Research report

# Facial expressions modulate the time course of long latency auditory brain potentials

Gilles Pourtois[a,b], Damien Debatisse[b], Paul-Andre Despland[c], Beatrice de Gelder[a,b,*]

[a]*Donders Laboratory for Cognitive and Affective Neuroscience, University of Tilburg, Tilburg, The Netherlands*
[b]*Laboratory of Neurophysiology, University of Louvain, Brussels, Belgium*
[c]*Neurology Unit, Centre Hospitalier Universitaire Vaudoix, Lausanne, Switzerland*

## Abstract

Long latency auditory brain potentials were recorded while subjects listened to bi-syllabic words spoken with an emotional expression and concurrently viewed congruent or incongruent facial expressions. Analysis of the auditory waveforms suggests the existence of a positive deflection around 240 ms post-stimulus with a clear posterior topography (the P2b component). This potential is subsequent upon the modality-specific auditory N1–P2 components and precedes the amodal N2–P3 complex. Congruent face-voice trials elicited an earlier P2b component than incongruent trials suggesting that auditory processing is delayed in the presence of an incongruent facial context. These electrophysiological results are consistent with previous behavioural studies showing an acceleration of reaction times for rating voice expressions that are part of congruent bimodal stimulus pairs. A source localisation analysis performed on the scalp EEG during the time-window corresponding to the P2b component disclosed a single dipole solution in the anterior cingulate cortex. © 2002 Elsevier Science B.V. All rights reserved.

*Keywords:* Multisensory object recognition; P2b; Evoked potential; Audio-visual integration; Facial expression; Cross-modal bias

## 1. Introduction

Affective information is often communicated simultaneously by different sensory channels of communication such as the face and the voice. Perceiving emotions is thus among the many perceptual skills that have correlates in more than one sensory modality [6,20]. Do face and voice inputs integrate at early perceptual stages or at late decisional stages that is, when the respective visual and auditory processes are fully completed? Laboratory studies have begun to address the phenomenon of audio-visual integration of emotion signals using a cross-modal bias paradigm [20]. Recently, de Gelder and Vroomen [8] used bimodal voice-face pairs in which varying degrees of semantic incongruency were created between the expression in the face and in the emotional tone of voice.

Subjects were instructed to judge the facial expression or the tone of voice and results indicated perceptual integration of the two sources of information. Identification of the emotion in the face was biased in the direction of the simultaneously presented tone of voice [8]. The reverse effect, a bias from the emotion in the face on judgment of the emotion in the voice, was also obtained. These results strongly suggest the existence of mandatory bidirectional links between affect recognition structures in face and voice. In the present experiment, we address the issue of the time-course of face-voice integration of emotion using the high temporal resolution of event-related brain potentials and we explored its neural correlate using a source localisation model.

Electrophysiological studies (either EEG or MEG) of cross-modal identification [3] have clearly demonstrated large amplitude effects either consisting of increase or decrease of early unimodal components like the auditory N1 or the visual P1 component (each generated around 100 ms for stimulus presentation in their respective modality) in normal subjects during the perception of multimodal stimuli [25,16,24]. Increase or amplification of the neural

*Corresponding author. Department of Psychology—Donders Laboratory for Cognitive and Affective Neuroscience, University of Tilburg, PO Box 90153, 5000 LE Tilburg, The Netherlands. Tel.: +31-13-466-2167; fax: +31-13-466-2370.
*E-mail address:* b.degelder@kub.nl (B. de Gelder).

signal in modality-specific cortex (see Refs. [7,13] for a discussion) appears therefore as an important electrophysiological correlate of cross-modal integration. This has been found with brain imaging data for audio-visual stimuli (i.e. activation of the auditory cortex during lipreading [4]; activation of the fusiform gyrus and the amygdala during the perception of affective stimuli [10], and visuo-tactile stimulus pairs (i.e. activation of the lingual gyrus during visuo-tactile perception [19]).

Only a few studies have directly investigated the time-course of face-voice integration of emotions using the high temporal resolution of EEG recordings. In a previous study in which facial expressions were presented concurrently with sentence fragments, de Gelder et al. [9] reported that face-voice pairs elicited an extra negative component around 170 ms post stimulus that was compatible with the electrophysiological parameters (e.g. latency and topography) of the mismatch negativity (MMN) [22]. Subsequently evidence for integrated perception of facial expressions paired with sentence fragments has been provided in a study where the auditory N1 component was significantly enhanced by congruent pairs [23]. Finally, in a third study [29], the pitch MMN has been shown to be influenced by the simultaneous presentation of positive non-facial stimuli combined with pure tones. All these studies have shown that the amplitude of early auditory processing as indexed by the N1 or the MMN could be modulated by the simultaneous presentation of a concurrent visual stimulus. In the present experiment a different methodology was used and we studied whether a facial expression could also influence the latency of the concurrent auditory process. Event-related brain potentials were recorded while subjects heard auditory fragments paired with either a congruent or an incongruent facial expression (happiness or fear).

