BRAIN RESEARCH 1070 (2006) 160–170

available at www.sciencedirect.com

SCIENCE DIRECT®

www.elsevier.com/locate/brainres

BRAIN RESEARCH

Research Report

Combined perception of emotion in pictures and musical sounds

Katja N. Spreckelmeyer[a], *Marta Kutas*[b,c], *Thomas P. Urbach*[b], *Eckart Altenmüller*[a], *Thomas F. Münte*[d,*]

[a]*Institute of Music Physiology and Musicians' Medicine, University of Music and Drama, Hannover, Hohenzollernstr. 47, D-30161 Hannover, Germany*
[b]*Department of Cognitive Science, University of San Diego, California, La Jolla, CA 90021-0515, USA*
[c]*Department of Neurosciences, University of San Diego, California, La Jolla, CA 90021-0515, USA*
[d]*Department of Neuropsychology, University of Magdeburg, Universitätsplatz 2, Gebäude 24, D-39106 Magdeburg, Germany*

ARTICLE INFO

Article history:
Accepted 19 November 2005
Available online 5 January 2006

Keywords:
Emotion
Crossmodal integration
Event-related potential
Affective processing

ABSTRACT

Evaluation of emotional scenes requires integration of information from different modality channels, most frequently from audition and vision. Neither the psychological nor neural basis of auditory–visual interactions during the processing of affect is well understood. In this study, possible interactions in affective processing were investigated via event-related potential (ERP) recordings during simultaneous presentation of affective pictures (from IAPS) and affectively sung notes that either matched or mismatched each other in valence. To examine the role of attention in multisensory affect-integration ERPs were recorded in two different rating tasks (voice affect rating, picture affect rating) as participants evaluated the affect communicated in one of the modalities, while that in the other modality was ignored. Both the behavioral and ERP data revealed some, although non-identical, patterns of cross-modal influences; modulation of the ERP-component P2 suggested a relatively early integration of affective information in the attended picture condition, though only for happy picture–voice pairs. In addition, congruent pairing of sad pictures and sad voice stimuli affected the late positive potential (LPP). Responses in the voice affect rating task were overall more likely to be modulated by the concomitant picture's affective valence than vice versa.

1. Introduction

Judging the emotional content of a situation is a daily occurrence that typically necessitates the integration of inputs from different sensory modalities—especially vision and audition. Although the combined perception of auditory and visual inputs has been studied for some years (McGurk and MacDonald, 1976; Stein and Meredith, 1993; Welch and Warren, 1986, see also Calvert, 2001 and Thesen et al., 2004 for reviews), the multisensory perception of emotion has only relatively recently come into focus. Those studies investigating the integration of affective information have typically used emotional faces paired with emotionally spoken words (Balconi and Carrera, 2005; de Gelder and Vroomen, 2000; de Gelder et al., 1999; Massaro and Egan, 1996; Pourtois et al., 2000). Behaviorally, face–voice pairs with congruent emotional

* Corresponding author.
E-mail address: thomas.muente@uni-magdeburg.de (T.F. Münte).

0006-8993/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2005.11.075

expressions have been found to be associated with increased accuracy and faster responses for emotion judgments compared to incongruent pairs. Massaro and Egan (1996), for example, used a computer-generated "talking head" with a male actor's voice saying 'please' in a happy, neutral or angry way, while the head's face displayed either a happy, neutral or angry expression. Participants made two-alternative forced choice judgments (happy or angry) on the audio-visual percept. Reaction times increased with the degree of ambiguity between the facial and vocal expressions. The probability of judging the audio-visual performance as angry was calculated for all conditions based on participants' responses. Overall, facial expression had a larger effect on judgments than the voice. However, when the facial expression was neutral, the combined percept was influenced considerably by the expression of the voice. The authors concluded that the influence of one modality on the emotion perception depended to a large extent on how ambiguous or undefined affective information in that modality was. de Gelder and Vroomen (2000) found an overall larger effect of voice on the ratings of audio-visual presentation than that reported by Massaro and Egan (1996). Besides a possible difference between angry and sad faces with respect to salience, the different visual presentation formats may help account for the somewhat different results. Specifically, the use of moving faces by Massaro and Egan may have led to visual dominance as in the ventriloquism effect (Stein and Meredith, 1993). This possibility is supported by de Gelder and Vroomen's (2000) observation that the effect of voice was reduced, although not completely eliminated when participants were instructed to selectively attend the face and ignore the voice. They also confirmed Massaro and Egan's finding that voice information had a greater impact when facial expressions were ambiguous.

Of particular interest in the realm of audio-visual integration is the question of timing, namely, when in the processing stream does the integration actually take place? Using event-related brain potentials (ERP) to examine the time course of integrating emotion information from facial and vocal stimuli, Pourtois et al. (2000) found a sensitivity of the auditory N1 (~110 ms) and P2 (~200 ms) components to the multisensory input: N1 amplitudes were increased in response to attended angry or sad faces that were accompanied by voices expressing the same emotion, while P2 amplitudes were smaller for congruent face–voice pairs than for incongruent pairs. By presenting congruent and incongruent affective face–voice pairs with unequal probabilities, de Gelder et al. (1999) evoked auditory mismatch negativities (MMN) in response to incongruent pairs as early as 178 ms after voice onset. Both of these results suggest that interactions between affective information from the voice and the face take place before either input has been fully processed.

Considerably less effort has been directed toward the integration of emotional information from more abstractly related inputs as they typically occur in movies, commercials or music videos (but see de Gelder et al., 2004 for discussion). Though music has been found to be suitable to alter a film's meaning (Bolivar et al., 1994; Marshall and Cohen, 1988), no attempt has been made to study the mechanisms involved in the integration of emotion conveyed by music and visually complex material. We assume that integration of complex affective scenes and affective auditory input takes place later than integration of emotional faces and voices because the affective content of the former is less explicit and less salient and thereby requires more semantic analysis before their affective meaning can begin to be evaluated. Although earlier components such as the N2 have been reported to be sensitive to emotional picture valence (e.g., Palomba et al., 1997), the most commonly reported ERP effect is modulation of P3 amplitude: pictures of pleasant or unpleasant content typically elicit a larger P3 (300–400 ms) and subsequent late positive potential (LPP) than neutral pictures (Diedrich et al., 1997; Johnston et al., 1986; Palomba et al., 1997; Schupp et al., 2000). LPP amplitude also has been found to vary with the degree of arousal; both pleasant and unpleasant pictures with highly arousing contents elicit larger LPP amplitudes than affective pictures with low arousal (Cuthbert et al., 2000). The finding that affective (compared to non-affective) pictures elicit a pronounced late positive potential which is enlarged by increasing arousal has been taken to reflect intensified processing of emotional information that has been categorized as significant to survival (Lang et al., 1997). The P3 in such studies has been taken to reflect the evaluative categorization of the stimulus (Kayser et al., 2000).

Support for the notion that an integration of affective pictures of complex scenes and affective voices takes place later than integration of affective faces and voices (de Gelder et al., 2002) comes from the demonstration that the auditory N1 to fearful voices is modulated by facial expressions even in patients with striate cortex damage who cannot consciously perceive the facial expression (de Gelder et al., 2002). In contrast, pictures of emotional scenes did not modulate early ERP components even though the patients' behavioral performance indicated that the picture content had, though unconsciously, been processed. The authors suggested that while non-striate neural circuits alone might be able to mediate the combined evaluation of face–voice pairs, integrating the affective content from voices and pictures is likely to require that cortico-cortical connections with extrastriate areas needed for higher order semantic processing of the picture content be intact.

