Emotional Cues during Simultaneous Face and Voice Processing: Electrophysiological Insights

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

PLoS one

Emotional Cues during Simultaneous Face and Voice Processing: Electrophysiological Insights

Taosheng Liu[1,2], Ana Pinheiro[2,3], Zhongxin Zhao[4]*, Paul G. Nestor[2,5], Robert W. McCarley[2], Margaret A. Niznikiewicz[2]*

1 Department of Psychology, Second Military Medical University, Shanghai, China, 2 Clinical Neuroscience Division, Laboratory of Neuroscience, Department of Psychiatry, Boston VA Healthcare System, Brockton Division and Harvard Medical School, Brockton, Massachusetts, United States of America, 3 Neuropsychophysiology Laboratory, CiPsi, School of Psychology, University of Minho, Braga, Portugal, 4 Department of Neurology, Neuroscience Research Center of Changzheng Hospital, Second Military Medical University, Shanghai, China, 5 University of Massachusetts, Boston, Massachusetts, United States of America

Abstract

Both facial expression and tone of voice represent key signals of emotional communication but their brain processing correlates remain unclear. Accordingly, we constructed a novel implicit emotion recognition task consisting of simultaneously presented human faces and voices with neutral, happy, and angry valence, within the context of recognizing monkey faces and voices task. To investigate the temporal unfolding of the processing of affective information from human face-voice pairings, we recorded event-related potentials (ERPs) to these audiovisual test stimuli in 18 normal healthy subjects; N100, P200, N250, P300 components were observed at electrodes in the frontal-central region, while P100, N170, P270 were observed at electrodes in the parietal-occipital region. Results indicated a significant audiovisual stimulus effect on the amplitudes and latencies of components in frontal-central (P200, P300, and N250) but not the parietal occipital region (P100, N170 and P270). Specifically, P200 and P300 amplitudes were more positive for emotional relative to neutral audiovisual stimuli, irrespective of valence, whereas N250 amplitude was more negative for neutral relative to emotional stimuli. No differentiation was observed between angry and happy conditions. The results suggest that the *general* effect of emotion on audiovisual processing can emerge as early as 200 msec (P200 peak latency) post stimulus onset, in spite of implicit affective processing task demands, and that such effect is mainly distributed in the frontal-central region.

Introduction

The ability to extract emotional salience from visual and/or auditory signals has important implications for effective functioning in social environment. Due to their emotional significance, emotional stimuli are thought to capture attention automatically when compared with neutral stimuli, both in visual and auditory modalities [1–5].

Studies in the last decades shed light on the neural processes underpinning emotional processing. Most of the studies examined emotion processing from unisensory stimuli, such as voice (prosody processing) [3,5,6] and faces [1,4,7]. There were also some studies that examined the simultaneous processing of emotional visual and auditory signals [8,9]. The use of event related potential (ERP) methodology for the study of emotional processing during face and/or voice is particularly advantageous since it affords tracking neurocognitive processes as they happen in real time from the millisecond the stimulus is presented and thus provides a window of inquiry into these processes before a response is made.

Studies on emotional face processing suggest that emotion modulates early processing stages and that the processing of emotional facial expressions is automatic, i.e. not dependent on directed attention [10–12]. The early effects of emotion on face processing point to a first differentiation between neutral and emotional signals but not between different types of emotion (from 110 to 250 ms, including P200 and N250 ERP components) [13], observed mainly at frontocentral electrodes [1]. The differentiation between specific emotions seems to occur at later stages, from 400 to 750 ms (including the Late Positive Potential-LPP component) [1]. Of note, the N250 has been proposed to index processes related to decoding of emotional content from faces [14,15], in contrast to the N170 that has been related more consistently to the structural encoding of faces [11,16,17]. Some of these studies indicate that the analysis of stimulus emotional salience and the structural encoding of faces (as indexed by the N170 ERP component) are partially independent processes [1,12,13]. Finally, emotion effects were also reported for P300 amplitude in studies using face stimuli [18–20].

ERP studies on auditory emotional processing (including stimuli such as speech prosody [3,21], music [22], or environmental sounds [23] have reported early modulatory effects of emotion on auditory sensory and perceptual processing. These effects were observed on N100 and P200 amplitude, including more negative N100 amplitude for neutral relative to emotional vocalizations [2],

more positive P200 for happy than neutral or sad voice stimuli [22], more positive P200 for neutral relative to emotional prosody [3], or more positive P200 for emotional relative to neutral vocalizations [2,24]. These effects were more pronounced at frontocentral electrodes [2]. Together, these findings suggest that N100 may index the earliest point at which emotional and non-emotional vocalizations can be distinguished, likely along sensory, acoustic dimentions, and that P200 is an index of the extraction of emotional salience from acoustic cues whether or not they contain linguistic information. In addition, effects of emotional valence were also consistently reported for P300 amplitude [1,25,26].

In spite of advances observed in recent years [8,9,22,27–33], there is still a dearth of data on how emotional information from one modality can influence processing in the other modality and how these two sources of information interact. Studies using audiovisual stimuli suggest that during affective processing inputs from one modality influence perception in another modality. Behavioral studies of audiovisual integration indicate that two congruent stimulus modalities may improve processing, as indexed by shorter reaction times and higher accuracy, when compared to unimodal stimulus [34–36]. The few ERP studies on multisensory integration of emotional information argue for an early (in the first 225 msec) integration of non-verbal emotional information from visual (face or picture) and auditory (voice or music) modalities. Reductions in the amplitude and latency of components of interest have been taken as evidence of the influence of one modality on the other, although enhanced amplitude and longer latencies have been also reported. Audiovisual integration was found to operate at the level of pre-attentive processes as indicated by mismatch negativity (MMN) study results [31], suggesting that the integration of emotional information from face and voice occurs automatically and before both have been fully processed independently. In addition, the evidence of early integration of emotional information was found at the level of N100, as reflected by its increased amplitude when the emotion in face and voice was congruent relative to when the emotion was incongruent [32]; this congruence effect did not occur when faces were presented upside-down. Giard and Peronnet [37], investigating multisensory integration of auditory and visual information from simple objects, found that audiovisual processing was associated with reduced early sensory components relative to the visual modality. Similar effects of increased processing speed and reduced effort in processing multisensory relative to unisensory non-emotional face and voice stimuli were reported by Brefczynski-Lewis and colleagues [38]. Of particular note, in the study of Pourtois et al. [33] that investigated the integration of congruent (happy and fearful) facial expressions and prosody, an earlier posterior P2b component was observed for congruent face-voice trials relative to incongruent trials, indicating delayed auditory processing for incongruent facial contexts. These findings suggested that the cross-modal integration of emotion occurs at early (perceptual) stages during auditory processing. Finally, Spreckelmeyer et al. [22] testing the combined perception of emotional cues (happy and sad vs. neutral) via the simultaneous presentation of matched or mismatched affective pictures (from the International Affective Pictures System - IAPS database - [39]) and affective sung notes, found a modulation of P200 and late positive potential (LPP) components. In particular, congruent happy pictures and voices led to more positive P200 amplitude in an attended relative to unattended condition, and congruent sad pictures and voices led to more positive-going late-positive potential (LPP).