## 2. Methods

### 2.1. Subjects

Ten native French-speaking subjects all right-handed with an average age of 20 years participated in the study. They were paid for their participation. Two subjects were not included in the statistical analyses because of excessive alpha band in the EEG.

### 2.2. Stimuli and procedure

Stimuli consisted of an auditory and a visual component. Visual materials consisted of six black and white static pictures taken from the Ekman series [14]. For each actor a happy and a fearful facial expression were selected yielding a total of 12 visual stimuli. Auditory materials consisted of 12 bi-syllabic spoken words which were obtained with the following procedure. Six male and six female actors all native French speakers were instructed to pronounce a neutral sentence ('they are travelling by plane') in an emotional tone of voice (either happy or fearful). Speech samples were recorded on a DAT recorder and subsequently digitized and amplified (using SoundEdit 16 1.0 b4 running on Macintosh). From these, only the tokens of the last word 'plane' were selected using SoundEdit. A total of 24 samples (12 actors×2 tones of voice) were obtained. In a pilot study, they were presented to eight volunteers (four males and four females) none of which participated in the experiment. They were instructed to label each fragment as 'happy', 'fearful', or 'don't know'. Based on recognition rates, 12 fragments (i.e. six actors pronouncing 'plane' either in a happy or fearful tone of voice) were selected (mean recognition rate 74% correct). Each spoken fragment was then combined with a facial expression in order to construct audio-visual stimulus pairs with either a congruent (12 pairs) or an incongruent affective content (12 pairs). Trials were generated using the Stim software running on a PC Pentium II. Sounds were delivered over two loudspeakers placed on each side of the screen at a mean sound level of 72 dB. Mean size of the face pictures was 6 cm width by 8 cm height. Mean luminance was 25 cd/m$^2$ and less than 1 cd/m$^2$ for the background and ambiant light.

A trial started with the presentation of the face. After a delay (900 ms) following the onset of the face, the voice fragment (duration 381±50 ms) was presented while the face picture remained on the screen till the end of the voice fragment. The delay between voice and face onsets was introduced in order to reduce interference of the visual brain response elicited by the faces. Intertrial interval (measured from the offset of the visual stimulus) was 0.5 s. Subjects were tested in a dimly lit room seated at 80 cm away from the screen. They were instructed to pay attention to the voices but no behavioral data were collected to avoid any response-related components in the ongoing EEG. Eight blocks of 24 audio-visual trials (half congruent and half incongruent) were presented randomly. The 12 fragments were identical for the congruent and incongruent conditions (i.e. the only difference was the emotion of the face they were paired with).

### 2.3. Electrophysiological recording and data processing

Auditory event-related brain potentials (AEPs) were recorded and processed using a Neuroscan 32 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 30 electrodes mounted in a quickcap (10–20 System) with a linked-mastoids reference, 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. After removal of EEG and EOG artefacts (epochs with EEG or EOG exceeding ±70 μV were excluded from the averaging), epoching was made 100 ms prior to auditory

stimulus onset and for 924 ms after stimulus presentation. Epochs were re-referenced off-line to a common average reference. Data were low-pass filtered at 30 Hz. Maximum amplitudes and mean latencies of AEPs were measured relative to a 100 ms pre-stimulus baseline and assessed using repeated measures Analyses of Variance (ANOVAs). Analyses were focused on the posterior positive deflection subsequent upon the auditory P2 component but elicited before the N2b–P3 complex. For each individual subject and each condition, the latency and amplitude of the P2b component were measured relative to the maximum positivity at three posterior electrodes (CP3, CPz and CP4) in the 210–270 ms interval. Source estimation [26] was performed on the time-window of the auditory waveforms (congruent and incongruent averaged together) corresponding to the P2b component (i.e. time window of 30 ms from 215 to 245 ms post-stimulus). A single unconstrained dipole solution (Advance Source Analysis) was first computed and compared with a second solution providing two symmetric dipoles (symmetry constraint) that were localized in the temporal lobe in a cortical region (i.e. the middle temporal gyrus, Brodmann area [21]

known to be involved in the perception of audio-visual events [21]. Moreover, a third solution providing a pair of dipoles with the mirror symmetry constraint only (that allows therefore for some movement of the pair of dipoles) was also computed. The dipole localization method is based on the single equivalent current dipole model which provides electrode locations in the 10–20 standard or more. A three-layer realistically shaped head model, describing the brain, skull, and scalp compartment, is applied in combination with the boundary-element method. The localization produces a set of dipole parameters that matches EEG forward solution at a given latency.