To examine the time course of integrating affective scene–voice pairs in healthy subjects, we recorded event-related brain potentials (ERP) while simultaneously presenting affective and neutral pictures with musical tones sung with emotional or neutral expression. Our aim was to assess when and to what extent the processing of affective pictures is influenced by affective information from the voice modality. In addition, we examined the relative importance of attention to this interaction by directing participants' attention to either the picture modality or the voice modality.

We hypothesized that affective information in the auditory modality can facilitate as well as impede processing of affective information in the visual modality depending on whether the emotion expressed in the voice matches the picture valence or not. Presumably congruent information enhances stimulus salience, while incongruent information leads to an ambiguous percept, thereby reducing stimulus salience. Given what is known from investigations of affective picture processing as well as from picture–voice integration in patients with striate damage, we do not expect integration to manifest in ERP components before 300 ms post-stimulus onset. Rather, we think it more likely that the simultaneously presented auditory information will have a modulating effect on the P3 and the subsequent late positive potential, assuming that significance of the pictures would be influenced by related additional information. We are less certain of what to expect when participants attend to the voice instead of the picture. The amplitude of the P3 to auditory (non-affective) oddball target stimuli co-occurring with visual stimuli is smaller in conjunction with affective faces (Morita et al., 2001) and affective pictures (Schupp et al., 1997) than with neutral visual stimuli. Such results have been interpreted as reflecting a re-allocation of attentional resources away from the auditory input to the affective pictures. Thus, it may be that the ERP pattern obtained in the attend-voice-task will differ significantly from that in the attend-picture-task.

2. Results

2.1. Behavioral results

Separate ANOVAs on two repeated measures (factor 'valence$_{att}$' [=valence in the attended modality (happy, neutral, sad)] and factor 'valence$_{unatt}$' [=valence in the unattended modality (happy, neutral, sad)]) were conducted for both rating tasks (for mean ratings and standard deviations in the 9 different conditions per task, see Table 1). In the *attend-picture-task*, we found a significant main effect of valence of the attended modality with mean ratings for happy, neutral and sad pictures being 5.71, 3.94 and 2.19, respectively (valence$_{att}$ $F(2,26) = 356.4$, $P < 0.001$). Post hoc analysis (Scheffé) revealed all categories differed significantly from each other (all $P < 0.01$). There was no main effect of the emotion expressed by the unattended voice stimuli on picture valence ratings (valence$_{unatt}$ $F(2,26) = 2.14$, $P = 0.15$) and picture valence and voice valence did not interact ($F(4,52) = 0.58$, $P = 0.64$).

In the *attend-voice-task*, mean ratings for happy, neutral and sad voice stimuli also differed as expected (4.83, 3.91 and 3.61, respectively; valence$_{att}$ $F(2,26) = 68.5$, $P < 0.001$). Post hoc analysis (Scheffé) revealed significant differences between all three categories (all $P < 0.001$). In contrast to the picture valence ratings, however, there was a significant main effect of the valence of the concurrently presented unattended picture on voice valence ratings (valence$_{unatt}$ $F(2,26) = 14.0$, $P < 0.001$). Happy voice stimuli were rated more positive when paired with a happy picture than when paired with a sad picture (5.07 versus 4.63; $t(13) = 4.77$, $P < 0.01$). The same was true for neutral voice stimuli (4.06 versus 3.65; $t(13) = 2.72$, $P < 0.05$). No reliable influence of picture valence was observed for sad voice stimuli. Nevertheless, voice valence and picture valence did not interact ($F(4,52) = 1.10$, $P = 0.36$).

Table 1 – Behavioral results

Attend-picture-task			Attend-voice-task		
Picture valence	Voice valence	Picture rating mean (SD)	Voice valence	Picture valence	Voice rating mean (SD)
Happy	Happy	5.77 (0.42)	Happy	Happy	5.07 (0.38)
Happy	Neutral	5.72 (0.45)	Happy	Neutral	4.79 (0.44)
Happy	Sad	5.65 (0.55)	Happy	Sad	4.63 (0.53)
Neutral	Happy	3.92 (0.21)	Neutral	Happy	4.06 (0.33)
Neutral	Neutral	3.92 (0.15)	Neutral	Neutral	4.01 (0.31)
Neutral	Sad	3.90 (0.20)	Neutral	Sad	3.65 (0.47)
Sad	Happy	2.19 (0.41)	Sad	Happy	3.79 (0.45)
Sad	Neutral	2.20 (0.38)	Sad	Neutral	3.61 (0.32)
Sad	Sad	2.18 (0.33)	Sad	Sad	3.42 (0.42)

Mean valence ratings for pictures in the attend-picture-task (left) and for voices in the attend-voice-task (right) for all possible picture–voice combinations.

2.2. ERP data

2.2.1. Valence effect

2.2.1.1. Attend-picture-task

2.2.1.1.1. Effect of (attended) picture valence. ERPs recorded in the attend-picture-task are depicted in Fig. 1. Responses to neutral, happy and sad pictures collapsed across voice valence are superimposed. Picture valence affected the amplitude of P2, P3 and N2b (valence$_{att}$ $F(2,26) = 8.86, 4.76, 7.23$, all $P < 0.05$) as well as the LPP ($F(2,26) = 18.78$, $P < 0.001$). Pair wise comparisons revealed that P2 was more pronounced for happy pictures than for neutral ($F(1,13) = 36.64$, $P < 0.001$) and sad ($F(1,13) = 5.42$, $P = 0.037$) pictures. Since P3, N2b and LPP effect interacted with caudality ($F(4,52) = 6.86, 3.75$, and 3.53, all $P < 0.01$), pair wise comparisons were conducted separately at prefrontal, fronto-central and parieto-occipital sites (see Table 2 for F values). Starting at 380 ms, the ERP was more positive going for happy pictures than for neutral and sad pictures at prefrontal sites. The pattern changed towards the back of the head and at parieto-occipital electrodes where both happy and sad pictures elicited equally greater positivities than did neutral pictures.

2.2.1.1.2. Effect of (unattended) voice valence. To determine what effect(s) the valence of the unattended voice stimuli had on the brain response to picture stimuli, ERPs elicited by pictures paired with different valence voices were superimposed separately for happy, neutral and sad pictures (shown for 3 midline sites in Fig. 2). A valence effect of the unattended voice modality was found for the N1 component; this effect varied with electrode location (valence$_{unatt}$ × caudality $F(4,52) = 3.90$, $P < 0.01$). At parieto-occipital sites pairing with sad voices led to reduction of the N1 amplitude compared to pairing with neutral ($F(1,13) = 11.43$, $P < 0.005$) or happy voices ($F(1,13) = 8.86$, $P = 0.011$). A main effect of voice valence was found for the P2 components (valence$_{unatt}$ $F(2,26) = 3.56$, $P = 0.043$). P2 amplitudes were larger for all pictures paired with happy than sad voices ($F(1,13) = 5.72$, $P = 0.033$) or with neutral voices (although this difference was marginally significant: $F(1,13) = 3.93$, $P = 0.069$). At fronto-central electrodes, congruent pairings of happy pictures with happy voices yielded the largest P2 amplitudes overall (compared to sad picture/happy voice: $F(1,13) = 10.05$, $P = 0.007$, and neutral picture/happy voice $F(1,13) = 36.02$, $P < 0.001$; interaction valence$_{att}$ × valence$_{unatt}$ × caudality $F(8,104) = 2.08$,

