p* = .001). P2 amplitude was higher for the posterior areas, as found for the general (see previous) analysis.
- *Happiness.* Both condition (*F* (1, 30) = 5.90, *p* = .022) and localization (*F*

(3, 30) = 6.63, *p* = .001), effect were significant. Particularly, condition analysis revealed an increased peak for congruous patterns. Moreover, P2 had its maximum value in the posterior areas (Oz and Pz).

- *Fear.* Main effect of condition was significant, with an increasing of the peak for congruous stimuli (*F* (1, 30) = 27.51, *p* < .001). Moreover, P2 was mainly distributed on the posterior localization.
- *Sadness.* Similarly to the other emotions, condition (*F* (1, 30) = 24.65, *p* = .001), and localization (*F* (1, 30) = 23.55, *p* = .001), effects were significant. A higher amplitude of P2 was registered for congruous patterns in the posterior sites.
- *Surprise.* P2 deflection was significantly affected by condition (*F* (1, 30) = 19.1, *p* = .001), and localization (*F* (3, 30) = 7.04, *p* = .001). Specifically, contrast analysis indicated a peak amplitude increasing for congruous condition and for posterior localizations.
- *Anger.* Condition (*F* (1, 30) = 6.22, *p* = .019), and localization (*F* (3, 30) = 11.30, *p* <. 001), differentiated the ERP profile. P2 deflection, indeed, has its maximum value in response to congruous stimuli and in the posterior areas.

b) Peak latency

N2 variation. The first general analysis revealed a localization effect (*F* (3, 30) = 6.09, *p* = .001). Post-hoc comparisons indicated that N2 is delayed in posterior areas, if compared with anterior ones and, above all, with occipital areas for all the emotions. The ANOVAs applied to each emotion showed analogous data. In particular, all emotions presented a delayed latency in posterior lobes, independently from the condition, for disgust (*F* (1, 30) = 3.21, *p* = .02), happiness (*F* (3, 30) = 5.17, *p* = .002), fear (*F* (3, 30) = 4.28, *p* = .004), anger (*F* (3, 30) = 6.61, *p* < .001), surprise (*F* (3, 30) = 7.01, *p* = .013), and sadness (*F* (3, 30) = 5.43, *p* = .002).

P2 variation. The ANOVA applied to the six emotions displayed a localization effect (*F* (3, 30) = 8.02, *p* =.001). In particular, contrast analysis indicated a retardation in posterior areas (Pz and Oz), as compared to central and anterior areas. Moreover, the significant interaction effect, *F* (3, 30) = 2.11, *p* =.008, showed differences between emotions as a function of the localization. Specifically, fear and anger showed a delayed peak in the posterior sites of the scalp, while the other emotions had a more homogeneous profile in terms of cortical distribution. Taking into account each emotion, we observed the main retardation of peak onset for congruous condition in comparison with in the incongruous one, respectively for disgust (*F* (1, 30) = 13.75, *p* = .001); happiness (*F* (3, 30) = 11.74, *p* = .002); fear (*F* (3, 30) = 10.70, *p* = .003); anger (*F* (3, 30) = 17.70, *p* < .001); surprise (*F* (3, 30) =

29.02, p = .001); (sadness F (3, 30) = 24.15, p <.001).

Behavioural data. In order to explore the presence of differences between the six congruous emotional patterns, a repeated measure ANOVA was applied to behavioural data. As showed in Fig. 6, the analysis indicated that the six emotions differed (F (3, 30) = 32.90, p < .001). In particular, as revealed by contrasts effects, RTs were slower for fear, anger and surprise, while sadness and disgust presented reduced response times. Happiness showed intermediate values if compared with the others emotions (no significant differences).