While the Spreckelmeyer et al. [22] and Pourtois et al. [33] studies clearly suggested interactions between auditory and visual channels in terms of modulating emotional and perceptual processes, their interpretation is difficult given the small sample size in these studies and a lack of a neutral condition in Pourtois et al. [33] that would allow a comparison between audiovisual effects in the presence and absence of emotional cues. Together, findings from unisensory and multimodal studies indicate early emotion effects (indexed by N100, P200, N250, and P300), suggesting that the detection of emotional signals is relatively automatic. However, it is still unclear how emotional cues presented in different sensory modalities (e.g., face and voice) are processed when presented simultaneously and what neural processes underpin these cognitive and perceptual events. To the best of knowledge, there are no published ERP studies that examine simultaneous emotion processing from (static) face and voice. Puce et al. [40] examined species specific congruency (using dynamic face and voice stimuli) effects, but the study focused on the flexible use of emotionally neutral information rather than on the processing of emotional information from multiple channels.

The current study aimed at investigating electrophysiological correlates of processing neutral vs. emotional (angry and happy) cues in simultaneously presented static faces and non-semantic sounds. Non-semantic sounds were chosen to avoid the confounding effects of linguistic (e.g., semantic) information on emotion processing. An additional motivation to use non-semantic emotional sounds was the fact that they constitute an important vehicle to convey emotional attitudes and mood states.

In addition, we were interested in examining the effects of emotional processing of audiovisual stimuli in the absence of task requirements that would focus on these emotional cues. Given the significance of emotion for survival, and the fact that in daily life emotion is conveyed from multisensory modalities in an automatic way and often without full conscious awareness [41], we expected that this design would provide us with a more ecologic account of multimodal emotional processing from auditory and visual modalities. In order to maintain attention to the task, participants were instructed to detect a monkey face from a series of faces that contained either neutral or emotional expressions.

We hypothesized that, consistent with previous studies on affective multimodal integration, we would observe a differentiation between neutral and emotional conditions in early components, particularly before 300 msec. Given the sensitivity of N100 and P200 amplitude to attention, and the relevance of emotional information to survival, we expected that the emotional conditions would capture more attentional resources than the neutral condition, which would be indexed by enhanced N100 and P200 amplitude. In addition, we hypothesized that later components would be also sensitive to emotional manipulations, in particular the N250 and P300 [1,14,15,25,26].

Based on previous accounts of increased positivity for emotional relative to neutral stimuli in the 300 msec latency window that spans both the N250 and P300 components [42], we expected reduced N250 and increased P300 for emotional relative to neutral stimuli. Given inconsistent findings related to specific emotion effects on the P300 component [10,42,43], we aimed at investigating whether P300 recorded to complex emotional audiovisual input distinguishes between emotional and neutral information, similarly to earlier components, or whether it distinguishes between specific emotional categories, here happy and angry faces. We reasoned that audiovisual emotion processing effects will be observed at fronto-central location and that parieto-occipital components associated with structural aspects of face processing (including N170) will not be sensitive to emotion processing effects, as suggested by previous studies [1,12,13].

Results

1. Behavioral data

The mean response time to targets was 702.98±192.05 ms with 97.94% response accuracy. The high response accuracy suggested that subjects stayed engaged by the task and that their attention was appropriately maintained.

2. ERP data

Figure 1 shows grand average waveforms for neutral and emotional conditions, at parieto-occipital electrodes. Figure 2 illustrates the topographic distribution of parieto-occipital components for each condition.

2.1. Parieto-occipital components

P100. No significant main effect or interaction involving the factor *condition* was found for P100 amplitude or latency ($p>0.05$).

N170. As for P100, there was no significant main effect or interaction involving the factor *condition* ($p>0.05$).

P270. No significant main effect or interaction involving *condition* factor was found for P270 ($p>0.05$). The lack of condition differences for all parietal-occipital components was retained when the data were re-analyzed using the average reference.

2.2. Frontal-central components

Figure 3 shows grand average waveforms for neutral and emotional conditions, at fronto-central electrodes. Figure 4 illustrates the topographic distribution of fronto-central components for each condition.

N100. N100 amplitude to Happy and Angry condition seemed less negative relative to neutral condition. However, the statistical results showed no significant main effect or interaction involving condition factor ($p>0.05$).

P200. The MANOVA on P200 amplitude yielded a significant main effect of region ($F_{(1,17)} = 11.532$, $p = 0.003$): P200 amplitude was more positive at frontal relative to central region. A significant main effect of condition ($F_{(2,16)} = 3.919$, $p = 0.041$) was also observed: P200 was less positive for neutral (mean = 2.58 μV) relative to happy ($p = 0.044$; mean = 4.013 μV) condition, while neither neutral or happy condition was significantly different from Angry (mean = 3.57 μV) condition.

No significant main effects or interactions were observed for P200 latency.

N250. The MANOVA on N250 amplitude yielded a significant main effect of condition ($F_{(2,16)} = 4.934$, $p = 0.021$): N250 was more negative for neutral (mean = −6.59 μV) relative to both Angry ($p = 0.048$; mean = −5.08 μV) and Happy ($p = 0.033$; mean = −5.15 μV) conditions, and no difference was found between Angry and Happy condition ($p = 1.000$).

The MANOVA on N250 latency yielded a significant main effect of condition ($F_{(2,16)} = 8.664$, $p = .003$): N250 peaked earlier for Happy (mean = 257.50 msec) relative to Angry ($p = 0.038$; mean = 261.81 msec) and Neutral ($p = 0.003$; mean = 264.29 - msec) conditions, with no difference between the two latter conditions.

P300. The MANOVA on P300 amplitude yielded a significant main effect of condition ($F_{(2,16)} = 7.697$, $p = 0.005$): P300 was less positive for neutral (mean = 2.21 μV) than for both angry ($p = 0.023$; mean = 4.26 μV) and happy ($p = 0.009$; mean = 4.47 μV) conditions, with no difference between Angry and Happy condition ($p = 1.000$). No significant main effects or interactions were observed for P300 latency.