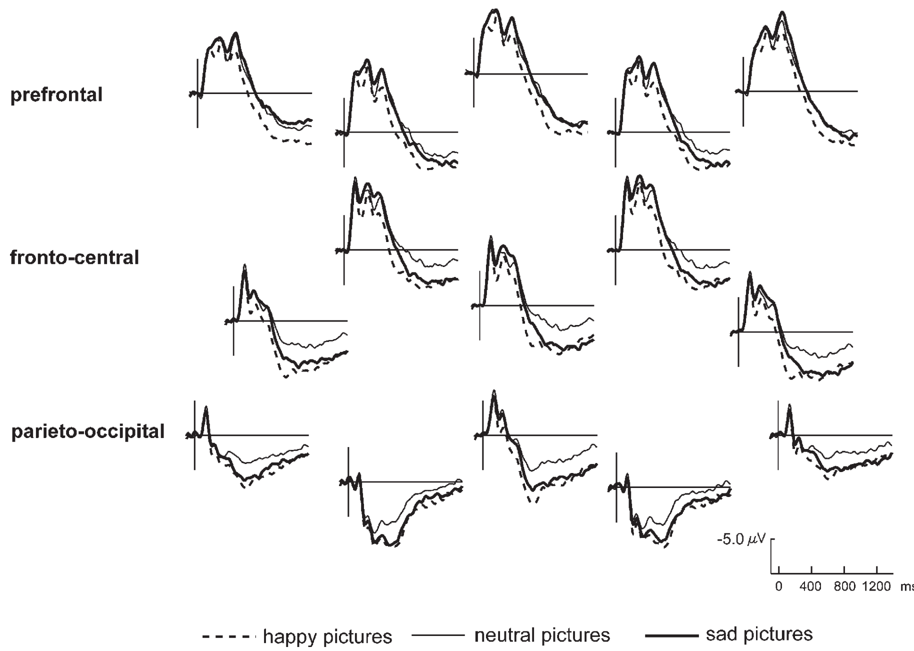

Fig. 1 – Effects of (attended) picture valence in the attend-picture-task: depicted are grand average ERPs to the three different categories of picture valence (happy, neutral, sad) at prefrontal (top two rows), fronto-central (middle two rows) and parieto-occipital electrodes (bottom two rows) (for specific locations, see Fig. 6).

P = 0.044). Finally, attended picture modality interacted with unattended voice modality between 500 and 1400 ms ($valence_{att} \times valence_{unatt}$ $F(4,52) = 2.72$, $P = 0.040$).This LPP was only affected by voice valence when sad pictures were presented. It was more pronounced in combination with a sad than with a neutral voice stimulus ($F(1,13) = 22.40$, $P = 0.000$). At prefrontal electrodes, sad pictures paired with happy voices also led to a more pronounced LPP than when paired with neutral voices (interaction with caudality ($F(2,26) = 3.54$, $P < 0.05$), but pair-wise comparison at prefrontal electrodes did not reach significance ($F(1,13) = 3.43$, $P = 0.087$).

2.2.1.2. *Voice-rating task*

2.2.1.2.1. *Effect of (unattended) picture valence*. When participants were asked to attend the voice instead of the picture, picture valence affected P3 amplitudes ($F(2,26) = 10.01$, $P < 0.001$) and N2b ($F(2,26) = 2.16$, $P < 0.05$) (see Fig. 3): P3 was greater for neutral pictures than for sad ($F(1,13) = 28.79$, $P = 0.000$) or happy ($F(1,13) = 5.62$, $P = 0.034$) pictures. The effect was largest over fronto-central electrodes (interaction with caudality ($F(4,52) = 5.32$, $P < 0.001$; see Table 2 for details). Sad pictures led to a larger N2b than happy and neutral pictures. This effect also interacted with caudality ($F(4,52) = 10.23$, $P < 0.000$), reflecting a larger effect at prefrontal sites than at

Table 2 – Effect of picture valence

		Attend-picture-task			Attend-voice-task		
		380–420 ms	420–500 ms	500–1400 ms	380–420 ms	420–500 ms	500–1400 ms
Prefrontal	Happy–neutral	7.15*	14.17**	8.86*	n.s.	n.s.	n.s.
	Happy–sad	9.69**	8.27*	6.88*	10.10**	12.96**	n.s.
	Neutral–sad	n.s.	n.s.	n.s.	25.54***	19.49**	6.29*
Fronto-central	Happy–neutral	n.s.	22.17***	81.23***	7.33*	n.s.	n.s.
	Happy–sad	11.16**	n.s.	n.s.	n.s.	n.s.	n.s.
	Neutral–sad	n.s.	n.s.	29.59***	22.53***	8.00*	n.s.
Parieto-occipital	Happy–neutral	8.45*	23.96***	18.00**	4.98*	n.s.	n.s.
	Happy–sad	n.s.	n.s.	n.s.	n.s.	n.s.	9.65**
	Neutral–sad	6.41*	7.56*	21.19**	5.11*	n.s.	14.69**

Pairwise comparison of ERP averages to pictures of different valence in the attend-picture- (left) and the attend-voice-task (right). Given are significant F values ($df = 1,13$) for comparison of mean amplitudes in the P3 (380–420 ms), N2b (420–500 ms) and LPP (500–1400 ms) time window at three levels of caudality (prefrontal, fronto-central and parieto-occipital).
n.s.—not significant.
* $P < 0.05$.
** $P < 0.01$.
*** $P < 0.001$.

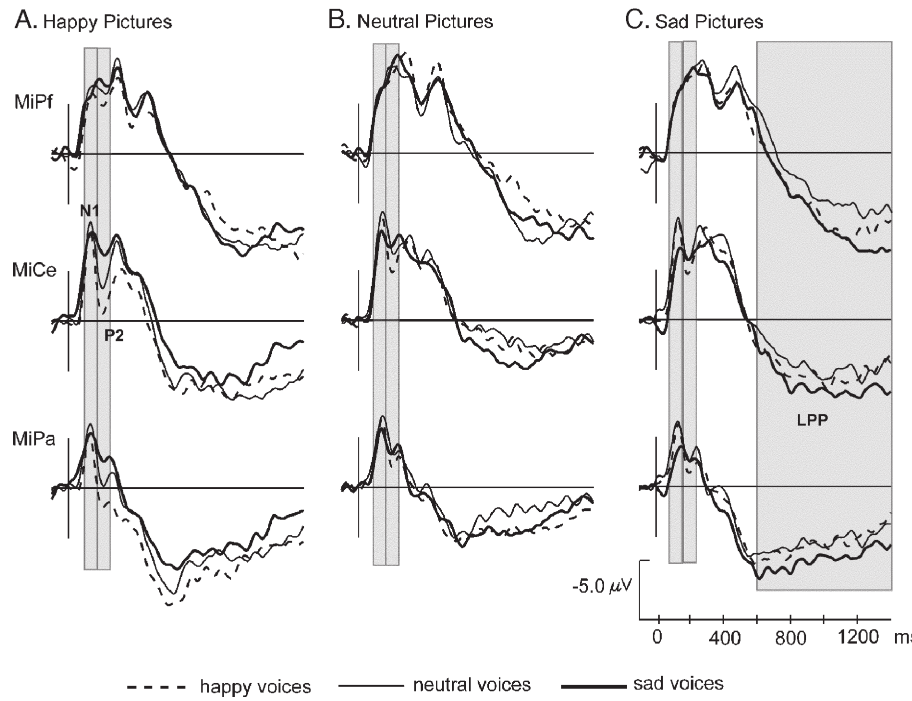

Fig. 2 – Effects of (unattended) voice valence in the attend-picture-task: grand average ERPs to the three different categories of voice valence (happy, neutral, sad), separately depicted for happy (A), neutral (B) and sad (C) pictures at three midline electrodes (MiPf = midline prefrontal, MiCe = midline central, MiPa = midline parietal). Time windows with significant effects of affective valence$_{unatt}$ or valence$_{att}$ × valence$_{unatt}$ – interaction are highlighted.

any other sites (see table for details). The LPP effect seen in the attended picture condition was reduced and interacted with caudality ($F(4,52) = 8.62, P < 0.000$). Prefrontally, neutral pictures led to a greater positive deflection than sad pictures, while parieto-occipitally, sad pictures led to a greater positivity than happy and neutral pictures (see Table 2 for details). No effect of picture valence was found for the P2 ($F(2,26) = 2.31$, n.s.).

