Investigating the Multimodal Nature of Human Communication

Insights from ERPs

Silke Paulmann[1], Sarah Jessen[2], and Sonja A. Kotz[2]

[1]McGill University, School of Communication Sciences and Disorders, Montreal, Quebec, Canada,
[2]Max Planck Institute for Human Cognitive and Brain Sciences, Research Group "Neurocognition of Rhythm in Communication," Leipzig, Germany

Abstract. The multimodal nature of human communication has been well established. Yet few empirical studies have systematically examined the widely held belief that this form of perception is facilitated in comparison to unimodal or bimodal perception. In the current experiment we first explored the processing of unimodally presented facial expressions. Furthermore, auditory (prosodic and/or lexical-semantic) information was presented together with the visual information to investigate the processing of bimodal (facial and prosodic cues) and multimodal (facial, lexic, and prosodic cues) human communication. Participants engaged in an identity identification task, while event-related potentials (ERPs) were being recorded to examine early processing mechanisms as reflected in the P200 and N300 component. While the former component has repeatedly been linked to physical property stimulus processing, the latter has been linked to more evaluative "meaning-related" processing. A direct relationship between P200 and N300 amplitude and the number of information channels present was found. The multimodal-channel condition elicited the smallest amplitude in the P200 and N300 components, followed by an increased amplitude in each component for the bimodal-channel condition. The largest amplitude was observed for the unimodal condition. These data suggest that multimodal information induces clear facilitation in comparison to unimodal or bimodal information. The advantage of multimodal perception as reflected in the P200 and N300 components may thus reflect one of the mechanisms allowing for fast and accurate information processing in human communication.

Keywords: P200, N300, ERPs, facilitation

Introduction

Human communication entails the processing of different information sources such as facial expressions, gestures, tone of voice, or words. While each of these sources – or information channels – can convey valuable information in their own right, it is a widely held belief that the combination of information channels can lead to improved and more successful communication. For instance, after delivering a sad message on the phone, the bearer of the message would often say "If I had only seen their face when they said they were fine," promoting the idea that the combined information from multiple sources would in fact facilitate understanding of the situation. The question is: Does empirical evidence support such an example, that is, do we really observe differences in neural correlates for multimodal and unimodal information processing? To answer this question, the present study investigated the time-course and neural correlates of unimodal, bimodal, and multimodal stimulus processing.

Typically, a setting in which more than one information source becomes relevant is *emotional* communication (see example above). For instance, we can convey our state of mind – whether we feel angry, happy, or neutral – by verbal (lexical-semantic) and nonverbal channels (gestures, facial expressions, prosody). There is a broad literature suggesting that so-called basic emotional (including neutral) categories can be successfully recognized by only one channel (e.g., face or prosody; see e.g., Borod et al., 2000; Ekman, 1992; Elfenbein & Ambada, 2002; Paulmann, Pell, & Kotz, 2008). In addition, there is evidence suggesting that recognizing different emotional categories is even more successful when the information is multimodal (de Gelder & Vroomen, 2000; Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007. This suggests that multiple information sources facilitate the forming of a unified representation of an event relative to a single information source. Given the growing body of literature on emotional communication, we chose to investigate unimodal, bimodal, and multimodal information processing with emotional (angry, happy) and neutral stimuli. This allowed investigating both the different levels of information processing and the possibility that emotional stimuli are processed differently from neutral stimuli when presented unimodal, bimodal, or multimodally.

Electrophysiological Investigations of (Emotional) Vocal Expressions

To ensure successful human (emotional) communication, incoming auditory information must be processed and integrated rapidly. ERPs have proven to be an excellent tool to investigate the time-course of these processes. Specifically, two ERP components have been identified to reflect early vocal expression processing, namely, the N100 and P200. The auditory N100 component, a negative-going ERP component occurring between approximately 75–150 ms poststimulus onset is associated with the processing of sensory information such as frequency and intensity of a stimulus. This component does not vary as a function of emotionality of a stimulus (Pourtois, de Gelder, Vroomen, Rossion, & Crommelinck, 2000). In contrast, the P200 is modulated by the emotional quality of a stimulus. Specifically, emotional utterances can be distinguished from neutral utterances 200 ms after sentence onset irrespective of speaker voice (Paulmann & Kotz, 2008a; Paulmann, Schmidt, Pell, & Kotz, 2008). This is true for utterances that do not contain emotional lexical-semantic information (Paulmann, 2006). The ones that do not contain lexical-semantic information offer the unique possibility to investigate pure (emotional) prosodic processing.