2.3. Correlation between P200 and N250

The relationship of P200 and N250 amplitude to each condition (Angry, Happy and Neutral) was examined using Pearson correlations at Cz. For Angry condition, there was a significant positive correlation between the P200 and N250 amplitudes ($r = 0.557$, $p = 0.013$); For Happy condition, this correlation was $r = 0.520$, $p = 0.022$; For Neutral condition, it was $r = 0.762$, $p = 0.000$. The results indicated that more positive P200 was associated with less negative N250.
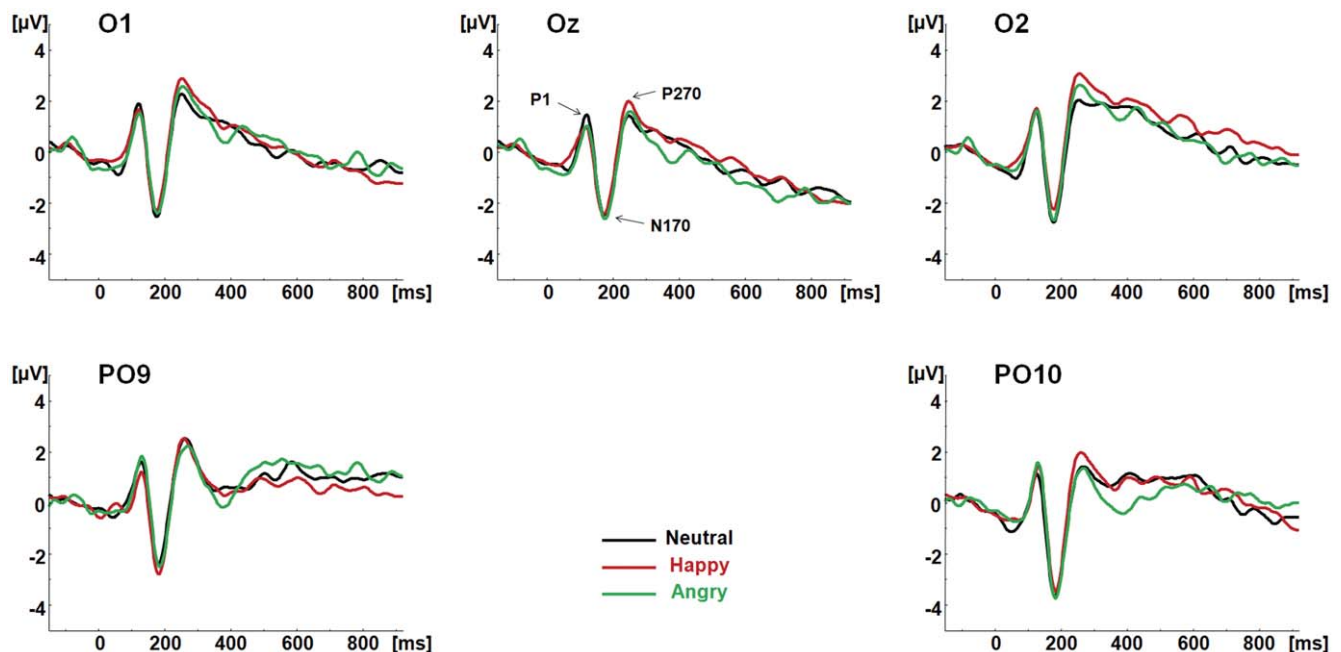
Figure 1. ERP waveforms to each condition at parieto-occipital electrodes. The parieto-occipital components P1, N170 and P270 are shown at O1, Oz, O2, PO9 and PO10 for the neutral (Black), happy (Red) and angry (Green) congruent audiovisual conditions (face and voice).
doi:10.1371/journal.pone.0031001.g001
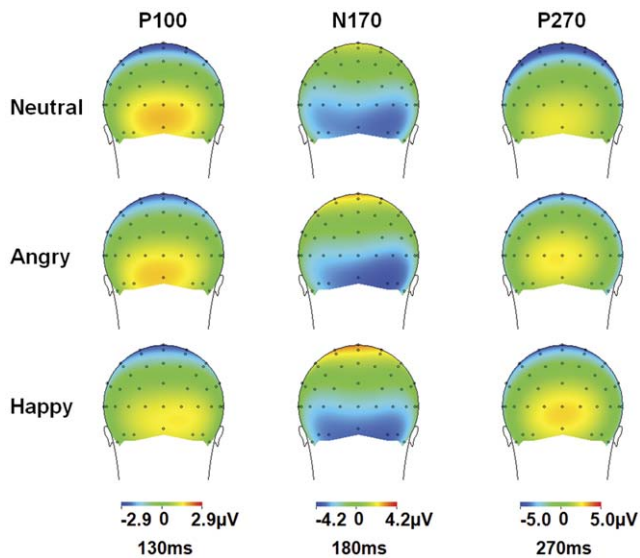
Figure 2. Scalp topographies for the parieto-occipital ERPs components. The voltage topographies of the parieto-occipital components on the scalp are shown for Neutral (top row), Angry (middle row) and Happy (bottom row) congruent audiovisual conditions (face and voice) at 130, 180, and 270 ms.
doi:10.1371/journal.pone.0031001.g002

Discussion

This study aimed at investigating the processing of emotional cues during simultaneous presentation of face and voice. To this end, congruent faces and non-semantic sounds were presented, while subjects were asked to perform a visual categorization task, i.e., deciding if they saw a monkey face. Results showed that the extraction of emotional cues from multimodal input was associated with differential ERP activity at frontal-central sites, and was indexed by P200, N250, and P300. P100, N170 and P270 observed at parieto-occipital sites were not modulated by emotional content.

Fronto-central negativities showed a differential amplitude pattern: while N100 did not distinguish between neutral and emotional conditions, N250 was less negative to emotional relative to neutral face-voice pairings. Fronto-central positive components, P200 and P300 were more positive for emotional relative to neutral face-voice pairings. No component distinguished between the type of emotion presented (happy and angry). The emotion effects observed for P200, N250 and P300 suggest that emotional information cannot be ignored even when it is not task-relevant [48]. These results also suggest that emotion modulates both perceptual (P200, N250), and attentional (N250, P300) processes in the two sensory modalities.

In the following sections, we will separately discuss main findings for each component, integrating our findings with recent evidence on affective and multimodal processing.

N100: the similar sensory processing of neutral and emotional audiovisual information

The audiovisual (face-voice) stimuli generated a robust N100 response; however, the N100 did not distinguish between the emotional types of stimuli.

N100 has been traditionally associated with sensory processes; it was found sensitive to physical characteristics of the stimuli and modulated by attention [49,50]. In previous studies investigating multimodal integration of affective information, N100 was found sensitive to congruency effect: it was larger for congruent relative to incongruent angry face-voice pairs [32], which suggested an early integration of information coming from different inputs. However, that study did not use neutral condition.

In the current study congruency of inputs was not investigated but rather, we focused on implicit processing of emotion conveyed
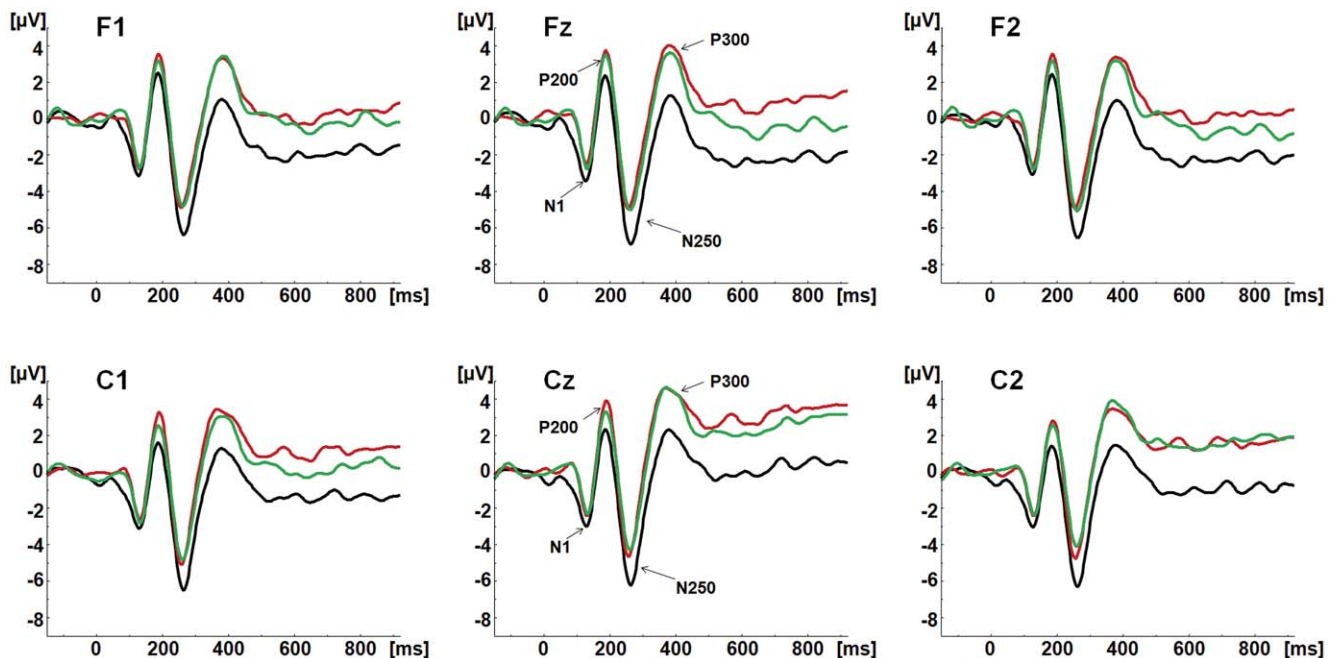
Figure 3. ERP waveforms to each condition at fronto-central electrodes. The fronto-central components N1, P200, N250 and P300 are shown at F1, Fz, F2, C1, Cz and C2 for the neutral (Black), happy (Red) and angry (Green) congruent audiovisual conditions (face and voice).
doi:10.1371/journal.pone.0031001.g003
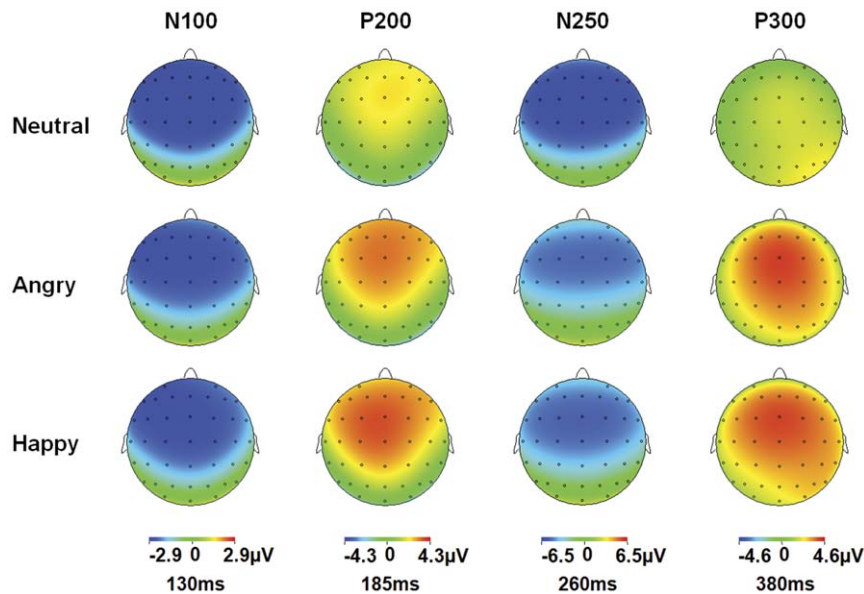
Figure 4. Scalp topographies for the fronto-central ERPs components. The voltage topographies of the fronto-central components on the scalp are shown for Neutral (top row), Angry (middle row) and Happy (bottom row) congruent audiovisual conditions (face and voice) at 130, 185, 260 and 380 ms.
doi:10.1371/journal.pone.0031001.g004