2.2.1.2.2. Effect of (attended) voice valence. The N1 effect of voice valence reported for the attend-picture-task did not reach significance ($F(2,26) = 2.53, P = 0.099$) in the attend-voice-

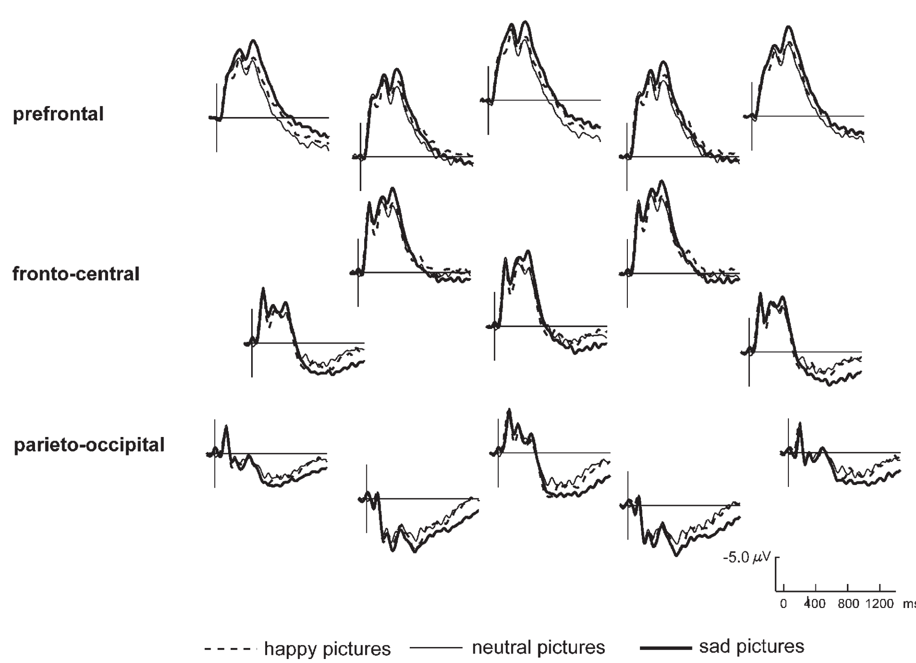

Fig. 3 – Effects of (unattended) picture valence in the attend-voice-task: depicted are grand average ERPs to the three different categories of picture valence (happy, neutral, sad) at prefrontal (top two rows), fronto-central (middle two rows) and parieto-occipital electrodes (bottom two rows) (for specific locations, see Fig. 1).

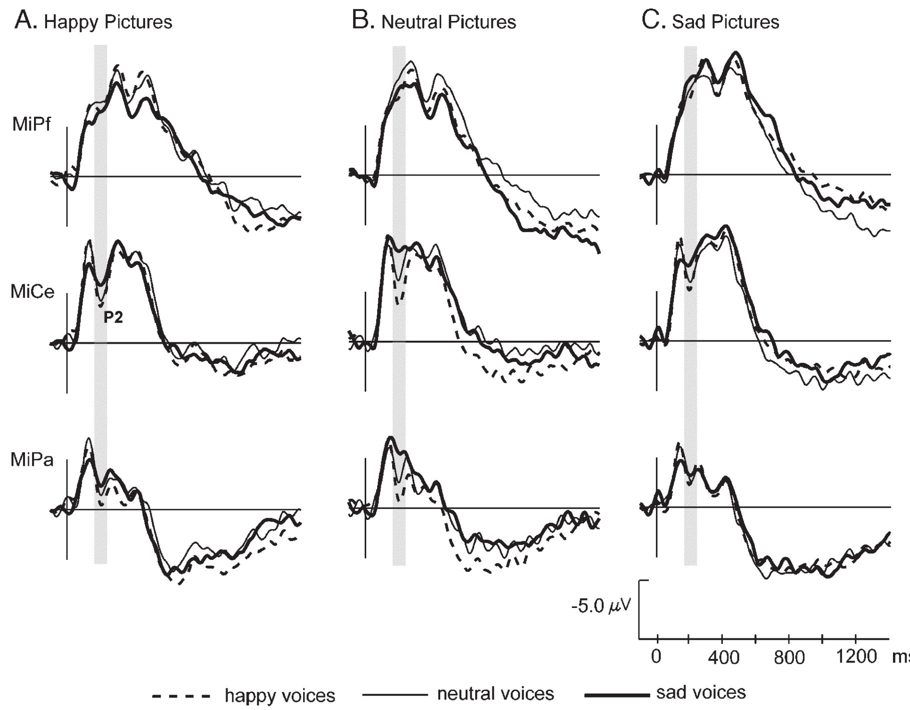

Fig. 4 – Effects of (attended) voice valence in the attend-voice-task: grand average ERPs to the three different categories of voice valence, separately depicted for happy (A), neutral (B) and sad (C) pictures at three midline electrodes (MiPf = midline prefrontal, MiCe = midline central, MiPa = midline parietal). Time windows with significant effects of affective valence$_{unatt}$ or valence$_{att}$ × valence$_{unatt}$ – interaction are highlighted.

task (Fig. 4). However, valence of the voice stimulus, now attended, had a significant main effect on P2 amplitude (valence$_{att}$ $F(2,26) = 6.19$, $P < 0.01$). Again, the P2 was more pronounced when happy voice stimuli were presented than when neutral ($F(1,13) = 7.29$, $P = 0.018$) or sad ($F(1,13) = 12.09$, $P = 0.004$) voices were presented. No effect of voice valence was found for the LPP ($F(2,26) = 1.84$, P = n.s.).

2.2.2. *Task effect*

ERPs were affected by the task manipulation. From 250 ms onwards, ERPs took a relatively more positive course when the picture was being rated than when the voice was being rated (F values for consecutive time windows starting at 250 ms (1,13): 18.93, 76.19, 148.38, 20.83, all $P < 0.000$). Between 250 and 500 ms, a main effect of caudality reflected greater positivity at parieto-occipital than at prefrontal and fronto-central leads in both tasks ($F(2,26) = 48.55, 46.08, 63.81$, all $P < 0.001$) (see Fig. 5). During the LPP, the caudality pattern interacted with task (interaction task × caudality $F(2,26) = 18.67$, $P < 0.001$), reflecting equipotential LPPs across the head in the voice-rating task and a more frontally distributed positivity in the picture rating task.