In addition, auditory (emotional) stimuli have also been reported to elicit a negativity, the N300. For instance, Bostanov and Kotchoubey (2004) investigated the recognition of affective prosody using emotional exclamations as stimuli. The authors report an enhanced N300 for contextually incongruous exclamations in contrast to congruous exclamations. In contrast to the P200 component, which has been argued to reflect detection and processing of stimulus properties/features such as pitch (Pantev, Elbert, Ross, Eulitz, & Terhardt, 1996), intensity (Picton, Woods, Barribeau-Braun, & Healy, 1977), valence (Paulmann & Kotz, 2008a), or arousal (Paulmann & Kotz, 2006), the N300 has been linked to a more subsequent in-depth, "meaning-related" evaluation of the stimulus (Bostanov & Kotchoubey, 2004; Lebib et al., 2004; Wu & Coulson, 2007).

Electrophysiological Investigations of (Emotional) Facial Expressions

According to recent research, the time-course of early facial expression processing is comparable to the time-course of early vocal expression processing. Facial expressions also elicit an N100 response, again interpreted to reflect the processing of physical components of a stimulus (Federmeier & Kutas, 2002; Luck & Hillyard, 1994; Vogel & Luck, 2000). Also, similar to the auditory modality, earlier ERP components (N100, P100) seem to be unaffected by the emotionality of the stimuli, while slightly later occurring components such as the P200 (Pizzagalli, Regard, & Lehman, 1999), the N200 (e.g., Campanella et al., 2002), and the N230 (e.g., Balconi & Pozzoli, 2003) are generally found to be responsive to the *emotional features* of a visual stimulus. Like the early N100, the P200 is found at anterior electrode-sites in the ERP and is associated with the extraction of basic visual features (Federmeier & Kutas, 2002). Comparable to the auditory P200, the visual P200 has been functionally linked to the emotional salience of a stimulus (Schutter, de Haan, & van Honk, 2004).

In addition, the processing of visual stimuli is also reported to elicit a later negative component, the anteriorly distributed N300, which is typically associated with the evaluation of the semantic content of a visual stimulus (Hamm, Johnson, & Kirk, 2002). However, Carrietié and colleagues also report the N300 reflecting the processing of arousing stimulus properties (Carrietié, Iglesias, & García, 1997) as well as a differentiation according to the valence of emotional stimuli (Carretié, Iglesias, García, & Ballesteros, 1997). In the context of the perception of facial emotional expressions, the N300 has been attributed to the evaluation of the emotional content of a facial expression (e.g., Carretié & Iglesias, 1995). Comparable to the auditory literature, it has thus been suggested that the P200 is indicative of detecting (emotional) stimulus properties, while the N300 is associated with any subsequent evaluation (Schutter et al., 2004).

Electrophysiological Investigations of Multimodally Presented (Emotional) Expressions

As previously outlined, visual and auditory emotional information alone provide a fairly reliable estimate of our counterpart's emotional state. Both modalities have been linked to specific neural correlates reflecting both early perceptual (P200, N200, N230) and, later, more cognitively based processing (N300). However, we often perceive information from several modalities at the same time, arguably leading to a more holistic perception of a stimulus. Indeed, results of several studies suggest that coherent and simultaneous presentation of multiple channels leads to *facilitated information processing*, reflected in shorter reaction times (e.g., de Gelder, Vrommen, & Pourtois, 1999) and increased accuracy during emotion classification (e.g., Kreifelts et al., 2007). Analogously, in the case of incoherent information processing, interference effects are observed, which generally manifest themselves in slower reaction times and decreased accuracy in emotion classification (for detailed reviews on multimodal perception see Calvert, Brammer, & Iversen, 1998; Campbell, 2007; Miller & D'Esposito, 2005).