## 3. Results

A detailed analysis of the auditory waveforms at posterior leads suggests the existence of a positive deflection following the N1–P2 complex and labeled here the P2b component (Fig. 1). Analyses were focused on this positive deflection (220–260 ms post-stimulus) which follows
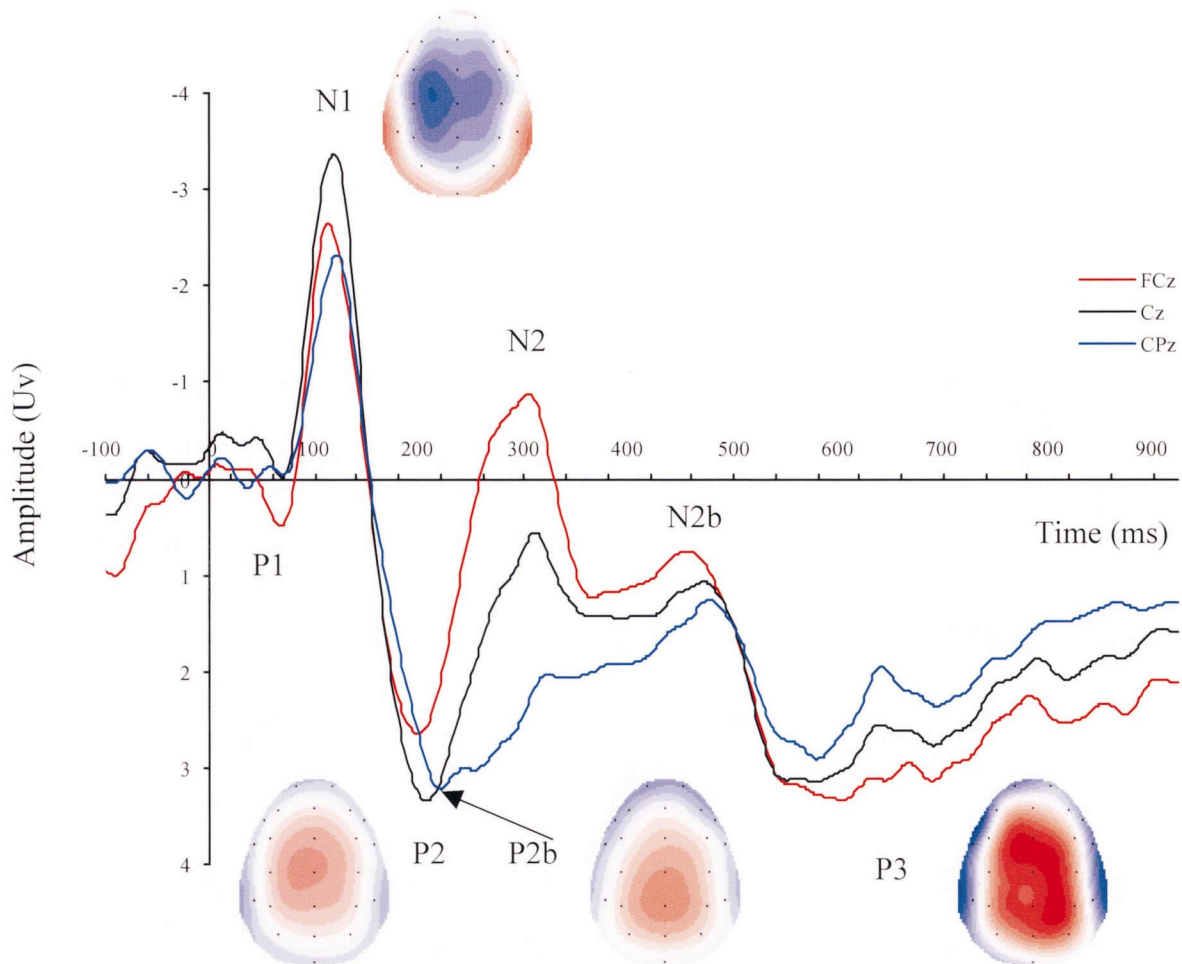