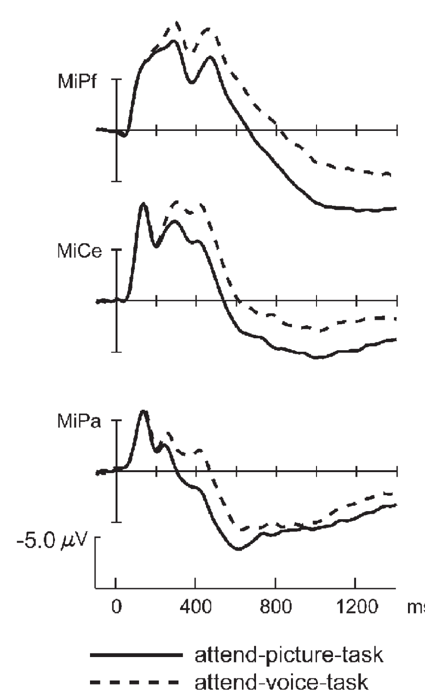

Fig. 5 – Task effect: comparison of grand average ERPs of attend-picture-task (dotted line) and attend-voice-task (solid line) at three midline electrodes (MiPf = midline prefrontal, MiCe = midline central, MiPa = midline parietal) collapsed over all conditions.

3. Discussion

While it may not be surprising that people combine facial expressions with voice tones to gauge others' emotional states, it does not necessarily follow that people's affective ratings or processing of pictures would be influenced in any way by the affective content of a concurrent but irrelevant sung note or vice versa. The current study, however, provides both behavioral and electrophysiological evidence for some interaction at the level of affect between simultaneously presented pictures and voices, even when only one of these modalities is actively attended (by instruction).

We had hypothesized that additional affective information in an unattended modality would have a certain potential to intensify or reduce affective impact of an emotional picture stimulus depending on whether its valence is congruent or incongruent with the picture valence. Although the rating of the pictures did not show a bias towards the valence of the concurrently presented voices, ERP responses indicate modified processing of picture–voice pairs with matching affective valence. Sad pictures evoked a more positive-going LPP when the accompanying voice was also sad. Congruent pairing of happy pictures and happy voices led to enlargement of the P2 component.

3.1. *P2 effect*

While we thought it likely to find modulations of ERP components known to reflect stimulus significance such as P3 and LPP, we had not, however, expected to find such an early effect of affective coherence as the P2 effect for happy picture–voice pairs. P2 is known to be an early sensory component that can be modulated by acoustical features of an auditory stimulus such as loudness or pitch (Antinoro et al., 1969; Picton et al., 1970). In fact, the main effects of voice valence on the early components N1 and P2, found in both tasks, can be linked to differences in the acoustic structure of the voice stimuli. Musical notes expressing sadness tend to have a slower tone attack, also described as longer rise time, than happy notes (see Juslin and Laukka, 2003 for a review), and increasing rise times are known to reduce the amplitude of auditory onset potentials (Elfner et al., 1976; Kodera et al., 1979). This explanation cannot, however, account for the striking asymmetry in P2 amplitude between congruent and incongruent happy picture–voice pairs. Obviously, the simultaneous presentation of the happy picture has led to enhanced processing of the happy voice, clearly indicating an early integration of the two modalities. Modulation of the P2 component has already been reported in audio-visual object recognition tasks. In designs comparing the ERP to simultaneous audio-visual presentation with the 'sum'-ERP of the unimodally presented stimuli, P2 is larger in the 'simultaneous'-ERPs (Giard and Peronnet, 1999; Molholm et al., 2002). The functional significance of this effect, however, remains unclear. Pourtois et al. (2000) reported modulation of P2 in response to emotional congruent face–voice pairs. However, the question arises: why did we find such an early effect for happy pictures but not for sad ones? It is possible that due to their specific physical structure (loud tone onset), happy voice stimuli are harder to ignore than sad or neutral voice stimuli and thus more likely to be integrated early in the visual perception process. Moreover, it is conceivable that happy pictures, too, are characterized by certain physical features such as a greater brightness and luminance than, e.g., sad pictures. It is known that certain sensory dimensions correspond across modalities, and that dimensional congruency enhances performance even when task irrelevant. For example, pitch and loudness in audition have been shown to parallel brightness in vision (Marks et al., 2003). Thus, loud and high pitched sounds that are paired with bright lights result in a better performance than incongruent pairing with dim lights. Findings that such cross-modal perceptual matches can already be made by small children has led researchers to assume similarity of neural codes for pitch, loudness and brightness (Marks, 2004; Mondloch and Maurer, 2004). However, the notion that P2 reflects such loudness–brightness correspondence would need to be studied in future experiments. The picture–voice–valence interaction vanished when the attention was shifted from pictures to voices in the attend-voice-task indicating that whatever caused the effect of picture valence on the auditory component was not an automatic process but required attending to the picture.

3.2. *LPP effect*

In line with our hypothesis, the LPP in the attend-picture-task was enhanced for sad pictures that were paired with sad voice stimuli. Based on the assumption that LPP amplitude increases with stimulus significance and reflects enhanced processing, it can be inferred that the additional congruent affective information has intensified the perceived sadness or at least made it less ambiguous. Happy pictures, too, gained enhanced processing when paired with happy voices, though only over visual areas at the back of the head. However, the latter effect did not become significant. Perhaps if the valence in the voices would have been more salient, it would have been more easily extracted automatically and had a greater influence on the ERPs to pictures. Nevertheless, our data imply that even affective information that is less naturalistically associated than faces and voices is integrated across channels. Thus, our results underline the role of emotional coherence as a binding factor.

3.3. *Effect of task*

The change of attentional focus from pictures to voices in the attend-voice-task had a considerable effect on the ERP with amplitude and topographical differences starting at around 250 ms. Both tasks elicited a late positivity starting at ~400 ms with a maximum at about 600 ms at parietal sites. Only at prefrontal and fronto-central electrodes the positivity continued to the end of the time window (1400 ms). A frontal effect with a similar time course has previously been described in response to emotional stimuli when the task specifically calls for attention to the emotional content (Johnston and Wang, 1991; Johnston et al., 1986; Naumann et al., 1992) and has been taken to reflect engagement of the frontal cortex in emotional processing (Bechara et al., 2000). However, shifting the attention away from the pictures in the voice-rating task resulted in an overall more negative going ERP. Particularly at prefrontal and frontal electrodes, P3 and LPP were largely reduced in the voice-rating task compared to the picture rating task. Naumann et al. (1992) reported a similar pattern after presenting affective words and asking two groups of participants to either rate the affective valence (emotion group) or to count the letters of the words (structure group). The resulting pronounced frontal late positive potential only present in the emotion group was interpreted as reflecting emotion specific processes. It thus seems that rating the voice valence was a suitable task to shift participants' attention away from the emotional content of the pictures. It also indicates that the frontal cortex is less involved in the evaluation of the affective voice stimuli than in evaluation of the picture. We will now discuss the effects of picture and voice valence when attention was drawn off the pictures.

The rating of the voices was considerably biased by the valence of the pictures. It seemed to have been much more difficult to fight off the impression of the picture than ignoring the voice. The bias of affective ratings of faces and voices has been reported to be stronger if the expression of the to be rated item was neutral (Massaro and Egan, 1996). Though we did not find such a relationship in the behavioral data of the voice-rating task, the ERP recording revealed larger P3 amplitudes for neutral than for happy or sad pictures. We think that this pattern reflects a shift of attentional resources. As has been suggested by others (Morita et al., 2001; Schupp et al., 1997), more attentional resources were available for the auditory stimulus (resulting in an enhanced P3) when the concurrently presented picture was not affective and/or arousing than when it was. As an additional effect of picture valence, sad pictures elicited a larger N2b than happy and neutral pictures over the front of the head. Enhanced N2b components over fronto-central electrode sites are typically observed when response preparation needs to be interrupted as in response to NoGo items in Go/NoGo tasks (Eimer, 1993; Jodo and Kayama, 1992; Pfefferbaum and Ford, 1988). Based on the finding that negative items are more likely than positive items to bias a multisensory percept (Ito and Cacioppo, 2000; Ito et al., 1998; Windmann and Kutas, 2001), we might speculate that sad pictures are more difficult to ignore and thus lead to a greater NoGo response.