Although still scarce, there is some electrophysiological evidence suggesting that the integration of different modalities, which may lead to faster and more accurate processing, already occurs during very early processing stages and

thus may be based on physical stimulus properties rather than the concept evaluation of a multimodal stimulus (e.g., Giard & Peronnet, 1999; Pourtois et al., 2000). For example, Stekelenburg and Vroomen (2007) observed an effect of multimodality on the N100 and the P200 component comparing speech and nonspeech. They report a *decrease* in amplitude and latency for the presentation of congruent auditory and visual stimuli (Experiment 1). In addition, a study by van Wassenhove, Grant, and Poeppel (2005) investigated the time-course of neural correlates for auditory and audio-visual speech and discovered *reduced* amplitudes for early auditory components (N1/P2 complex) for audio-visual speech in contrast to audio-only speech.

There is also some evidence for early channel integration in the emotional literature. For instance, Pourtois et al. (2000) compared the processing of sad and angry facial expressions, paired with either prosodically matching or mismatching utterances. The authors observed a *larger* amplitude of the auditory N100 component for congruent sad and angry audiovisual stimuli in contrast to incongruent audiovisual stimuli, suggesting an early valence independent integration of the two modalities. In fact, given that the N100 is known to be a component primarily influenced by physical attributes of the stimuli, one may hypothesize that this early integration is only physically motivated. Similar evidence for an early integration of emotional audiovisual information was found by de Gelder, Böcker, Tuomainen, Hensen, and Vroomen (1999), who observed a mismatch negativity effect (MMN) around 180 ms after stimulus onset for the combination of angry voices with sad faces in comparison to a combination of angry voices with angry faces. This integration seems to be a mandatory process as it is observed independent of attention (de Gelder et al., 1999). Finally, in a recent study, Pourtois, Debatisse, Despland, and de Gelder (2002) presented emotionally intoned words with either congruent or incongruent emotional facial expressions. Their results revealed a *shorter latency* for the P2b component for congruent in contrast to incongruent pairs.

Taken together, all these studies provide evidence for an early valence and emotional category independent integration of visual and auditory information. This early integration was found for both unattended (MMN studies) and attended (N100, P2b) stimuli. Consequently, it is assumed that this early integration may lead to faster and more accurate processing of stimuli with more than one information channel. A review of the studies above reveals that the studies investigating very early integration phenomena (as reflected in the N100) report *increased amplitudes* for congruent stimuli in contrast to incongruent stimuli, while studies investigating integration at a later point in time (after physical stimulus properties have been analyzed) report *amplitude reductions and shorter latencies* for channel congruent in contrast to channel incongruent stimuli (but see Experiment 2 in Stekelenburg & Vroomen, 2007). The authors report stronger P200 amplitude reduction effects (as measured at CZ) for bimodal incongruent stimuli in contrast to unimodal stimuli, than for bimodal congruent stimuli in contrast to unimodal stimuli. Interestingly, latency differences are not observed, suggesting that interference effects do not manifest themselves in latency differences.

It is worth noting that facilitation effects for multimodal stimuli are also observed at slightly later stages of processing (when arguably a more in-depth evaluation of the stimulus has already taken place). For instance, Lebib and colleagues (2004) report that incongruently dubbed audiovisual stimuli elicit a larger N300 component than congruently dubbed stimuli. The increase in amplitude is considered to reflect the difficulty in integrating the two incoherent information sources. Similar evidence comes from Wu and Coulson (2007), who report N300 amplitude reductions in response to probes that were congruent to co-speech gestures as opposed to probes that were only congruent to the preceding speech stream but not the accompanying gesture. In short, multimodal information processing in both emotional and neutral contexts seems to enact a facilitation in comparison to unimodal information processing, as reflected in behavioral (e.g., de Gelder & Vroomen, 2000; Kreifelts et al., 2007) and electrophysiological studies (e.g., de Gelder et al., 1999; Lebib et al., 2004; Pourtois et al., 2000; Wu & Coulson, 2007).