Fig. 1. Grand averaged auditory waveforms (congruent trials) and corresponding topographies (horizontal axis) at FCz (frontal), Cz (central) and CPz (parietal) electrodes. The latency for the N1 topography is 116 ms, 188 ms for the P2, 224 ms for the P2b and 580 ms for the P3. For each topographical map, the time interval is 20 ms and the amplitude scale goes from $-6$ $\mu$V (blue color) to $+6$ $\mu$V (red color).

the N1–P2 modality-specific components post auditory stimulus (as well as the MMN) but is elicited before the N2b–P3 complex. The N100 (100–120 ms) and P200 (190–210 ms) exogenous components are late cortical auditory components [17] with a fronto-central topography, each composed of multiple subcomponents. The N1 is modulated by auditory selective attention (i.e. enlarged N1 elicited by attended stimuli). In the present study, the N1 component is not modulated by face congruency (Fig. 2) and the methodology used was not appropriate to study the MMN. The N2b–P3 complex [27] occurs at a later latency range than the P2b component, when modality-specific processing is completed. This amodal complex would represent the updating of working memory (see Refs. [12,30] for a discussion). The P2b is clearly separate from the P2 component (see Fig. 1 and Table 1) because of its later time window and its more posterior topography (maximum at electrode CPz). The P2b component is also different from the P3 component (e.g. the P3b) because of the non-overlapping time-windows between these two positive deflections.

Analysis of the waveforms at posterior leads (CP3, CPz and CP4) comparing congruent and incongruent trials (see Fig. 2) shows a strong latency effect on the P2b com-

Table 1
Latency (in ms) and amplitude (in μV) of the N1, P2 and P2b component for the different conditions

|     |     | Happy | | Fear | |
| --- | --- | --- | --- | --- | --- |
|     |     | Congruent | Incongruent | Congruent | Incongruent |
| **N1** | FC3 | 112/−1.75 | 110/−2.33 | 122/−2.33 | 118/−1.75 |
|     | FCz | 110/−2.34 | 112/−3.72 | 114/−2.46 | 116/−2.32 |
|     | FC4 | 108/−1.73 | 112/−2.12 | 114/−1.47 | 124/−1.57 |
| **P2** | C3 | 208/2.31 | 204/2.07 | 186/1.25 | 218/1.66 |
|     | Cz | 206/4.25 | 204/3.34 | 210/2.36 | 216/2.81 |
|     | C4 | 198/1.54 | 204/1.08 | 212/1.3 | 194/0/97 |
| **P2b** | CP3 | 252/2.4 | 266/2.69 | 218/1.46 | 248/2.64 |
|     | CPz | 230/3.05 | 250/3.84 | 218/2.73 | 246/3.76 |
|     | CP4 | 226/1.14 | 244/1.56 | 218/1.91 | 244/2.2 |