The greater LPP amplitude for affective versus non-affective pictures that is characteristic for affective picture processing (Cuthbert et al., 2000; Ito et al., 1998; Palomba et al., 1997; Schupp et al., 2000) and which had been observed in the attend-picture-task appeared to be largely reduced if attention was directed away from the visual toward the auditory modality. Diedrich et al. (1997), likewise, did not find a difference between affective and neutral pictures when participants' were distracted from attending to the emotional content of the pictures by a structural processing task. In the present study, however, the effect of valence on the LPP while reduced was not completely eliminated. Prefrontally, neutral pictures were associated with a greater positive deflection than sad pictures, while parieto-occipitally, sad pictures were associated with a greater positivity than happy and neutral pictures. Against the theoretical background that LPP amplitudes to affective stimuli reflect their intrinsic motivational relevance (Cuthbert et al., 2000; Lang et al., 1997), both the parietal as well as the prefrontal effect seem to be related to the perceived valence of the multisensory presentation. However, perceived valence was not always dominated by the valence of the to-be-attended voice modality. The prefrontal effect bears some similarity to the P3 effect of picture valence discussed earlier. The valence of the voices could only be adequately processed if the evaluation was not disturbed by arousing content of affective pictures. While the dominant (sad) picture valence influences neural responses mainly over primary visual areas at the back of the head, detection of happy and sad voice tones is accompanied by enhanced positivities over prefrontal sites which, if taken at face value, reflect activity of brain areas known to be involved in the processing of emotional vocalizations (Kotz et al., 2003; Pihan et al., 2000; Wildgruber et al., 2004) as well as emotion in music (Altenmuller et al., 2002; Schmidt and Trainor, 2001). The different topographies, thus, implicate at least two separate processes, each related to modality-specific processing of affect.

4. Conclusion

We have delineated the time course of integration of affective information from different sensory channels extracted from stimuli that are only abstractly related. Our data indicate that integration of affective picture–voice pairs can occur as early as 150 ms if the valence information is salient enough. Congruent auditory information evokes enhanced picture processing. We thus demonstrated that audio-visual integration of affect is not reduced to face–voice pairs but also occurs between voices and pictures of complex scenes. Probably because the human voice is a particularly strong emotional stimulus, affective information is automatically extracted from it even if it is not task relevant. Our data further highlight the role of attention in the multisensory integration of affective information (de Gelder et al., 2004), indicating that integration of picture and voice valence require that pictures are attended.

4.1. *Notes*

Pictures used from the IAPS were 1463, 1610, 1710, 1920, 2040, 2057, 2080, 2150, 2160, 2311, 2340, 2530, 2550, 2660, 4220, 5480, 5760, 5910, 7580, 8190, 8470, 8540, 2840, 2880, 2890, 7160, 4561, 5510, 5531, 6150, 7000, 5920, 7002, 7004, 7009, 7010, 7020, 7035, 7050, 7185, 7233, 7235, 7950, 8160, 2205, 2710, 2750, 2800, 2900, 3180, 3220, 3230, 3350, 6560, 6570, 9040, 9050, 9181, 9220, 9340, 9421, 9433, 9560, 2590, 2661, 3300.

5. Experimental procedure

5.1. *Stimuli*

5.1.1. *Picture stimuli*

Picture stimuli were 22 happy, 22 neutral and 22 sad pictures from the International Affective Picture System (IAPS) (Lang et al., 1995).

Because the experimental setup required that the pictures be presented for very short durations (300–515 ms), a preexperiment was conducted to assure that the pictures could still be recognized and evaluated similarly to the reported ratings (Lang et al., 1995) even with presentation times as short as 300 ms. In the preexperiment, a larger pool of IAPS pictures (30 per emotion category) was presented to 5 different volunteers (all PhD students, age 25 to 30 years, 4 female) with duration times randomized between 302 and 515 ms. Participants were asked to rate the pictures with regard to emotional valence and arousal on 7-point scales. Participants were additionally asked to note whenever they thought the picture was too hard to recognize or too shocking. Pictures were excluded whenever any one participant's valence rating did not match Lang et al.'s rating (e.g., happy instead of sad or vice versa) or whenever anyone noted that a picture was too difficult to recognize or repulsive. The mean valence ratings of the remaining 22 pictures per category were 5.90 (SD 0.39) for happy pictures, 4.02 (SD = 0.36) for neutral pictures and 1.80 (SD 0.58) for sad pictures. Valence ratings among the three categories differed significantly as tested with an one-way ANOVA ($F(2,63) = 447.27$, $P < 0.001$) and post hoc Scheffé tests ($P < 0.001$ for all comparisons). Analogous to Lang et al. (1995), arousal ratings were higher for both happy and sad than for neutral pictures (4.29 (SD = 0.82), and 4.07 (SD = 0.84) versus 2.15 (SD = 1.21); $F(2,63) = 31.78$, $P < 0.001$; post hoc (Scheffé): $P < 0.001$ for sad versus neutral and for happy versus neutral).

5.1.2. *Voice stimuli*

Voice stimuli were generated from 10 professional opera singers and advanced singing students (5 women) asked to sing the syllable 'ha' with a happy, sad or neutral tone. From 200 different tones, twenty-two were selected for each emotional category based on the valence ratings of 10 raters (age 21–30, 5 female) on a 7-point scale (1 = extremely sad to 7 = extremely happy). The selected stimuli met the following criteria: their mean ratings were within the category boundaries (rating <3 sad, >5 happy, between 3 and 5 neutral), and they were consistently rated as happy (responses had to be 5,6 or 7), neutral (responses had to be 3,4 or 5) or sad (responses had to be 1,2 or 3) by at least 7 of 10 raters. All tones were also rated by these same participants for arousal on a 7-point scale (1 = 'not arousing at all' to 7 = 'extremely arousing'). Mean valence ratings by category were 5.23 (SD = 0.35) for happy, 3.91 (SD = 0.28) for neutral and 2.81 (SD = 0.44) for sad notes. Mean ratings between all three categories were significantly different as tested with an one-way ANOVA ($F(2,63) = 247.03$, $P < 0.001$) and post hoc Scheffé tests ($P < 0.001$ for all comparisons). Mean arousal ratings for happy, neutral and sad notes on a 7-point scale were 2.62 (SD = 0.37), 2.18 (SD = 0.28) and 2.51 (SD = 0.27), respectively. As for pictures, arousal ratings were higher for both happy and sad than for neutral notes ($F(2,63) = 12.07$, $P < 0.001$; post hoc (Scheffé): $P < 0.01$ for sad versus neutral and for happy versus neutral). Between valence categories, notes were matched for length (mean = 392 ms, SD = 60 ms) and pitch level (range: A^2–A^4). A total of 66 voice stimuli were digitized with a 44.1-kHz sampling rate and 16-bit resolution. The amplitude of all sounds was normalized to 90% so the maximum peak of a waveform was equally loud across all the notes.