by auditory and visual modalities. The results of our study suggest that N100 does not provide a robust distinction between neutral and emotional audiovisual stimuli. Even though a distinction between neutral and emotional stimuli has been observed in previous studies with unimodal stimuli [51], this was not observed in our study with complex audiovisual input. Because of the sensitivity of N100 amplitude to attention, these results may suggest that both neutral and emotional audiovisual stimuli captured similar amount of attention at this early processing stage. Alternatively, given the implicit nature of emotion perception in this task, attention may not have played a major role in the early distinction between neutral and emotional information.

P200: extraction of emotional salience and increased attention to relevant (emotional) audiovisual stimuli

The P200 was obtained to all stimuli and was the first component that indicated processing differences between emotional and non-emotional face-voice stimuli.

Traditionally, P200 has been associated with early categorization processes, is sensitive to physical features of stimuli and modulated by attention [52–55]. Studies on affective prosody processing have pointed to P200 sensitivity to extracting emotional salience from the auditory signal [3] and valence-specific orienting to relevant stimuli [56]. Studies on visual processing have also reported the relevance of P200 as an index of emotional salience detection [12,13,57]. Unimodal studies (both *visual* and *auditory*) suggest that the emotional differentiation irrespective of valence occurs around 200 msec after stimulus onset (e.g., *auditory* – [3]; *visual* – [12,13]). In addition, this early differentiation may be independent of attention, occurring under implicit processing demands, as suggested by previous studies [32].

In the current study, increased P200 was found to emotional (both angry and happy) relative to neutral condition, even though the difference between angry and neutral condition did reach statistical significance. These results have been obtained independent of attentional focus. They suggest that the differentiation

between emotional and non-emotional cues may happen outside of the focus of attention.

It is noteworthy that the differentiation between neutral and emotional stimuli but not between specific types of emotions at the level of P200 has been previously reported in the studies of prosody processing where electrophysiological correlates of emotional and neutral prosody were investigated [58]. In that study, P200 was proposed to index the extraction of emotional salience from the auditory signal. Thus, the results of the current study, in conjunction with the previous studies' results, suggest that the processes indexed by P200 and related to extracting emotional salience are not modality specific but rather operate on both auditory and audiovisual material [22]. This finding is consistent with the role of P200 in categorization processes that are likely not modality specific (differentiating if a stimulus is emotionally salient or not is a type of categorization).

N250: decoding emotional cues for discrimination of audiovisual stimuli

While early components index more sensory aspects related to early stimulus categorization, later components are argued to index more complex cognitive and affect-related processes [56]. The fronto-central distributed N250 has been proposed to index early perceptual recognition processes and decoding of emotional cues (such as the extraction of emotional cues from faces [59]. Using a task of facial emotional recognition and perception of blurred faces and objects, Streit et al. [59] found a more pronounced N250 component elicited by the faces in the emotion recognition task, whereas this component was virtually absent to the faces in the blurred object condition. Similar results were also reported by Wynn et al. [60]: these authors used three tasks – a gender identification task, an emotion identification task, and a building identification task. The results indicated the largest N250 amplitude for emotion, followed by gender, and then by buildings. Unfortunately, both studies did not include a contrast between neutral and emotional faces.

In our study, less negative N250 amplitude was observed for emotional (both happy and angry) relative to neutral condition. These results are consistent with previous studies of emotional face processing reporting reduced amplitude of N250, VPP (P200) and P300 to emotional relative to neutral faces, irrespective of valence [12,42].

Of note, in a study using dynamic audiovisual speech stimuli [61], a negative deflection peaking around 300 msec with a similar scalp distribution as the N250 in our study, was found to be more negative to incongruent than to congruent audiovisual speech stimuli. Also, another study showed that incongruous relative to congruous faces in a face identity-matching task elicited an enhanced negativity peaking around 350 msec with a frontocentral maximum [62]. This seems to indicate that difficulties in integration may be related to a more negative N250 amplitude. Thus, we may argue that our findings suggest that greater emotional salience of emotional audiovisual information leads to less difficulty in its processing. Consistent with this hypothesis, the N250 latency was earlier in the happy relative to neutral condition. Thus, while emotional audiovisual stimuli are related to enhanced stimulus elaboration (more positive P200), they also seem to be related to easier integration processes (less negative N250).

It is also possible that the two processes are closely related and influence each other such that the reduced N250 for emotional stimuli is due to the preceding enhancement of the P200 [63]. Of note here is the observation of a strong positive correlation between the anterior P200 and the N300 in an investigation of the differential processing of angry, happy, and neutral faces [56]. More positive P200 for emotional faces was associated with less negative N300 for the same stimuli. In our study, we observed a similar relationship between P200 and N250.

P300: discrimination of emotional audiovisual information for enhanced elaboration and evaluation

Processes indexed by P300 are associated with processes such as stimulus recognition and discrimination [64]. In addition, the P300 has been shown to index the amount of attentional resources allocated to a stimulus [65]. Also, P300 amplitude can be modulated by valence: the reported *emotion effect* (i.e., more positive-going ERPs for emotional relative to neutral stimuli around 300 msec, including the P300 and N300 components – [66–68]) has been described mostly in visual studies.