ponent in the sense that incongruent trials trigger a delayed P2b component compared to congruent trials, suggesting a slowing of the auditory processing around 240 ms when incongruent audio-visual pairs are presented. This is indicated by an extra shift in time for the positive deflection starting around 180 ms and peaking around 250 ms with incongruent pairs (Fig. 2). These observations were confirmed by a repeated measures ANOVA on the
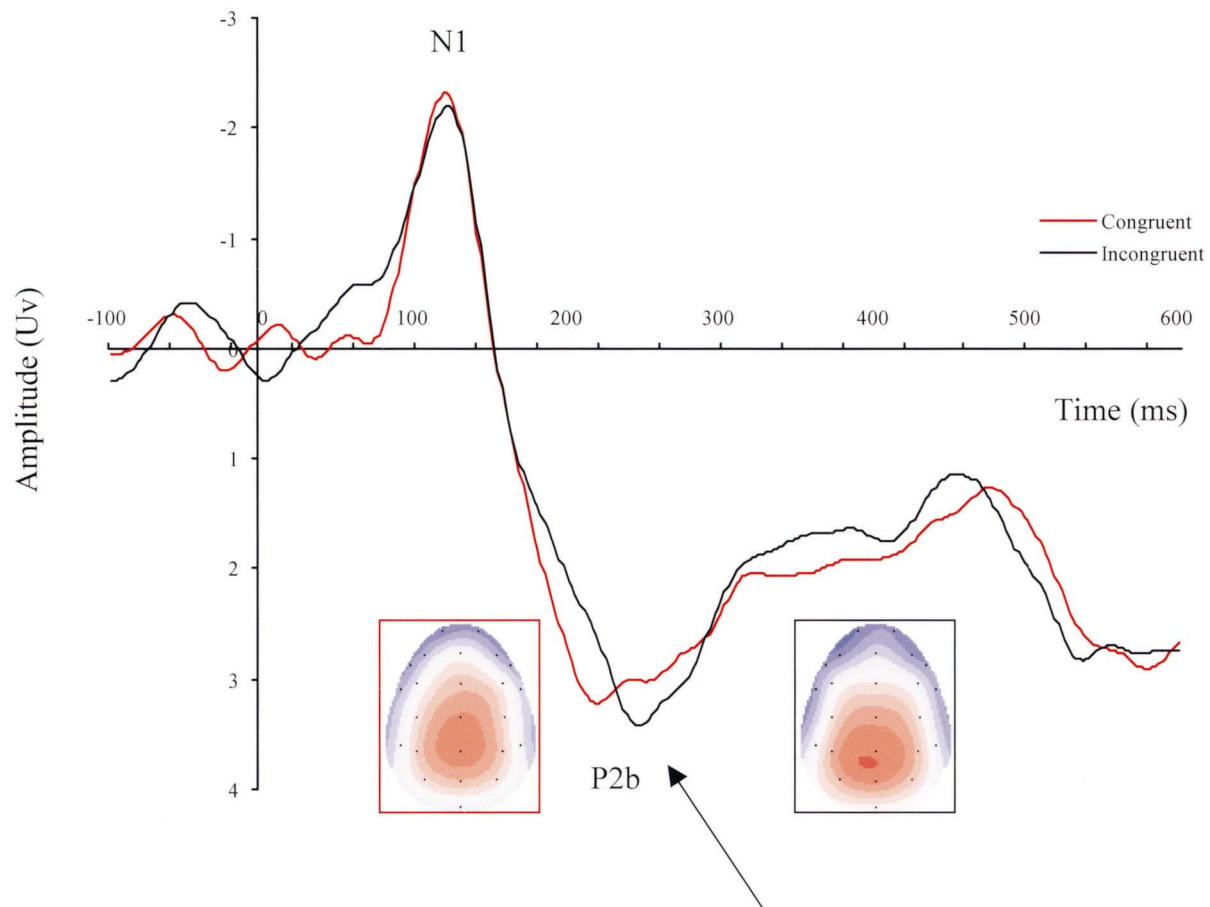


Fig. 2. Grand averaged auditory waveforms at CPz electrode measured during the presentation of congruent and incongruent audio-visual stimulus pairs (and corresponding topographies at 224 and 242 ms in the congruent and incongruent condition, respectively).

latencies of the P2b at three posterior leads with the factors *Emotion* (happy vs. fearful), *Congruency* (congruent vs. incongruent pairs) and *Electrode Position* (left, mid-line or right): the analysis only revealed a significant effect of *Congruency* [$F(1,7) = 7.68$, $P = 0.015$] indicating that incongruent pairs (mean latency at CPz = 248 ms) delay the P2b component compared to congruent pairs (mean latency at CPz = 220 ms). This congruency effect is not modulated by emotion. The same statistical analysis carried out on the maximum amplitudes of the P2b component revealed a significant effect of *Electrode Position* [$F(2,14) = 5.16$, $P = 0.02$] indicating that the amplitude of the P2b component is higher at the central electrode than at the left [$t(7) = 2.74$, $P = 0.012$] or right [$t(7) = 2.82$, $P = 0.014$] electrode [the difference between left and right electrode being non-significant, $t(7) < 1$].

The unconstrained source localization disclosed an origin in the cingulum for the auditory P2b component (Fig. 3) which is very stable for the early time window (210–230 ms) of this component. At 216 ms, the mean coordinate ($x$ is antero-posterior, $y$ is left–right and $z$ is vertical) of the source in the realistic head shape model is $8.9x$ $0.5y$ $47.2z$ (3% residual variance). At 226 ms, it is $8.8x$ $3.9y$ $46.1z$ (3.9% residual variance) and $6.1x$ $6.8y$ $44.1z$ (4.7% residual variance) at 234 ms suggesting that the dipole solution in the anterior cingulate cortex is quite steady within the time window corresponding to the P2b component at the level of the scalp. During the 200–250 ms time window, the overall correlation is 0.96 and the overall goodness of fit is 0.92 for the single dipole solution in the anterior cingulate cortex. A second source analysis performed during the 200–250 ms time-window with the symmetry constraint and with two dipoles localized in the middle temporal gyrus ($-30x$ $-50y$ $-10z$ in the left hemisphere and $-30x$ $50y$ $-10z$ in the right hemisphere) reveals a weaker solution than the single dipole solution in the anterior cingulate cortex with an overall correlation of 0.86 and an overall goodness of fit of 0.73. Moreover, this second solution was not stable within the 200–250 ms time-window. Finally, a third source analysis was performed during the 200–250 ms time-window (to be fit to the P2b distribution) with the symmetry constraint only which allows for some movement of one pair of dipoles. This revealed again a weaker solution (overall correlation of 0.87 and an overall goodness of fit of 0.74) than the single dipole solution in the anterior cingulate cortex. In this latter case, the two dipoles were localized in the inferior frontal gyrus (Brodmann area 45, $25x$ $-50y$ $7z$ in the left hemisphere and $25x$ $50y$ $7z$ in the right hemisphere).