The Present Investigation

While previous studies have clearly provided valuable information about the *interplay* of different channels and the concept of integration in particular, we are still in need of studies that investigate whether the *time-course* differs for the neural correlates for unimodal, bimodal, and multimodal processing. This is particularly true for emotional communication contexts, since an integration of different channels has sometimes been observed to occur before (N100; Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005) emotional stimulus characteristics (valence, arousal, meaning) could have been evaluated (P200, N300; Lebib et al., 2004; Paulmann & Kotz, 2006, 2008a; Wu & Coulson, 2007). Hence, the question arises whether unimodal, bimodal, and multimodal processing differs between emotional and neutral communication at later stages (P200, N300), a question addressed by the present study. To our knowledge, previous studies have failed to provide insight into this issue since they seldom present unimodal, bimodal, and multimodal stimuli simultaneously from different emotional and neutral categories. The current study investigated the nature and the time-course of emotional and neutral processing in multimodal, bimodal, and unimodal conditions. In particular, we added communication channels stepwise in order to investigate how the addition of modalities influences perception in emotional and neutral communication settings. We investigated one-channel information processing by means of facial expressions, pri-

marily because these stimuli have been argued to be more easily identified and processed than auditory information (e.g., Adolphs, 2002). In a second step, we presented pseudo-sentences (that is sentences that convey no lexical-semantic information) in combination with the same pictures. Hence, we provided information by two channels, that is, via facial expressions and via prosody. As a last step, we presented facial expressions with vocal expressions containing lexical information. Information was thus presented via a total of three channels: the facial, prosodic, and lexical channel. If adding channels of congruent information causes facilitation not only during offline processing (e.g., de Gelder et al., 1999; Kreifelts et al., 2007), but also during online processing, one would predict increased facilitation with a rising number of information channels, which should be indicated by a change in latency (the more channels, the faster a stimulus is processed) and amplitude (the more channels, the lower the amplitude) of associated components, that is, the P200 (first perceptual processing; see Stekelenburg & Vroomen, 2007; van Wassenhove et al., 2005) and subsequent ERPs such as the N300 (evaluation process). Finally, by presenting both emotional and neutral stimuli, we were also able to investigate whether processing information from multiple sources follows a different time-course in emotional and nonemotional communication settings. If so, one could expect emotional stimuli to be processed more rapidly than nonemotional stimuli, since perception of emotional stimuli are argued to engage a faster, more direct processing route than of nonemotional stimuli perception (e.g., LeDoux, 2000).

Participants and Methods

Participants

Twenty-four right-handed native speakers of German (12 female, mean age: 25.7, *SD* = 2.1) participated in the study, though one participant had to be excluded due to excessive movement artefacts. None of the participants reported any hearing impairments, and all had normal or corrected-to-normal vision. Participants gave their written informed consent, and the experiment was approved by the Ethics Committee. All participants were compensated financially for their participation.

Stimulus Material

Portrayals were elicited from two native German speakers (1 male, 1 female) who had experience in acting. Recordings were made with a digital camcorder connected to a high quality clip-on microphone. During the recording session, actors produced lexical or so-called pseudo-sentences belonging to either a positive (happy), negative (angry), or neutral category. Pseudo-sentences are sentences that contain prosodic information in the absence of semantic content. Stimuli were prepared to fit one of three conditions defined by the number of information channels: A *unimodal condition* in which only one channel of information was present (face), a *bimodal condition* in which two channels were present (prosody + face, and as a control, or filler condition also prosody + lexic, though the latter did not enter statistical analysis), and a *multimodal condition* which presented all three channels of information simultaneously (prosody + lexic + face).

Unimodal Condition

Face stimuli were constructed by extracting still images from the short videoclips. Five different black-and-white photographs of different facial expressions were created, amounting to a total of 30 pictures (5 male-angry, 5 male-happy, 5 male-neutral, 5 female-angry, 5 female-happy, 5 female-neutral); see Figure 1 for examples. The expressions were rated by 30 participants in a rating study. In the rating study, facial expressions from six different emotions (anger, disgust, fear, happy, pleasant surprise, sad) and neutral were presented in a forced-choice paradigm. Emotional facial recognition in this study was clearly above chance (at least 4× above chance) for all emotions. The mean percentage correct for angry expressions was 63.33% (*SD* = 12.3), for happy the mean percentage correct was 93.83% (*SD* = 9.3), and the mean percentage correct for neutral was 89.83% (*SD* = 10.0).