Consistent with the role of P300 amplitude as an index of attention- and valence-oriented processes, the larger P300 amplitude for emotional relative to neutral condition found in our study suggests that emotional information has privileged access to processing resources [69,70] and that participants use more attentional resources when integrating emotional audiovisual information than when integrating neutral audiovisual information. Also, the more positive-going amplitude for emotional stimuli is consistent with the sensitivity of P300 to stimulus valence, reported for pictures [71] and sounds [72].

Together, these results point to three major conclusions:

(1) Given the results of the current study, we conclude that even under implicit affective processing demands, emotion automatically attracts attention and modulates perceptual (P200, N250) and attentional (N250, P300) processes devoted to audiovisual stimuli.

(2) Given the results of the current study in conjunction with the results reported in the previous studies, we concluded that the processes indexed by N100, P200, N250, and P300 components do not seem to be modality specific but rather operate on sensory percepts and their higher order representations regardless of modality. This result suggests several possibilities: cross talk between visual and auditory modalities, the feedback loop from

amodal regions of emotion processing back to primary visual and auditory cortices, and direct influence of frontally distributed regions including orbitofrontal and inferior frontal gyri devoted to higher order processing of emotional information on P200, N250, and P300.

The neural generators of these components are primarily confined to temporal structures [28,63], so it is unlikely that they might reflect activity in the frontal regions devoted to emotion processing. Furthermore, it is believed that processes supported by the frontal regions are related to emotion identification and make distinction between specific emotions possible. Thus, the lack of a distinction between specific emotions for all the components discussed here speaks against the possibility that these higher order regions contributed to the scalp recorded amplitudes in a meaningful way.

Therefore, it is more likely that the current results reflect cross talk between visual and auditory modalities, the influence of early multimodal integration regions such as superior temporal sylcus on unimodal processing areas, or both. Several recent studies indicate direct communication between visual and auditory modalities as well as modulatory impact of early multimodal integration regions. For example, Saint-Amour et al., [73] reported the presence of McGurk MMN to mismatches between auditorily presented 'ba' and visually presented 'va'. The first response was observed at 175 msec, followed by responses around 290 msec, and 350–400 msec that roughly correspond to effects observed in the current study in P200, N250 and P300 latency range. The effects in the Saint-Amour study were localized to the STG regions. Kayser et al. [74] reported the results of recordings from the auditory cortex in alert monkeys exposed to naturalistic audiovisual stimuli. The authors demonstrated that firing rates and timed spike patterns were more reliable when monkeys were watching congruent audiovisual events. Werner and Noppeney [75] using functional MRI approach showed three regions where evidence pointed to multisensory integration: primary auditory cortex, superior temporal and intraparietal sulcus, and ventrolateral prefrontal cortex. Functional connectivity analyses pointed to both direct contact between primary sensory cortices and feedback influences from the STS to primary cortices.

(3) Thus, the current results where distinction between neutral and emotional audiovisual facial expressions has been reflected across three components, P200, N250 and P300, suggest a sequence of processes of growing complexity with both feed-forward and feedback communication patterns between primary sensory regions and multimodal secondary areas. Presumably, the outputs of these processes will be analyzed further by prefrontal regions where emotional valence will be assigned. However, the ERP data obtained in this study did not reflect these processes. As suggested by Paulmann et al. [21], ERPs are not sensitive to aspects of emotion processing that are supported by frontal regions.

This study used an implicit task of emotion detection given its greater ecological validity. It is more frequent that people make judgments regarding emotional states of others within the context of other events rather than within the context of making explicit choices about emotions observed in other people. Nonetheless, it would be interesting to examine how, and to what extent, the processes observed here would change if the task of emotion identification was explicit rather than implicit. It would be also interesting to contrast congruent and incongruent emotional states expressed by faces and voices. Finally, we have tested only two emotions (angry and happy). More studies are needed to explore the differential processing of different emotion types presented through audiovisual channels.

Conclusions

In this study, we presented evidence for early and late effects of the implicit processing of affective face-and- non-semantic vocalizations. Neutral and emotional (happy and angry) stimuli were differentiated at fronto-central electrode sites as reflected in P200, N250, and P300 components.

Neutral and emotional cues were not distinguished at the level of N100, suggesting a similar sensory processing of neutral and emotional information in audiovisual channels. More positive P200 amplitude for emotional relative to neutral condition indicated increased mobilization of attentional resources towards affective stimuli, regardless of valence. These findings suggested that P200 is an index of extraction of emotional salience from stimuli, allowing a first categorization between stimuli but not the discrimination between affective categories. The increased N250 for neutral relative to emotional stimuli indexed the decoding of emotional cues for subsequent discriminative processes, possibly also suggesting more difficult integration of neutral relative to emotional cues. Finally, the increased P300 for emotional relative to neutral condition suggested that both angry and happy cues provided in audiovisual input are related to enhanced processing and attentional resources, when compared with neutral stimuli, potentially related to their motivational value.

This study provided, for the first time, ERP evidence on the time course of implicit processing of emotional cues from concurrently presented faces and voices. This is particularly relevant for the understanding of social interactions where emotional signals are often processed in an implicit way and without full conscious awareness.

Methods

1. Subjects

Twenty-two subjects participated in the study. Inclusion criteria included (a) right handedness (Edinburgh Inventory, [44]); (b) no history of neurological illness; (c) no history of alcohol or drug abuse; (d) no history of psychiatric disorder in oneself or in first-degree relatives; (e) no current medication for medical disorders that would have effects on electroencephalogram (EEG) morphology or consequences at the level or neurological and/or cognitive functioning; (f) verbal intelligence quotient (IQ) above 75; (g) no alcohol use in the 24 hours before testing; (h) no major sensory impairment that would prevent study participation (i.e., normal audition and normal or corrected vision); (i) an ability and desire to cooperate with the experimental procedures, as demonstrated by a signed given informed consent, following Harvard Medical School and Veterans Affairs Boston Healthcare System guidelines.

The subjects were recruited from local community by internet advertisements and paid for participation. Four participants were excluded from the sample due to excessive artifact in EEG data, leaving 18 subjects for subsequent data analyses. Table 1 shows the demographics characteristic of these subjects.

This study has been approved by the Institutional Review Board of the Veterans Administration Boston Health Care System and by the Institutional Review Board of the Harvard Medical School. The approved informed consent form has been signed by all subjects.

2. Stimuli

All stimuli were audiovisual and consisted of the simultaneous presentation of a face and a voice. The experimental stimuli were human faces matched with a human voice based on emotion (angry, happy or neutral), gender (male and female) and age.

Table 1. Socio-demographic characterization of participants.

Variable	Value
Age (years)	44.61±7.84
Gender (M;F)	16;2
Education (years)	14.69±2.02
Subject's SES*	2.11±0.68
Parental SES	2.45±0.86
Verbal IQ**	105.39±11.28
Full Scale IQ	103.94±13.72

*SES = socioeconomic status.
**IQ = intelligence quotient.
doi:10.1371/journal.pone.0031001.t001

Audiovisual controls were primate faces and species-specific vocalization.