## 4. Discussion

A detailed analysis of the auditory waveforms at posterior leads suggest the existence of a positive deflection following the N1–P2 complex and labeled here the P2b component (the present P2b is not equivalent to the brainstem auditory response known in the older literature as the P2b [11]. Interestingly, we demonstrated that congruent face-voice trials elicit an earlier P2b component than incongruent trials suggesting that the processing of auditory words is delayed when performed in the presence of an incongruent facial context. At the electrophysiological level, the P2b component follows the auditory P2 component with a more posterior topography and is thought to represent the perceptual integration of the auditory fragment with the concurrent facial context. The latency and topography properties of the P2b component suggest that this component does not overlap with later EEG components (e.g. the N2–P3 complex) known to be involved in cognitive processes at later decisional stages. Three arguments allow us to conclude that the P2b component does not overlap with the P3 component and that the P2b component is not a P3 component with an earlier time-course. A detailed analysis of the waveforms and topographies suggests that the P2b component clearly precedes the putative N2–P3 complex (see Fig. 1). A passive task was used and this manipulation is less appropriate to yield a N2–P3 complex in the EEG than overt responses. Finally, we did not use the classical oddball paradigm used in neurophysiology to elicit in the EEG a clearcut P3 component. The P2b component and its generator in the anterior cingulate cortex are partly similar to another late positive potential occuring in the auditory modality (as well as in the visual modality) and known in the literature as the error positivity (see Ref. [15] for a recent review) but the underlying cognitive process at stake in the present study is quite different from the error detection mechanisms indexed by the error positivity. The error positivity is an ERP component on reaction errors which is subsequent upon the error negativity. Our results are also consistent with recent studies based either on functional magnetic resonance imaging [5] and on magnetoencephalography [24] reporting that the phonemic (auditory) and graphemic (visual) counterparts of alphabetical letters converge around 200 ms within the Superior Temporal Sulcus.