Bimodal Condition

The prosody + face condition was created by simultaneously presenting prosodic stimuli and face stimuli (see above). The prosodic stimuli were created by extracting 30 phonotactically and morpho-syntactically legal pseudo-sentences without meaningful lexical-semantic information from the recordings (example: Mon set die Brelle nogeferst and ingerafen), amounting to a total of 90 sentences (30 sentences per category, 15 each spoken by the male and 15 each spoken by the female speaker). These sentences were rated by 24 participants (12 female). Again, a forced-choice paradigm was applied. The mean percentage correct for angry was 94.2% (*SD* = 6.45), for happy the mean percentage correct was 69.4% (*SD* = 8.55), and the mean percentage correct for neutral was 96.9% (*SD* = 3.09). Further rating details can be found in Pell, Monetta, Paulmann, and Kotz (2009).

In addition, we created a filler, or control, condition that also contained two channels of information, namely, prosodic and semantic information. These stimuli were constructed by extracting the audio track of the videos, in which the actors produced lexical sentences (e.g., anger: Er hat das Paar gereizt und aufgebracht [He has teased and upset the couple]; happy: Sie hat die Trauung verkündet

Table 1. Acoustic features of auditory stimuli used in the experiment. Pitch range (range f0) is given in Hz, amplitude range in dB, and speech rate is given in syllables spoken per second. Standard deviations of measurements are in brackets

	Emotion	Range f0	Range amplitude	Speech rate
Lexical sentences	Anger	213.49 (43.53)	56.04 (3.28)	4.13 (0.4)
	Happiness	213.31 (48.14)	53.32 (0.81)	4.27 (0.42)
	Neutral	118.10 (45.89)	53.09 (0.8)	3.78 (0.41)
Pseudo-sentences	Anger	203.43 (46.2)	49.68 (4.58)	4.04 (4.48)
	Happiness	288.08 (44.68)	50.41 (4.95)	4.31 (4.35)
	Neutral	157.25 (36.41)	44.03 (2.59)	3.47 (3.59)

und gelächelt [She has announced the wedding and smiled]; neutral: Er hat die Pflanzen gegossen und beschnitten [He has watered the plants and cut them]). Again, for each category (happy, angry, neutral), 30 different sentences were extracted, 15 spoken by a male and 15 spoken by a female speaker, amounting to a total of 90 recordings. All sentences were syntactically similar (SVO) and the verbs as well as nouns of the sentences were controlled for word letter length, syllable length, word frequency, initial sounds, and plosive consonants. The recordings were rated in a forced-choice paradigm for emotional valence by 64 listeners (32 female). Angry sentences were recognized correctly at 97.7% (*SD* = 4.2), happy sentences at 79.9% (*SD* = 8.6) correct and neutral sentences at 97.9% (*SD* = 3.0) (see Paulmann, Pell, & Kotz, 2008, for rating details).

Multimodal Condition

The prosody + lexic + face stimuli consisted of the audio track from the actors producing lexically meaningful sentences belonging to one of the emotional categories or neutral paired with the respective matching still images. See Table 1 for details on acoustic properties of auditory stimuli and Figure 1 for examples of facial stimuli.

Because each condition consisted of 90 stimuli, a total of 360 stimuli were presented. Each pseudo-sentence was