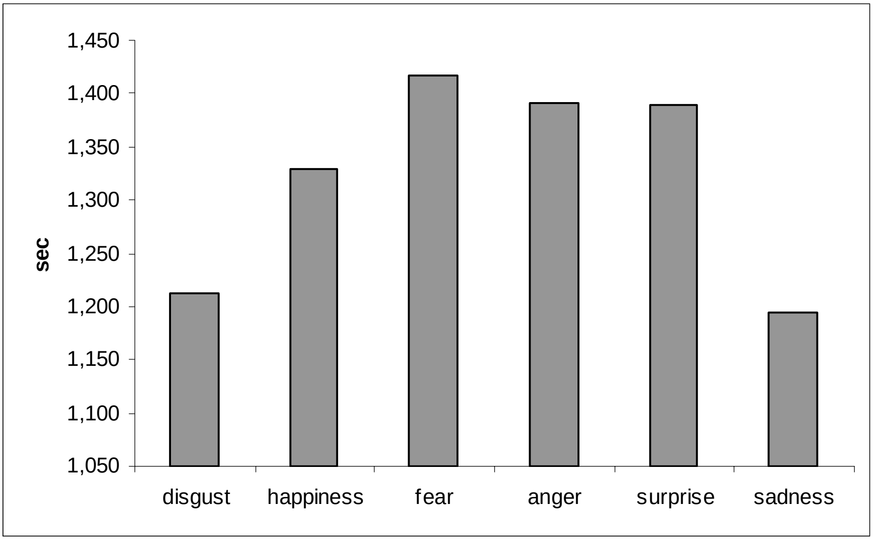

Figure 6 . Response times (RTs) for congruous pairs

In a second set of ANOVAs, each emotion was analyzed comparing congruous and incongruous condition. Statistical results indicated that for all emotions, except for happiness, condition effect was significant. Nonetheless, descriptive data revealed an opposite trend for two emotional subcategories. While fear (F (1, 30) = 26.23, p < .001), anger (F (1, 30) = 16.26, p < .001), and surprise (F (1, 30) = 14.58, p = .001), revealed an increased RT for congruous condition, sadness (F (1, 30) = 43.51, p < .001) and disgust (F (1, 30) = 22.53, p < .001) showed an opposite trend, with delayed response for incongruous condition.

4. DISCUSSION

Based on our experimental results, we can summarize four main points. In particular, they concern: 1) ERP profile differentiation as a function of the *cortical localization*; 2) The systematic presence of *long-latency ERP effect*, that was supposed to be the cognitive marker of the cross-modal integration; 3) The relevance of this cortical index, able to differentiate *congruous* and *incongruous patterns*; 4) A different subjects' response as a function of the specific *emotional correlates*.

Firstly, as revealed by ERP morphological analysis applied to a long-latency time-window, there were two different deflections, one positive (P2) and one negative (N2). They were characterized by a heterogeneous distribution on the scalp. Specifically, P2 was mainly localized in the parietal areas, since it arose its higher amplitude at the back site of the scalp (Pz). An analogous differentiation was observed for N2 peak variation, with a different distribution between anterior and posterior areas. Indeed, it was more localized in anterior (Fz) rather then posterior (Pz and Oz) cortical sites. These differences in ERP localization suggested that the two deflections may account for *different cognitive phenomena*. This hypothesis was greatly supported by the different sensitivity shown by the two ERP deflections to localization in interaction with the condition effect. In fact, the increasing of the discrepancy between anterior/posterior localization in the incongruence condition for both P2 and N2 is a main point to be discussed.

Moreover, in the case of P2, data revealed a peak amplitude variability as a function of experimental condition, with increased values for congruous patterns. On the contrary, N2 presented a similar peak profile for both conditions. P2 sensitivity to congruence condition suggests that this ERP variation could be a long-latency index of multisensory integration process for emotional stimuli. Specifically, it would be a relevant marker of the redundant and convergent information processing from multiple emotional channels (visual and auditory). Instead, its reduction in case of incongruence would account for an unsuccessful convergence of information supported by face and voice. On the contrary, N2 deflection did not reveal to be directly related to cross-modal integration in case of convergence or divergence. Indeed, it appeared to keep the same profile in both experimental conditions. Although being out of the main interest of cross-modal integration analysis, we can state that this variation could be related to the emotional content of the stimuli. Indeed, previous studies demonstrated its sensitivity to the emotional patterns (types of emotion), as a function of some parameters, such as the arousal parameter (Balconi & Pozzoli, 2003; Balconi, 2004a; Sato *et al.*, 2001).