Source localisation carried out on the time-window of the P2b component disclosed a single dipole solution in the anterior cingulate cortex, an area known to be selectively implicated in processing congruency or conflict between stimuli [18]. The contribution of anterior cingulate cortex in dealing with congruency has been shown in many previous brain imaging studies [2]. The present results are compatible with involvement of the cingulate cortex in face-voice pairing around 220 ms. The specific anatomical situation of the cingulate cortex which is close and parallel to the median line may yield a single dipole solution that better explains the corresponding scalp EEG than multiple dipoles solutions when putative generators are localized in this region.
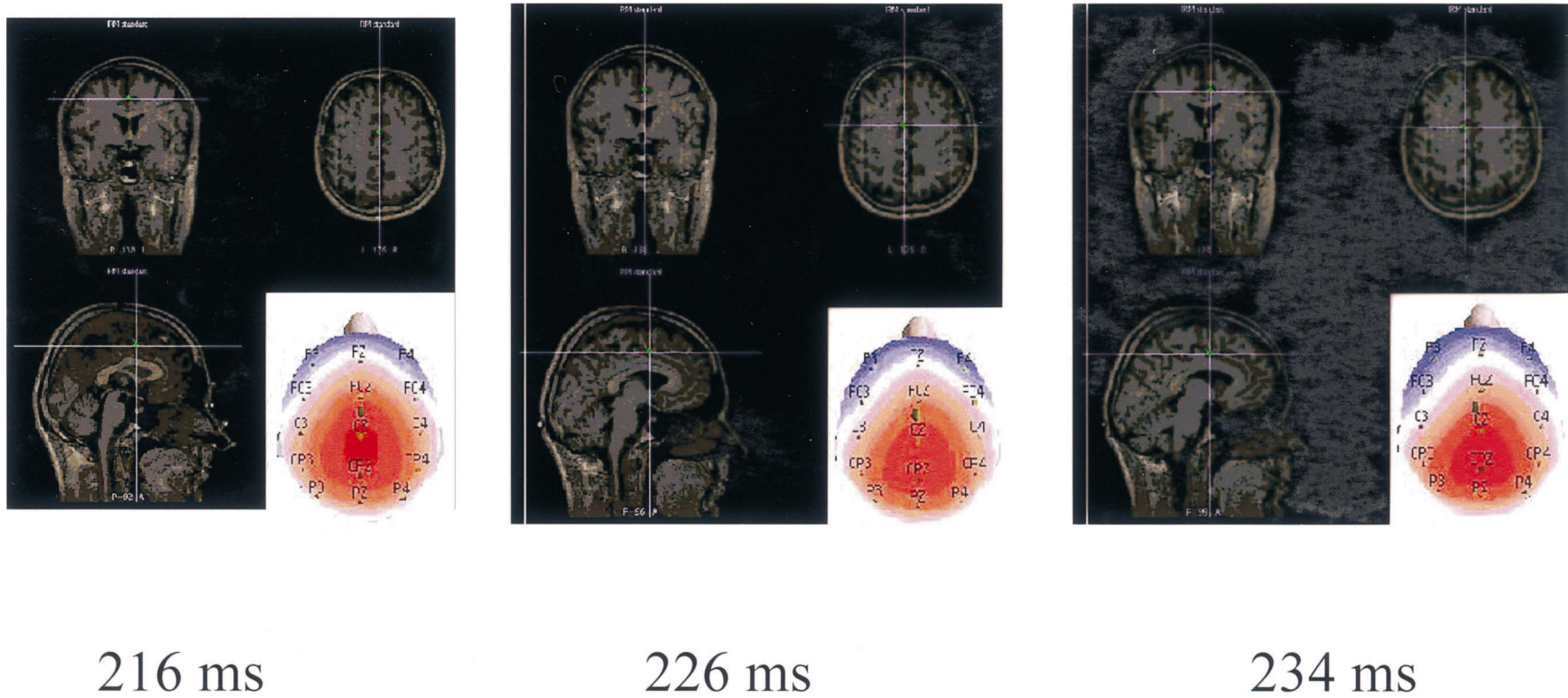
216 ms          226 ms          234 ms

Fig. 3. Coronal, axial and sagittal sections (and corresponding topographies on the horizontal axis) of a normalized brain showing the single dipole solution in the anterior cingulate cortex for 216, 226 and 234 ms time-window post stimulus.

Here we showed that audio-visual integration of emotion is taking place at perceptual stages during the auditory processing (i.e. around 220 ms post-stimulus at posterior leads) and not at later decisional stages when modality-specific processing is presumably completed as indexed by the N2–P3 complex. Previous studies suggested that audio-visual pairing was first manifested in the scalp EEG at early stages (100 ms) by an amplitude increase of modality-specific components [16,23]. In the present study, we have not shown any modulation in amplitude of the auditory N1 component as previously found [9,23] since we used a different methodology which is inappropriate to investigate early amplitude effects [22] and we specifically assessed whether a facial expression could also influence the latency of the concurrent auditory process at a later stage. Our previous EEG studies did not show changes in the latency parameter of the auditory related components but it used a different design (an oddball design), different auditory materials and only negative emotions. Early amplitude effects in the auditory modality are probably followed within a later time-window (240 ms) by the appearance of components (like the P2b component) characterized as sensitive to the content of the audio-visual pairs. In this respect, the perception of multimodal events would be manifested in the EEG by an amplification process going on in modality-specific cortices (like the secondary auditory cortex or the extra-striate visual cortex) as well as by the activation of a cortical network (including cortical areas such as the posterior parietal cortex and the middle temporal gyrus) characterized as multimodal [28,1,21].

## Acknowledgements

## References

[1] R.A. Andersen, L.H. Snyder, D.C. Bradley, J. Xing, Multimodal representation of space in the posterior parietal cortex and its use in planning movements, Annu. Rev. Neurosci. 20 (1997) 303–330.

[2] R. Cabeza, L. Nyberg, Imaging cognition II: An empirical review of 275 PET and fMRI studies, J. Cogn. Neurosci. 12 (2000) 1–47.

[3] G.A. Calvert, M.J. Brammer, S.D. Iversen, Crossmodal identification, Trends Cognit. Sci. 2 (1998) 247–260.