The face stimuli that were a part of the audiovisual stimulus set included 108 human colored face photographs selected from the Radboud Faces Database [45] (www.rafd.nl). They consisted of 36 adult models (18 female) each displaying either an angry, happy or neutral facial expression (frontal gaze, 90°). All expressions were static. Pictures showed no facial hair, no glasses, makeup or jewelry. Face stimuli were equalized in brightness [45]. Mean valence of face emotional expressions was 2.03 (SD = 0.18) for angry faces, 4.27 (SD = 0.25) for happy faces, and 3.18 (SD = 0.33) for neutral faces (according to the validation scores published by Langner et al. [45]). Eleven primates faces (gorillas, macaques, and chimpanzees) were presented as controls. These were downloaded from free websites (image.Google.com; www.copyright-free-photos.org.uk). All faces had neutral expression and frontal gaze. Each photograph was trimmed so that only the face area was shown, and then adjusted to the same size (W×H: 326×425pixels) by using Adobe Photoshop CS4. All images were displayed centrally on a 20 inches CRT computer screen for 1500 msec at 7.4° vertical visual angle and 9.6° horizontal visual angle.

Auditory stimuli that were a part of the audiovisual stimulus set included 108 human and 7 primate digitized voices. The human voices were non-verbal vocalizations uttered by 9 male and 9 female individuals not participating in the experiment, and recorded through a microphone sitting on Sony MDR-7506 Modified Headset. The human voice set included 36 neutral, 36 happy and 36 angry sounds congruent with each of 3 emotional expressions in the human faces. For neutral sounds, 9 male and 9 female individuals were asked to make the sound 'mmm'; for happy sounds, a fragment of a happy laugh was recorded when the individuals watched a funny movie clip; and for angry sounds, each individual was instructed to growl 'humph' angrily. Each individual was asked to produce 4 samples of each type of voice. Three judges were selected to assess the sounds. The two sounds of each category that were consensually rated by the three judges as the best representative of its category (neutral, happy, angry) were used in the study. All human voices were normalized (44 kHz, 16-bit, stereo, WAV-format) by using GoldWave v5.18, and adjusted to 72 dB in sound intensity and to 1500 msec duration. Mean fundamental frequency (F0) for angry voices was 246.30 Hz (SD = 64.80), for happy voices was 272.09 Hz (SD = 80.9), and for neutral voices was 168.35 Hz (SD = 49.3). A main effect of emotion was found for F0 ($F(2,34) = 90.478$, $p<0.01$): F0 was lower for neutral relative to both angry and happy voices ($p<0.01$). Subsequently, the same sounds were presented to 10 (3 females; mean age = 31.9±9.5) subjects who did not participate in the ERP

experiment. These subjects were asked to assess the valence and arousal of each sound, by using a 9-point scale as in the Self-Assessment Manikin [46]. Mean valence ratings for angry sounds were 2.43 (SD = 1.11), for happy sounds were 7.07 (SD = 1.17), and for neutral sounds were 4.96 (SD = 0.48). Mean arousal ratings for angry sounds were 6.88 (SD = 1.12), for happy sounds were 5.89 (SD = 1.30), and for neutral sounds were 2.91 (SD = 1.51). In addition, participants were asked to indicate if each sound was "angry", "happy", "neutral", or belonged to "other" emotional category. Angry sounds were judged as "angry" by 98.15% of participants, happy sounds were rated as "happy" by 100% of participants, and neutral sounds were rated as "neutral" by 98.15% of participants.

The auditory control stimuli that were a part of monkey face-voice stimulus set included primate vocalizations (coos and shrieks). Five of primate vocalizations were downloaded from the website http://www.monkeymania.co.uk/, and two were bought from http://www.soundsnap.com/.

For all sound files, the first and last 10 msec segments were faded out to avoid the on- and offset clicking transients. Auditory stimuli were presented binaurally through Sennheiser HD 380 PRO Headphones, at a comfortable sound level for each participant.

Each human face or voice was used one time; there were 27 monkey and 108 human audiovisual combinations. The monkey audiovisual stimuli served as response targets so that the processing of face-voice pairings would not be contaminated by a motor response, and the presence of a target would insure that subjects paid attention to each stimulus. The ERPs to monkey faces were not subjected to analysis. A total of one hundred and eight human audiovisual stimuli made up three emotionally *congruent* experimental conditions: angry (36), happy (36), and neutral (36).

All stimuli (one hundred eight face-voice parings and 27 monkey face-voice pairings) were presented in random order via SuperLab 4.2. The onset of (static) face and voice were synchronized to each other: both face and voice were presented at the same time and displayed for the same duration (1500 msec).

3. Procedure

A trial started with a 200 msec fixation cross in the centre of the screen, followed by a 200 msec blank screen. Then, an audiovisual stimulus was presented for 1500 msec (face and voice presented simultaneously, with the same onset time), followed by a 1000 msec inter-stimulus interval (ISI) consisting of a blank screen (see Figure 5).

The experiment was conducted in a dimly-lit, sound-attenuated, and electrically shielded room. Subjects were seated 100 cm distance from the CRT computer screen and instructed to press a button using their right forefinger when they saw a monkey face.

4. EEG Recording and processing

Sixty-four channel EEG was collected using custom designed electrode caps from BioSemi system (Active 2). EEG was acquired in a continuous mode at a digitization rate of 512 Hz, with a bandpass of 0.01 to 100 Hz, and stored for later analysis. Blinks and eye movements were monitored via electrodes placed on left and right temples and one above the left eye. Electrode impedances were kept below 5 Kohms.

5. ERP Data Analysis

EEG data were processed offline using the BrainVision Analyzer package (Brain Products GmbH, Munich, Germany) and referenced off-line to the mean of the right and left mastoids.
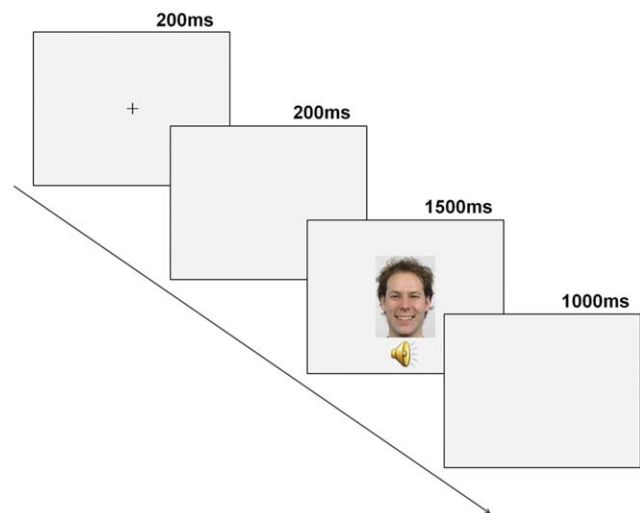
Figure 5. Schematic illustration of an experimental trial.
doi:10.1371/journal.pone.0031001.g005

Individual ERP epochs were constructed starting 150 msec pre-stimulus onset and ending 925 msec after the stimulus onset. Eye blink and movement artifacts were corrected by the method of Gratton, Coles, and Donchin (BrainVision Analyzer package) [47]. Trials containing excessive eye movements, blinks, muscle activity or amplifier blocking were rejected off-line before averaging (single-trial epochs with voltage exceeding +100/ −100 μV were rejected from further analysis). Individual averages were only considered for further analysis if at least 70% of segments available for a given condition passed the artifact rejection (all the cases reported met this criterion). Separate averages were calculated for each condition (Angry, Happy and Neutral), after subtraction of the 150 msec pre-stimulus baseline.

The mean number of segments in individual ERP averages for Angry, Happy and Neutral condition was 27.4±6.82, 28.5±5.74, and 29.2±5.14, respectively, and a one-way ANOVA did not show differences between conditions (p>0.05).

Based on visual inspection of ERP waveforms, distinct positive and negative components were identified in two different regions; in the parieto-occipital region these were: P100, N170 and P270 (see Figure 2 and 3); in the frontal-central region, N100, P200, N250 and P300 were identified (see Figures 4 and 5).