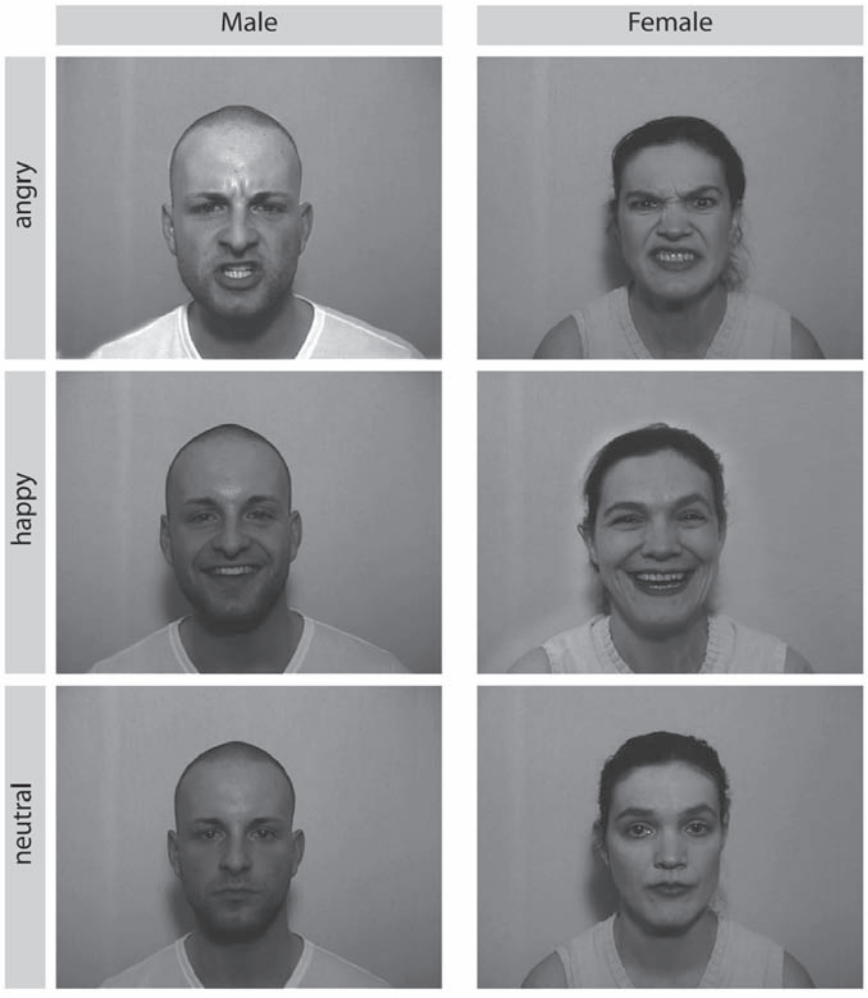

Figure 1. Examples of the photographs that were used as visual stimuli.

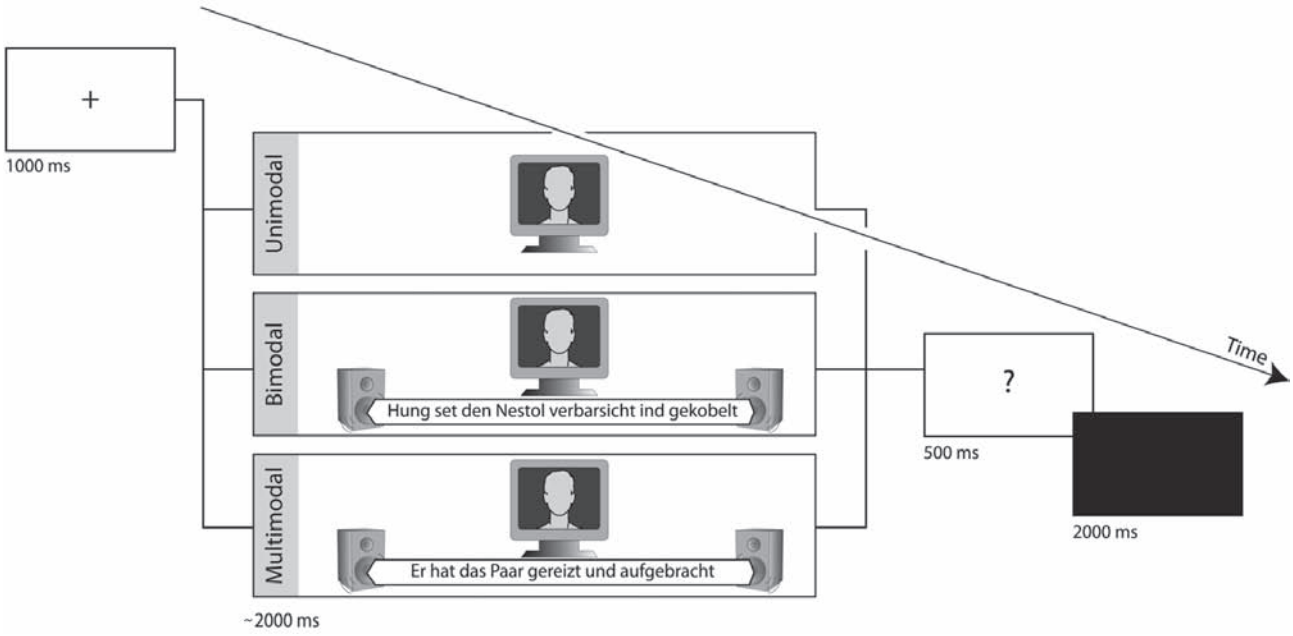

Figure 2. Visualization of a trial sequence. After a fixation cross (present for 1 s), stimuli were presented for up to 2000 ms before a fixation mark prompted participants for their answers.

presented once, each lexical sentence twice (once in the bimodal condition and once in the multimodal condition and each picture was presented nine times altogether, three times in the unimodal face condition, and three times each in the bimodal face + prosody and multimodal face + prosody + lexic conditions, paired with different but emotional category matching sentences each time.