[4] G.A. Calvert, E.T. Bullmore, M.J. Brammer, R. Campbell, S.C. Williams, P.K. McGuire, P.W. Woodruff, S.D. Iversen, A.S. David, Activation of auditory cortex during silent lipreading, Science 276 (1997) 593–596.

[5] G.A. Calvert, R. Campbell, M.J. Brammer, Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex, Curr. Biol. 10 (2000) 649–657.

[6] B. de Gelder, J. Vroomen, Hearing smiles and seeing cries: The

[7] B. de Gelder, Crossmodal perception: there is more to seeing than meets the eye, Science 289 (2000) 1148–1149.

[8] B. de Gelder, J. Vroomen, Perceiving emotions by ear and by eye, Cognit. Emot. 14 (2000) 289–311.

[9] B. de Gelder, K.B.E. Böcker, J. Tuomainen, M. Hensen, J. Vroomen, The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses, Neurosci. Lett. 260 (1999) 133–136.

[10] R.J. Dolan, J.S. Morris, B. de Gelder, Crossmodal binding of fear in voice and face, Proc. Natl. Acad. Sci. USA 98 (2001) 10006–10010.

[11] E. Donchin, Event-related brain potentials: A tool in the study of human information processing, in: H. Begleiter (Ed.), Evoked Potentials and Behavior, Plenum, New York, 1979, pp. 13–75.

[12] E. Donchin, M.G.H. Coles, Is the P300 component a manifestation of context updating?, Behav. Brain Sci. 11 (1988) 357–374.

[13] J. Driver, C. Spence, Multisensory perception: beyond modularity and convergence, Curr. Biol. 10 (2000) 731–735.

[14] P. Ekman, W.V. Friesen, Pictures of Facial Affect, Consulting Psychologists Press, Palo-Alto, 1976.

[15] M. Falkenstein, J. Hoormannm, S. Christm, J. Hohnsbein, ERP components on reaction errors and their functional significance: a tutorial, Biol. Psychol. 51 (2000) 87–107.

[16] M.H. Giard, F. Peronnet, Auditory-visual integration during multimodal object recognition in humans: a behavioral and electro-physiological study, J. Cogn. Neurosci. 11 (1999) 473–490.

[17] S.A. Hillyard, G.R. Mangun, M.G. Woldorff, S.J. Luck, Neural systems mediating selective attention, in: M.S. Gazzaniga (Ed.), The Cognitive Neurosciences, The MIT Press, Cambridge, MA, 1995, pp. 665–681.

[18] C.M. MacLeod, P.A. MacDonald, Interdimensional interference in the Stroop effect: uncovering the cognitive and neural anatomy of attention, Trends Cogn. Sci. 4 (2000) 383–391.

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

[20] D.W. Massaro, P.B. Egan, Perceiving affect from the voice and the face, Psychonom. Bull. Rev. 3 (1996) 215–221.

[21] M.M. Mesulam, From sensation to cognition, Brain 121 (1998) 1013–1052.

[22] R. Näätänen (Ed.), Attention and Brain Function, Lawrence Erlbaum Associates, Hillsdale, NJ, 1992, p. 494.

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

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

[25] M. Sams, R. Aulanko, M. Hamalainen, R. Hari, O.V. Lounasmaa, S.T. Lu, J. Simola, Seeing speech: visual information from lip movements modifies activity in the human auditory cortex, Neurosci. Lett. 127 (1991) 141–145.

[26] M. Scherg, Fundamentals of dipole source potential analysis, in: F. Grandori et al. (Eds), Auditory Evoked Magnetic Fields and Potentials, Vol. 6 (Advances in Audiology), Karger, 1990, pp. 40–69.

[27] K.C. Squires, N.K. Squires, S.A. Hillyard, Decision-related cortical potentials during an auditory signal detection task with cued intervals, J. Exp. Psychol.: Human Perception and Performance 1 (1975) 268–279.

[28] B.E. Stein, M.A. Meredith (Eds.), The Merging of the Senses, The MIT Press, Cambridge, MA, 1993, p. 211.

[29] V. Surakka, M. Tenhunen-Eskelinen, J.K. Hietanen, M. Sams, Modulation of human auditory information processing by emotional visual stimuli, Cogn. Brain Res. 7 (1998) 159–163.

[30] R. Verleger, Event-related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of P3, Behav. Brain Sci. 11 (1988) 343–427.

bimodal perception of emotions, Bulletin of the Psychonomic Society, 1995.