The peak amplitude and latency were measured for each component. ERP components' amplitude and latency were measured as the most positive (for positive components) or the most negative (for negative components) data point in a specified latency window – at parieto-occipital region: P100: 80–180 msec; N170: 150–230 msec; P270: 220–330 msec; at fronto-central region: N100: 80–160 msec; P200: 150–240 msec; N250: 210–320 msec; P300: 310–450 msec.

Amplitude and latency of parieto-occipital components were separately subjected to multivariate analyses (MANOVAs) with condition (angry, happy and neutral) and electrode (O1/Oz/O2, PO9/PO10, PO7/PO8) as within-subject factors.

Amplitude and latency of fronto-central components were separately subjected to MANOVA analyses with region (frontal, central), condition (angry, happy and neutral) and electrode (Fz, F1/F2, F3/F4, Cz, C1/C2, C3/C4) as within-subject factors. Wilks' Lambda was used in all MANOVA analyses. Main effects were followed with planned comparisons with Bonferroni correction.

6. Comparison between different referencing schemes

In addition to the use of the mean of the right and left mastoids as a reference, we have also analyzed all data using the average reference to examine if the posterior electrodes in particular were affected by the choice of the reference (see Table S1 and Figure S1). As can be seen in Table S1 that compares the statistical results as a function of the reference, and from Figure S1, the results for the amplitudes of the components at posterior sites, which indeed were reduced by the use of the mean of the left and right mastoid reference, did not differ between the two referencing schemes. The amplitudes of the components at fronto-central sites were affected by the use of the average reference: they were reduced with the use of the average reference. Since our primary interest was in the centro-frontal sites given the evidence that these sites are sensitive to emotional manipulations, we report and discuss the results from the mean of the left and right mastoid referencing scheme.

Supporting Information

Figure S1 The comparison of ERP waveforms obtained with the mean of left and right Mastoid-Reference (Mastoid Ref) and Average-Reference (Average Ref). Left: ERPs waveforms to each condition at Cz (top) and PO10 (bottom) created with the use of the mean of the right and left mastoids as a reference method; Right: ERPs waveforms to each condition at Cz (top) and PO19 (bottom) created with the use of common average as a reference method. Compared with Average-Ref, the use of Mastoid-Ref method yielded larger components (N1, P200, N250 and P300) at Cz, but smaller components (P1, N170 and P270) at PO10. (TIFF)

Table S1 Comparison of the MANOVA results from Mastoid-Ref and Average-Ref (Mastoid-Ref: the use of the mean of the right and left mastoids as the reference; Average-Ref: the use of the average of the active channels as reference). (DOC)

References

1. Eimer M, Holmes A (2007) Event-related brain potential correlates of emotional face processing. Neuropsychologia 45: 15–31.
2. Liu T, Pinheiro AP, Deng G, Nestor PG, McCarley RM, et al. (2012) Electrophysiological insights into processing nonverbal emotional vocalizations. NeuroReport 23: 108–12.
3. Paulmann S, Kotz SA (2008) Early emotional prosody perception based on different speaker voices. Neuroreport 19: 209–213.
4. Posamentier MT, Abdi H (2003) Processing faces and facial expressions. Neuropsychol Rev 13: 113–43.
5. Wildgruber D, Ackermann H, Kreifelts B, Ethofer T (2006) Cerebral processing of linguistic and emotional prosody: fMRI studies. Prog Brain Res 156: 249–268.
6. Kotz SA, Meyer M, Paulmann S (2006) Lateralization of emotional prosody in the brain: an overview and synopsis on the impact of study design. Prog Brain Res 156: 285–294.
7. Paulmann S, Pell MD (2009) Facial expression decoding as a function of emotional meaning status: ERP evidence. Neuroreport 20: 1603–8.
8. Focker J, Gondan M, Roder B (2011) Preattentive processing of audio-visual emotional signals. Acta Psychol (Amst) 137: 36–47.
9. Muller VI, Habel U, Derntl B, Schneider F, Zilles K, et al. (2011) Incongruence effects in crossmodal emotional integration. Neuroimage 54: 2257–2266.
10. Batty M, Taylor MJ (2003) Early processing of the six basic facial emotional expressions. Brain Res Cogn Brain Res 17: 613–620.
11. Bentin S, Allison T, Puce A, Perez A, McCarthy G (1996) Electrophysiological studies of face perception in humans. J Cogn Neurosci 8: 551–565.
12. Eimer M, Holmes A (2002) An ERP study on the time course of emotional face processing. Neuroreport 13: 427–431.
13. Ashley V, Vuilleumier P, Swick D (2004) Time-course and specificity of event-related potentials to emotional expressions. NeuroReport 15: 211–215.
14. Streit M, Ioannides AA, Liu L, Wolwer W, Dammers J, et al. (1999) Neurophysiological correlates of the recognition of facial expressions of emotion as revealed by magnetoencephalography. Brain Res Cogn Brain Res 7: 481–491.
15. Streit M, Wolwer W, Brinkmeyer J, Ihl R, Gaebel W (2001) EEG-correlates of facial affect recognition and categorisation of blurred faces in schizophrenic patients and healthy volunteers. Schizophr Res 49: 145–155.
16. Eimer M, McCarthy RA (1999) Prosopagnosia and structural encoding of faces: evidence from event-related potentials. Neuroreport 10: 255–259.
17. Liu J, Harris A, Kanwisher N (2002) Stages of processing in face perception: an MEG study. Nat Neurosci 5: 910–916.
18. An SK, Lee SJ, Lee CH, Cho HS, Lee PG, et al. (2003) Reduced P3 amplitudes by negative facial emotional photographs in schizophrenia. Schizophr Res 64: 125–135.
19. Lang PJ, Bradley MM, Cuthbert BN (1990) Emotion, attention, and the startle reflex. Psychological Review 97: 377–398.
20. Turetsky BI, Kohler CG, Indersmitten T, Bhati MT, Charbonnier D, et al. (2007) Facial emotion recognition in schizophrenia: when and why does it go awry? Schizophr Res 94: 253–263.
21. Paulmann S, Seifert S, Kotz SA (2010) Orbito-frontal lesions cause impairment during late but not early emotional prosodic processing. Soc Neurosci 5: 59–75.
22. Spreckelmeyer KN, Kutas M, Urbach TP, Altenmuller E, Munte TF (2006) Combined perception of emotion in pictures and musical sounds. Brain Res 1070: 160–170.
23. Thierry G, Roberts MV (2007) Event-related potential study of attention capture by affective sounds. Neuroreport 18: 245–248.
24. Sauter DA, Eimer M (2010) Rapid detection of emotion from human vocalizations. J Cogn Neurosci 22: 474–481.
25. Everhart DE, Demaree HA, Shipley AJ (2006) Perception of emotional prosody: moving toward a model that incorporates sex-related differences. Behav Cogn Neurosci Rev 5: 92–102.
26. Delplanque S, Silvert L, Hot P, Sequeira H (2005) Event-related P3a and P3b in response to unpredictable emotional stimuli. Biol Psychol 68: 107–120.
27. Ethofer T, Anders S, Erb M, Droll C, Royen L, et al. (2006) Impact of voice on emotional judgment of faces: an event-related fMRI study. Hum Brain Mapp 27: 707–714.
28. Kreifelts B, Ethofer T, Grodd W, Erb M, Wildgruber D (2007) Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. Neuroimage 37: 1445–1456.
29. Massaro DW, Egan PB (1996) Perceiving affect from the voice and the face. Psychon Bull Rev 3: 215–221.
30. de Gelder B, Vroomen J (2000) The perception of emotions by ear and by eye. Cogn Emot 14: 289–311.
31. de Gelder B, Bocker KB, Tuomainen J, Hensen M, Vroomen J (1999) The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses. Neurosci Lett 260: 133–136.
32. 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.
33. Pourtois G, Debatisse D, Despland PA, de Gelder B (2002) Facial expressions modulate the time course of long latency auditory brain potentials. Brain Res Cogn Brain Res 14: 99–105.
34. Frassinetti F, Bolognini N, Làdavas E (2002) Enhancement of visual perception by crossmodal visuo-auditory interaction. Exp Brain Res 147: 332–343.
35. Bolognini N, Frassinetti F, Serino A, Làdavas E (2005) "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. Exp Brain Res 160: 273–282.
36. McDonald JJ, Teder-Sälejärvi WA, Hillyard SA (2000) Involuntary orienting to sound improves visual perception. Nature 407: 906–908.
37. Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci 11: 473–490.
38. Brefczynski-Lewis J, Lowitszch S, Parsons M, Lemieux S, Puce A (2009) Audiovisual non-verbal dynamic faces elicit converging fMRI and ERP responses. Brain Topogr 21: 193–206.
39. Lang PJ, Bradley MM, Cuthbert BN (2008) International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8 University of Florida, Gainesville, FL.
40. Puce A, Epling JA, Thompson JC, Carrick OK (2007) Neural responses elicited to face motion and vocalization pairings. Neuropsychologia 45: 93–106.