Procedure

After the preparation for EEG recording, the participants were seated in an electrically shielded chamber at a distance of 115 cm in front of a monitor. The visual stimuli were presented in the center of the monitor, and the auditory stimuli were presented via loudspeakers positioned directly to the left and right side of the monitor. The 360 stimuli were pseudorandomized and presented to the participants split into 6 blocks of 60 stimuli each. After each stimulus, the participant had to decide whether the person he or she saw (or heard in the case of auditory stimuli) was male or female by pressing the left or right button of a three-button response panel. The gender of the person speaking and presented in the picture was always congruent. Half the participants had to press the right button for female and the left button for male, and the order was reversed for the other half of the participants. Before the onset of each stimulus, a fixation cross was presented in the center of the screen for 1000 ms. Following the presentation of each stimulus, a question mark appeared on the screen for 500 ms, prompting the participant to respond. After the response, an interstimulus interval (ISI) of 2000 ms followed, before the next stimulus was presented (see Figure 2 for a visualization of a trial sequence). After each block, the participant paused for a self-determined duration before proceeding (approx. 2–3 min).

ERP Recording and Data Analysis

ERP Recording

The electroencephalogram (EEG) was recorded from 64 Ag-AgCl electrodes mounted on a custom-made cap (Electro-Cap International, Eaton, OH, USA) according to the modified expanded 10–20 system (Nomenclature of the American Electroencephalographic Society, 1991). Signals were recorded continuously with a bandpass between DC and 70 Hz and digitized at a sampling rate of 250 Hz (Brain Amp amplifier, Brain Products, Gilching, Germany). The reference electrode was placed on the left mastoid. Bipolar horizontal and vertical EOGs were recorded for artifact rejection purposes. Electrode resistance was kept below 10 KΩ. Data was rereferenced offline to linked mastoids. The data was inspected visually in order to exclude trials containing extreme artefacts and drifts, and all trials containing EOG-artifacts above 30.00 μV were rejected automatically. In total, approximately 25% of the data were rejected. Because of very strong artifacts, one participant had to be excluded from the analysis. For further analysis, only trials in which the participants responded correctly were used. For all conditions the trials were averaged over a time range of 200 ms before stimulus onset to 2000 ms after stimulus onset.

Data Analysis

The main aim of the identity identification task was to ensure that participants attended actively to the stimulus material; therefore, no prior hypotheses were made with regard to the effect that gender selection may have on multimodal processing. Still, accuracy rates and reaction times (RTs; correct responses only) were analyzed by means of a repeated measurements ANOVA using the within-subject factors condition (COND: unimodal, bimodal, multimodal), and emotion (EMO: happy, angry, and neutral).

For the ERP analysis, the electrodes were grouped according to left and right hemisphere (left: AF7, AF3, F7, F5, F3, FT7, FC5, FC3, TP7, CP5, CP3, P7, P5, P3, PO7, PO3; right: AF8, AF4, F8, F6, F4, FT8, FC6, FC4, TP8, CP6, CP4, P8, P6, P4, PO8, PO4), and anterior and posterior region (anterior: AF7, AF3, F7, F5, F3, FT7, FC5, FC3, AF8, AF4, F8, F6, F4, FT8, FC6, FC4; posterior: TP7, CP5, CP3, P7, P5, P3, PO7, PO3, TP8, CP6, CP4, P8, P6, P4, PO8, PO4). Based on visual inspection two time windows were selected for analysis: 120 to 200 ms poststimulus (P200 component), and 200 to 320 ms poststimulus (N300 component). The mean amplitude was computed for both time windows. In addition, a peak-to-peak analysis was conducted (see Steinhauer, Alter, & Friederici 1999). To this end, the points in time for the peak analysis of the anteriorly distributed P200 and N300 components were determined at anterior electrode sites (for a list of electrodes see above), and then the time-windows encompassing these time points were chosen. To investigate the peak-to-peak latency for the P200/N300 components, that is, the difference in peak latency for the two components, the following time windows were chosen: 140 ms–170 ms and 170 ms–260 ms (see Steinhauer et al., 1999, for similar approach).