Another main result of the present research is related the ERP *latency effect*. In comparison with previous research (Pourtois *et al.*, 2002), our study revealed that latency of the peak was not differentiated as a consequence of the experimental condition, but only in terms of localization. It suggested that, mainly for the P2 ERP effect, incongruence condition doesn't require a greater amount of time to be processed. At this regard, it is important to note that peak delay for incongruous stimuli observed in previous studies was found for early (perceptual) and not for later (cognitive) components. A further result in line with this trend concerns RTs. Indeed, the time required to process incongruous patterns was the same of the congruous ones. Subjects' ability to early decode the incongruous patterns likely responds to a basic ability that allow the subjects to detect in an automatic and sudden way cognitive discrepancy in emotional information processing from different communicative channels. On the other hand, it must be underlined that congruous information comprehension doesn't represent a facilitation condition for the integrative process. In other words, this "facilitation effect" would be related to the first phase of stimulus elaboration (perceptive phase) but not the successive cognitive mechanisms.

On the basis of these results it is useful to analyze the effect of the single emotional pattern on the ERP profile variations. In fact, emotion-by-emotion data analysis allowed us to explore the cognitive significance of such ERP variations for cross-modal integration. In particular, we observed a N2 modification as a function of the different emotions, independently from the experimental condition. A substantial dichotomy was observed between a first group of emotions, including anger, fear, surprise and disgust and a second group represented by happiness and sadness. In the first case, the negative deflection was characterized by a higher peak amplitude. As suggested by previous studies, N2 could be a valid index of the emotional content of the stimulus in terms of arousal (high-low) and hedonic valence (positive-negative) parameters (Balconi, 2004b; Balconi & Pozzoli, 2003). The different profile of the ERPs as a function of the emotional valence may indicate the sensitivity of the negative-wave variation to the "semantic" value of the pattern. Similar ERP profiles were observed for happy and sad expressions, differentiated from the negative high-arousal emotions (fear, anger and surprise). It has been suggested that subjects reaction to negative and high-arousal emotional patterns could be more intense (emotional response) than less involving emotions (Lang, Nelson & Collins, 1990).

An opposite profile characterized P2. Primarily, it was revealed to be homogeneously distributed in a specific and circumscribed posterior area (Pz). Secondly, on the whole it was sensible to condition but not to the emotional content. Indeed, it is was not differentiated across the emotions. Rather it represented a cognitive marker of the congruence/incongruence of the emotional patterns: it was reduced in amplitude in response to incongruous stimuli and, on the contrary, increased in response to congruous stimuli. Previous researches revealed an analogous "amplification effect" of a positive-going wave (P2b) in case of congruous patterns (Pourtois *et al.*, 2002), generally subsequent to the early-latency ERP effects (N1-P1) and preceding a later-latency positive deflection (P3).

Nevertheless, as already underlined, previous studies applied their analysis exclusively to a narrow range of emotional correlates (generally one positive and one negative emotion), fact that did not allow to exclude the existence of a relation between P2 effect and the emotional content of the stimuli. In the present study, thanks to the ampler range of emotions, we can exclude with greater confidence an effect due to emotional content, and confirm the "cross-modal integration" significance of this effect. Moreover, the cognitive meaning of P2 can be compared to the role of an analogous ERP variation, the N4 effect. This late component was previously observed in response to conditions of semantic anomaly or in case of the necessity to reconfigure the semantic context because of stimulus-background divergence, independently from perceptive modality (it was observed for images and words) (Balconi & Lucchiari, 2005; Holcomb & McPherson, 1994). Thus, it is useful to compare P2 and N4 variations, since they seem to be correlated to the congruence of the stimulus, although in an antithetic manner. In the first case, N4 is indicative of the semantic incongruence (incongruence marker), while P2 signals the semantic convergence (congruence marker) between cross-modal information.