41. Critchley HD, Daly EM, Phillips M, Brammer M, Bullmore E, et al. (2000) Explicit and implicit neural mechanisms for processing of social information from facial expressions: A functional magnetic resonance imaging study. Hum Brain Mapp 9: 93–105.

42. Eimer M, Holmes A, McGlone F (2003) The role of spatial attention in the processing of facial expression: An ERP study of rapid brain responses to six basic emotions. Cogn Affect Behav Neurosci 3: 97–110.

43. Pourtois G, Grandjean D, Sander D, Vuilleumier P (2004) Electrophysiological correlates of rapid spatial orienting towards fearful faces. Cereb Cortex 14: 619–33.

44. Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9: 97–113.

45. Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, et al. (2010) Presentation and validation of the Radboud Faces Database. Cogn Emot 24: 1377–1388.

46. Bradley MM, Lang PJ (2007) The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual. Technical report B-3 University of Florida, Gainesville, Fl.

47. Gratton G, Coles MG, Donchin E (1983) A new method for off-line removal of ocular artifact. Electroencephalogr Clin Neurophysiol 55: 468–484.

48. Vuilleumier P, Armony JL, Driver J, Dolan RJ (2001) Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. Neuron 30: 829–841.

49. Mishra J, Hillyard SA (2009) Endogenous attention selection during binocular rivalry at early stages of visual processing. Vision Res 49: 1073–80.

50. Woods DL (1995) The component structure of the N1 wave of the human auditory evoked potential. Electroencephalogr Clin Neurophysiol Suppl 44: 102–9.

51. Luo W, Feng W, He W, Wang NY, Luo YJ (2010) Three stages of facial expression processing: ERP study with rapid serial visual presentation. Neuroimage 49: 1857–67.

52. Crowley KE, Colrain IM (2004) A review of the evidence for P2 being an independent component process: age, sleep and modality. Clin Neurophysiol 115: 732–744.

53. Garcia-Larrea L, Lukaszewicz AC, Mauguiere F (1992) Revisiting the oddball paradigm. Non-target vs neutral stimuli and the evaluation of ERP attentional effects. Neuropsychologia 30: 723–741.

54. Lenz D, Schadow J, Thaerig S, Busch NA, Herrmann CS (2007) What's that sound? Matches with auditory long-term memory induce gamma activity in human EEG. Int J Psychophysiol 64: 31–38.

55. Novak G, Ritter W, Vaughan HG (1992) The chronometry of attention-modulated processing and automatic mismatch detection. Psychophysiology 29: 412–430.

56. Schutter DJ, de Haan EH, van Honk J (2004) Functionally dissociated aspects in anterior and posterior electrocortical processing of facial threat. Int J Psychophysiol 53: 29–36.

57. Pizzagalli DA, Lehmann D, Hendrick AM, Regord M, Pascual-Marqui RD, et al. (2002) Affective judgments of faces modulate early activity (approximately 160 ms) within the fusiform gyri. NeuroImage 16: 663–677.

58. Paulmann S, Kotz SA (2008) An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical-sentence context. Brain Lang 105: 59–69.

59. Streit M, Wolwer W, Brinkmeyer J, Ihl R, Gaebel W (2000) Electrophysiological correlates of emotional and structural face processing in humans. Neurosci Lett 278: 13–16.

60. Wynn JK, Lee J, Horan WP, Green MF (2008) Using event related potentials to explore stages of facial affect recognition deficits in schizophrenia. Schizophr Bull 34: 679–687.

61. Lebib R, Papo D, Douiri A, de Bode S, Gillon Dowens M, et al. (2004) Modulations of 'late' event-related brain potentials in humans by dynamic audiovisual speech stimuli. Neurosci Lett 372: 74–79.

62. Munte TF, Brack M, Grootheer O, Wieringa BM, Matzke M, et al. (1998) Brain potentials reveal the timing of face identity and expression judgments. Neurosci Res 30: 25–34.

63. Williams LM, Palmer D, Liddell BJ, Song L, Gordon E (2006) The 'when' and 'where' of perceiving signals of threat versus non-threat. Neuroimage 31: 458–67.

64. Curran T, Tanaka JW, Weiskopf DM (2002) An electrophysiological comparison of visual categorization and recognition memory. Cogn Affect Behav Neurosci 2: 1–18.

65. Polich J, Pitzer A (1999) P300 and Alzheimer's disease: oddball task difficulty and modality effects. Electroencephalogr Clin Neurophysiol Suppl 50: 281–287.

66. Carretie L, Iglesias J, Garcia T (1997) A study on the emotional processing of visual stimuli through event-related potentials. Brain Cogn 6:

67. Carretie L, Iglesias J, Garcia T, Ballesteros M (1997) N300, P300 and the emotional processing of visual stimuli. Electroencephalogr Clin Neurophysiol 103: 298–303.

68. Johnston VS, Miller DR, Burleson MH (1986) Multiple P3s to emotional stimuli and their theoretical significance. Psychophysiology 23: 684–694.

69. Dolcos F, Cabeza R (2002) Event-related potentials of emotional memory: encoding pleasant, unpleasant, and neutral pictures. Cogn Affect Behav Neurosci 2: 252–263.

70. Weinstein AM (1995) Visual ERPs evidence for enhanced processing of threatening information in anxious university students. Biol Psychiatry 37: 847–58.

71. Olofsson JK, Nordin S, Sequeira H, Polich J (2008) Affective picture processing: an integrative review of ERP findings. Biol Psychol 77: 247–265.

72. Dominguez-Borras J, Garcia-Garcia M, Escera C (2008) Emotional context enhances auditory novelty processing: behavioural and electrophysiological evidence. Eur J Neurosci 28: 1199–1206.

73. Saint-Amour D, De Sanctis P, Molholm S, Ritter W, Foxe JJ (2007) Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. Neuropsychologia 45: 587–597.

74. Kayser C, Logothetis NK, Panzeri S (2010) Visual enhancement of the information representation in auditory cortex. Curr Biol 20: 19–24.

75. Werner S, Noppeney U (2010) Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. J Neurosci 30: 2662–75.