5.1.3. *Picture–voice pairings*

Picture and voice stimuli were combined such that each picture was paired once with a happy, once with a neutral and once with a sad voice. Likewise, each voice stimulus was paired with a happy, a neutral and a sad picture. Thus, all pictures and all sung notes were presented three times, each time in a different combination. Picture–voice pairs were created randomly for each participant. To increase the overall number of trials, the resulting set of 198 pairs was presented twice in the experiment, each time in a different randomized order.

5.2. *Participants*

Fourteen right-handed students (age range 18–27 years, mean = 21 years (SD = 2.75), 8 women) received either money or course credit for their participation in the experiment. None of the participants considered him- or herself a musician, though some reported having learned to play a musical instrument at some point. Participants gave informed consent, and the study was approved by the UCSD Human Subjects' Internal Review Board. Prior to the experiment, participants were given a hearing test to allow for an individual adjustment of audio volume.

5.3. *Task procedure*

Participants were tested in a sound attenuating, electrically shielded chamber. They were seated 127 cm in front of a 21-in. computer monitor. Auditory and visual stimuli were presented under computer control. Each trial started with a black screen for 1600 ms. Picture and voice pairs were presented simultaneously following the presentation of a crosshair, orienting participants toward the centre of the screen. The interval between cross onset and stimulus onset was jittered between 800 and 1300 ms to reduce temporal predictability. Voice stimuli were presented via two loudspeakers suspended from the ceiling of the testing chamber approximately 2 m in front of the subjects, 0.5 m above and 1.5 m apart. Each picture remained on screen as long as the concomitant auditory stimulus (ranging from 302 to 515 ms) lasted. Pictures subtended 3.6 × 6.3° of visual angle (width × height).

Two different tasks were alternated between blocks. In the *attend-picture-task*, participants were asked to rate picture valence on a 7-point scale (ranging from 1 = very sad to 7 = very happy) while ignoring the voice stimulus. In the *attend-voice-task*, participants were asked to rate the emotional expression of the voice (sung note) on the same scale while ignoring the picture stimulus. Participants gave their rating orally after a prompt to do so appeared on the screen 1500 ms after stimulus offset. After their response had been registered, the next trial was started manually by the experimenter. Trial durations ranged between 4102 and 4815 ms. Order of task blocks was counterbalanced. Prior to the experiment, participants took part in a short practice block.

5.4. *ERP recording*

The electroencephalogram (EEG) was recorded from 26 tin electrodes mounted in an elastic cap (see Fig. 6) with reference electrodes at the left and right mastoid. Electrode impedance was kept below 5 kΩ. The EEG was processed through amplifiers set at a bandpass of 0.016–100 Hz and digitized continuously at 250 Hz.

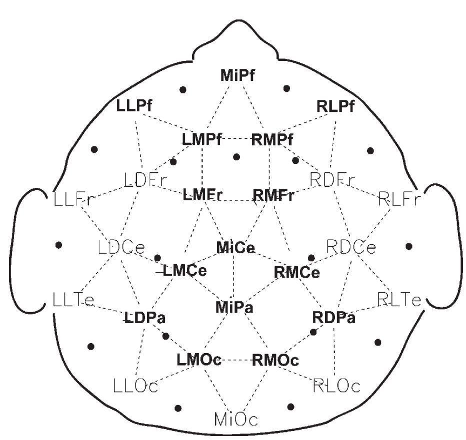

Fig. 6 – Distribution of electrode locations over the head as seen from above. Electrodes used for statistical analysis are printed in bold. Filled circles mark positions where electrodes of the 10–20 system would be (MiPf corresponds to Fpz, MiCe to Cz, MiPa to Pz and MiOc to Oz).

Electrodes were referenced on-line to the left mastoid and re-referenced off-line to the mean of the right and left mastoid electrodes. Electrodes placed at the outer canthus of each eye were used to monitor horizontal eye movements. Vertical eye movements and blinks were monitored by an electrode below the right eye referenced to the right lateral prefrontal electrode. Averages were obtained for 2048-ms epochs including a 500 ms prestimulus baseline period. Trials contaminated by eye movements or amplifier blocking or other artifacts within the critical time window were rejected prior to averaging.

ERPs were calculated by time domain averaging for each subject and each valence combination (picture–voice: happy–happy, happy–neutral, happy–sad, neutral–happy, neutral–neutral, neutral–sad, sad–happy, sad–neutral, and sad–sad) in both tasks (voice rating, picture rating).

These average ERPs were quantified by mean amplitude measures using the mean voltage of the 500 ms time period preceding the onset of the stimulus as a baseline reference. Time windows for the statistical analyses were set as follows: N1 (50–150 ms), P2 (150–250 ms), N2 (250–350 ms), P3 (380–420 ms) and N2b (420–500 ms), followed by a sustained late positive potential (LPP, 500–1400 ms). Electrode sites used for the analysis (Fig. 6, bold print) were midline prefrontal (MiPf), left and right lateral prefrontal (LLPf and RLPf) and medial prefrontal (LMPf and RMPf), left and right medial frontal (LMFr and RMFr) and medial central (LMCe and RMCe), midline central (MiCe), midline parietal (MiPa), left and right mediolateral parietal (LDPa and RDPa) and medial occipital (LMOc and RMOc).

The resulting data were entered into ANOVAs (analysis of variance). Separate ANOVAs on 4 repeated measures with within factors '$valence_{att}$' [=valence in the attended modality (happy, neutral, sad)], '$valence_{unatt}$' [=valence in the unattended modality (happy, neutral, sad)], 'laterality' (left-lateral, left-medial, midline, right-medial and right-lateral) and 'caudality' (prefrontal, fronto-central and parieto-occipital) were conducted on data from each task, followed by comparisons between pairs of conditions. To test for effects of task an additional ANOVA on 3 repeated measures [two levels of task (picture rating, voice rating), 5 levels of laterality (left-lateral, left-medial, central, right-medial and right-lateral) and 3 levels of caudality (prefrontal, fronto-central and parieto-occipital)] were performed.

Whenever there were two or more degrees of freedom in the numerator, the Huynh-Feldt epsilon correction was employed. Here, we report the original degrees of freedom and the corrected P values.

Acknowledgments

This research was supported by grants from the National Institutes of Health (HD22614 and AG08313) to MK, grants of the Deutsche Forschungsgemeinschaft to TFM (MU1311/3-3, MU1311/11-2) and EA (AI269/5-2), the BMBF (contract 01GO0202) and grants to KNS by the Gottlieb Daimler- and Karl Benz-Foundation and the Deutsche Studienstiftung.

REFERENCES

Altenmuller, E., Schurmann, K., Lim, V.K., Parlitz, D., 2002. Hits to the left, flops to the right: different emotions during listening to music are reflected in cortical lateralisation patterns. Neuropsychologia 40, 2242–2256.

Antinoro, F., Skinner, P.H., Jones, J.J., 1969. Relation between sound intensity and amplitude of the AER at different stimulus frequencies. J. Acoust. Soc. Am. 46, 1433–1436.

Balconi, M., Carrera, A., 2005. Cross-modal perception of emotion by face and voice: an ERP study. 6th International Multisensory Research Forum, Trento, Italy.

Bechara, A., Damasio, H., Damasio, A.R., 2000. Emotion, decision making and the orbitofrontal cortex. Cereb. Cortex 10, 295–307.

Bolivar, V.J., Cohen, A.J., Fentress, J.C., 1994. Semantic and formal congruency in music and motion pictures: effects on the interpretation of visual action. Psychomusicology 13, 28–59.

Calvert, G.A., 2001. Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb. Cortex 11, 1110–1123.