Finally, as regard to behavioural data (RT), we previously underlined the absence of a congruence/incongruence effect *per se* (absence of a RT retardation for incongruence; absence of a RT facilitation for congruence). Nonetheless, it is interesting to notice the presence of different response profiles for each emotion, even if in an unattended way. Indeed, "redundancy effect" produced by channel congruence seems to engender an "interference effect" for some patterns (fear, anger and surprise) and a "facilitation effect" for other emotions (above all sadness). Only in this second case, RTs were reduced in case of congruence. It is possible to suppose the presence of a cognitive facilitation in the elaboration of emotions which have a low degree of salience (low arousal level) and an opposite, distracting effect, for emotions which have a greater emotional impact (high arousal level). Paradoxically, in the last case, subjects presented longer response times in case of congruence, since we suppose that for these emotions subjects are more able to rapidly detect incongruence, and on the contrary that they are hindered in their comprehension if information is redundant and not required (such as in case of congruence). In other words, these emotions would be better recognized at minimum informational level but not at maximal (and exceeding) informational level. An opposite trend is supposed for the second type of emotions (later acquired in the development), such as sadness, which could benefit from an "informational reinforce", guaranteed by channel redundancy.

REFERENCES

Balconi, M. (2006). *Neuropsychology and cognition of emotional face comprehension*. Kerala: Research Signpost.

Balconi, M. (2004a). *Neuropsicologia delle emozioni*. Roma: Carocci.

Balconi, M. (2004b). Decoding dell'espressione facciali delle emozioni, arousal a valenza edonica. Analisi mediante potenziali evocati corticali (ERPs). *Ricerche di Psicologia*, *2*, 7-34.

Balconi, M., Carrera, A. (2006). Lessicalizzazione e concettualizzazione nella comprensione della mimica emotiva. Analisi mediante script emotivi. *Età Evolutiva*, *83*, 18-29.

Balconi, M., Lucchiari, C. (2005). Consciousness, emotion and face: An event-related potentials (ERP) study. In R.D. Ellis & N. Newton (Eds.), *Consciousness and emotion. Agency, conscious choice, and selective perception* (pp. 121-135). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Balconi, M., Pozzoli, U. (2003). Face-selective processing and the effect of pleasant and unpleasant emotional expressions on ERP correlates. *International Journal of Psychophysiology*, *49*, 67-74.

Callan, D.E., Callan, A.M., Kroos, C., Vatikiotis-Bateson, E. (2001). Multimodal contribution to speech perception revealed by independent component analysis: A single-sweep EEG case study. *Cognitive Brain Research*, *10*, 349-353.

Calvert, G.A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*, *11*, 1110-1123.

Calvert, G.A., Brammer, M.J., Iversen, S.D. (1998). Crossmodal identification. *Trends in Cognitive Sciences*, *7*, 247-253.

Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, M., Williams, S.C.R. (1997). Activation of auditory cortex during silent lipreading. *Science*, *276*, 593-596.

Calvert, G.A., Campbell, M., Brammer, R.J. (2000). Evidence from functional magnetic resonance imaging in crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*, 649-657.

De Gelder, B. (2000). Recognizing emotions by ear and by eye. In R. D. Lane & L. Nadel (Eds.), *Cognitive neuroscience of emotion* (pp. 84-105). New York: Oxford University Press.

De Gelder, B., Böcker, K. B. E., Tuomainen, J., Hensen, M., Vroomen, J. (1999). The combined perception of emotion from voice and face: Early interaction revealed by human electric brain responses. *Neuroscience Letters*, *260*, 133-136.

De Gelder, B., Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion*, *14*, 289-311.

De Gelder, B., Vroomen, J., Bertelson, P. (1998). Upright but not inverted faces modify the perception of emotion in the voice. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, *17*, 1021-1031.

Dolan, R. J., Morris, J. S., de Gelder, B. (2001). Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Science*, *98*, 10006-10010.

Driver, J., Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology*, *10*, 731-735.

Ekman P., Friesen W.V. (1976). *Pictures of facial affects.* Consulting Psychologist Press, Palo Alto.

Fort, A., Delpuech, C., Pernier, J., Giard, M.-H. (2002). Early auditory-visual interactions in human cortex during nonredundant target identification. *Cognitive Brain Research*, *14*, 20-30.

Ghazanfar, A.A., Maier, J.X., Hoffman, K.L., Logothetis, N.K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *Journal of Neuroscience*, *25*, 5004-5012.

Giard, M.H., Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, *11*, 473-490.

Graziano, M.S., Hu, X.T., Gross, C.G. (1997). Visuospatial properties of ventral premotor cortex. *Journal of Neurophysiology*, *77*, 2268-2292.

Hadjikhani, N., Roland, P.E. (1998). Cross-modal transfer of information between the tactile and the visual representations in the human brain: A positron emission tomographic study. *Journal of Neuroscience*, *18*, 1072-1084.

Holcomb, P.J., McPherson, W.B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition*, *24*, 259-276.

Johnstone, T., van Reekum, C.M., Oakes, T.R., Davidson, R.J. (2006). The voice of emotion: an fMRI study of neural responses to angry and happy vocal expressions. *Social Cognitive and Affective Neuroscience*, *1*, 242-249.

Lang, S.F., Nelson, C.A., Collins, P.F. (1990). Event-related potentials to emotional and neutral stimuli. *Journal of Clinical and Experimental Neuropsychology*, *12*, 946-958.

Levänen, S., Sams, M. (1997). Disrupting human auditory change detection: Chopin is superior to white noise. *Psychophysiology*, *34*, 258-265.

Macaluso, E., Frith, C.D., Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science*, *289*, 1206-1208.

Massaro, D. W., Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review*, *3*, 215-221.

MacLeod, C.M., MacDonald, P.A. (2000) Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention, *Trends in Cognitive Sciences*, *4*, 383-391.

Mesulam, M.M. (1998). From sensation to cognition. *Brain*, *121*, 1013-1052.

Näätänen, R. (1992). *Attention and brain function*. Mahwah: Lawrence Erlbaum Associates.

Pourtois, G., Debatisse, D., Despland, P.-A., de Gelder, B. (2002). Facial expressions modulate the time course of long latency auditory brain potentials. *Cognitive Brain Research*, *14*, 99-105.

Pourtois, G., de Gelder, B., Bol, A., Crommelinck, M. (2005). Perception of facial expressions and voices and of their combination in the human brain. *Cortex*, *41*, 49-59.

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport*, *11*, 1329-1333.

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

Rugg, M.D., Coles, M.G.H. (1997). *Electrophysiology of mind: Event-related brain potentials and cognition*. Oxford: Oxford University Press.

Sato, W. Takanori, K., Sakiko, Y., Michikazu, M. (2001). Emotional expression boots early visual processing of the face: ERP recording and its decomposition by independent component analysis. *Neuroreport*, *12*, 709-714.

Scherer, K.R. (1984). On the nature and function of emotion; A component process approach. In K.R. Scherer, P. Ekman (Eds.), *Approaches to emotion* (pp. 293-317). Hillsdale: Erlbaum.

Schroeder, C.E., Foxe, J.J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cognitive Brain Research*, *14*, 187-198.

Schroeder, C. E., Smiley, J., Fu, K. J., McGinnis, T., O'Connell, M. N., Hackett, T. A. (2003). Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. *International Journal of Psychophysiology*, *50*, 5-17.

Surakka, V., Tenhunen-Eskelinen, M., Hietanen, J. K., Sams, M. (1998). Modulation of human auditory information processing by emotional visual stimuli. *Cognitive Brain Research*, *7*, 159-163.

Teder-Sälejärvi, W. A., McDonald, J. J., Di Russo, F., Hillyard, S. A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research*, *14*, 106-114.