Cuthbert, B.N., Schupp, H.T., Bradley, M.M., Birbaumer, N., Lang, P.J., 2000. Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. Biol. Psychol. 52, 95–111.

de Gelder, B., Vroomen, J., 2000. The perception of emotions by ear and by eye. Cogn. Emot. 14, 289–311.

de Gelder, B., Bocker, K.B., 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.

de Gelder, B., Pourtois, G., Weiskrantz, L., 2002. Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. Proc. Natl. Acad. Sci. U. S. A. 99, 4121–4126.

de Gelder, B., Vroomen, J., Pourtois, G., 2004. Multisensory perception of affect, its time course and its neural basis. In: Calvert, G.A., Spence, C., Stein, B.E. (Eds.), Handbook of Multisensory Processes. MIT, Cambridge, MA, pp. 581–596.

Diedrich, O., Naumann, E., Maier, S., Becker, G., Bartussek, D., 1997. A frontal positive slow wave in the ERP associated with emotional slides. J. Psychophysiol. 11, 71–84.

Eimer, M., 1993. Effects of attention and stimulus probability on ERPs in a Go/Nogo task. Biol. Psychol. 35, 123–138.

Elfner, L.F., Gustafson, D.J., Williams, K.N., 1976. Signal onset and task variables in auditory evoked potentials. Biol. Psychol. 4, 197–206.

Giard, M.H., Peronnet, F., 1999. Auditory–visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J. Cogn. Neurosci. 11, 473–490.

Ito, T., Cacioppo, J.T., 2000. Electrophysiological evidence of implicit and explicit categorization processes. J. Exp. Soc. Psychol. 36, 660–676.

Ito, T.A., Larsen, J.T., Smith, N.K., Cacioppo, J.T., 1998. Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations. J. Pers. Soc. Psychol. 75, 887–900.

Jodo, E., Kayama, Y., 1992. Relation of a negative ERP component to response inhibition in a Go/No-go task. Electroencephalogr. Clin. Neurophysiol. 82, 477–482.

Johnston, V.S., Wang, X.T., 1991. The relationship between menstrual phase and the P3 component of ERPs. Psychophysiology 28, 400–409.

Johnston, V.S., Miller, D.R., Burleson, M.H., 1986. Multiple P3s to emotional stimuli and their theoretical significance. Psychophysiology 23, 684–694.

Juslin, P.N., Laukka, P., 2003. Communication of emotions in vocal expression and music performance: different channels, same code? Psychol. Bull. 129, 770–814.

Kayser, J., Bruder, G.E., Tenke, C.E., Stewart, J.E., Quitkin, F.M., 2000. Event-related potentials (ERPs) to hemifield presentations of emotional stimuli: differences between depressed patients and healthy adults in P3 amplitude and asymmetry. Int. J. Psychophysiol. 36, 211–236.

Kodera, K., Hink, R.F., Yamada, O., Suzuki, J.I., 1979. Effects of rise time on simultaneously recorded auditory-evoked potentials from the early, middle and late ranges. Audiology 18, 395–402.

Kotz, S.A., Meyer, M., Alter, K., Besson, M., von Cramon, D.Y., Friederici, A.D., 2003. On the lateralization of emotional

prosody: an event-related functional MR investigation. Brain Lang. 86, 366–376.
Lang, P.L., Bradley, M.M., Cuthbert, B.N., 1995. International Affective Picture System (IAPS): Technical Manual and Affective Ratings. The Center for Research in Psychophysiology, University of Florida, Gainesville, FL.
Lang, P., Bradley, M.M., Cuthbert, B.N., 1997. Motivated attention: affect, activation and action. In: Lang, P., Simons, R.F., Balaban, M. (Eds.), Attention and Orienting: Sensory and Motivational Processes. Erlbaum, Hillsdale, NJ, pp. 97–136.
Marks, L.E., 2004. Cross-modal interactions in speeded classification. In: Calvert, G.A., Spence, C., Stein, B.E. (Eds.), Handbook of Multisensory Processes. MIT Press, Cambridge, MA, pp. 85–106.
Marks, L.E., Ben-Artzi, E., Lakatos, S., 2003. Cross-modal interactions in auditory and visual discrimination. Int. J. Psychophysiol. 50, 125–145.
Marshall, S.K., Cohen, A.J., 1988. Effects of musical soundtracks on attitudes toward animated geometric figures. Music Percept. 6, 95–112.
Massaro, W., Egan, P.B., 1996. Perceiving affect from the voice and the face. Psychodin. Bull. Rev. 3, 215–221.
McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. Nature 264, 746–748.
Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., Foxe, J.J., 2002. Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. Brain Res. Cogn. Brain Res. 14, 115–128.
Mondloch, C.J., Maurer, D., 2004. Do small white balls squeak? Pitch-object correspondences in young children. Cogn. Affect. Behav. Neurosci. 4, 133–136.
Morita, Y., Morita, K., Yamamoto, M., Waseda, Y., Maeda, H., 2001. Effects of facial affect recognition on the auditory P300 in healthy subjects. Neurosci. Res. 41, 89–95.
Naumann, E., Bartussek, D., Diedrich, O., Laufer, M.E., 1992. Assessing cognitive and affective information processing functions of the brain by means of the late positive complex of the event-related potential. J. Psychophysiol. 6, 285–298.
Palomba, D., Angrilli, A., A, M., 1997. Visual evoked potentials, heart rate responses and memory to emotional pictorial stimuli. Int. J. Psychophysiol. 27, 55–67.
Pfefferbaum, A., Ford, J.M., 1988. ERPs to stimuli requiring response production and inhibition: effects of age, probability and visual noise. Electroencephalogr. Clin. Neurophysiol. 71, 55–63.
Picton, T.W., Goodman, W.S., Bryce, D.P., 1970. Amplitude of evoked responses to tones of high intensity. Acta Otolaryngol. 70, 77–82.
Pihan, H., Altenmuller, E., Hertrich, I., Ackermann, H., 2000. Cortical activation patterns of affective speech processing depend on concurrent demands on the subvocal rehearsal system. A DC-potential study. Brain 123 (Pt. 11), 2338–2349.
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.
Schmidt, L.A., Trainor, L.J., 2001. Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. Cogn. Emot. 15, 487–500.
Schupp, H.T., Cuthbert, B.N., Bradley, M.M., Birbaumer, N., Lang, P.J., 1997. Probe P3 and blinks: two measures of affective startle modulation. Psychophysiology 34, 1–6.
Schupp, H.T., Cuthbert, B.N., Bradley, M.M., Cacioppo, J.T., Ito, T., Lang, P.J., 2000. Affective picture processing: the late positive potential is modulated by motivational relevance. Psychophysiology 37, 257–261.
Stein, B.E., Meredith, M.A., 1993. The Merging of the Senses. The MIT Press, Cambridge, MA.
Thesen, T., Vibell, J.F., Calvert, G.A., Österbauer, R.A., 2004. Neuroimaging of multisensory processing in vision, audition, touch and olfaction. Cogn. Process. 5, 84–93.
Welch, R.B., Warren, D.H., 1986. Intersensory interactions. In: Boff, K.R., Kaufman, L., Thomas, J.P. (Eds.), Handbook of Perception and Human Performance. Sensory Processes and Perception, vol. I. Wiley, New York, pp. 25/1–25/36.
Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., Ackermann, H., 2004. Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. Cereb. Cortex 14, 1384–1389.
Windmann, S., Kutas, M., 2001. Electrophysiological correlates of emotion-induced recognition bias. J. Cogn. Neurosci. 13, 577–592.