Comparable to the behavioral analysis, a repeated measurements ANOVA was computed for all the conditions using the within-subject factors hemisphere (HEMI: left, right), region (REG: anterior, posterior), condition (COND: unimodal, bimodal, multimodal[1]), and emotion (EMO: happy, angry, and neutral). For the peak-to-peak analysis, the factor REG was not included since only frontal electrode sites were chosen to be analyzed. Customized tests of hypotheses (posthoc tests) were carried out using a modified Bonferroni procedure correction for multiple comparisons when appropriate (see Keppel, 1991). Therefore, in these cases, the α level was set at $p < .033$ and not at $p < .05$. Also, comparisons with more than one degree of freedom in the numerator were corrected for nonsphericity using the Greenhouse-Geisser correction (Greenhouse & Geisser, 1959). The graphs displayed were filtered with a 7 Hz lowpass filter. In order to estimate the effect size, Ω^2 was calculated. Ω^2 is the coefficient of determination which measures the part of variance in the dependent variable that can be explained by the independent variable.

Statistical Analyses

Only significant interactions involving the critical factors EMO and/or COND are reported in stepdown analyses.

Behavioral Analysis

The main analysis for RTs revealed a significant effect of EMO, $F(2, 44) = 5.55$, $p < .01$, $\Omega^2 = 0.17$, suggesting different response times for the different emotions during the identity identification process. Indeed, posthoc contrasts revealed faster gender recognition for both happy, $F(2, 44) = 5.25$, $p < .05$, $\Omega^2 = 0.16$, and angry, $F(2, 44) = 11.30$, $p < .01$, $\Omega^2 = 0.31$, stimuli in contrast to neutral stimuli. Gender selection was 12 ms faster for angry stimuli and 7 ms faster for happy stimuli in contrast to neutral stimuli. No other effects were significant.

The analysis for accuracy rates revealed no significant results, suggesting that gender recognition is not influenced by the amount of channels (one, two, or three) in which the information is presented and is also not influenced by the valence of the stimuli (all p values > .37).

P200 Mean Amplitudes

In the early time window a significant effect of EMO, $F(2, 44) = 6.28$, $p < .01$, $\Omega^2 = 0.133$, was found, revealing a more pronounced amplitude rise for angry stimuli than neutral stimuli, $F(1, 22) = 9.71$, $p < .01$, $\Omega^2 = 0.159$. In addition, EMO interacted with REG, $F(2, 44) = 14.27$, $p < .0001$, $\Omega^2 = 0.161$, revealing that at anterior electrode sites, the EMO effect was most pronounced, $F(2, 44) = 11.93$, $p < .0001$, $\Omega^2 = 0.322$. At these sites, both angry, $F(1, 22) = 20.71$, $p < .001$, $\Omega^2 = 0.300$, and happy, $F(1, 22) = 7.18$, $p < .05$, $\Omega^2 = 0.118$, stimuli elicited stronger amplitudes than neutral stimuli.

A highly significant effect of COND, $F(2, 44) = 13.38$, $p < .0001$, $\Omega^2 = 0.264$, was also observed. Posthoc contrasts revealed a significant difference between unimodal and bimodal stimuli, $F(1, 22) = 5.01$, $p = .0356$, $\Omega^2 = 0.080$, bimodal and multimodal stimuli, $F(1, 22) = 14.23$, $p < .001$, $\Omega^2 = 0.223$, as well as between unimodal and multimodal stimuli, $F(1, 22) = 20.84$, $p < .0001$, $\Omega^2 = 0.431$. Specifically, the unimodal stimuli elicited the largest amplitude, followed by the bimodal and then multimodal stimuli.

Moreover, a marginal significant interaction between EMO and COND was found, $F(4,88) = 2.57$, $p = .06$, $\Omega^2 = 0.023$. Posthoc tests revealed that for angry stimuli,

[1] Note that we chose to use pictures combined with pseudo-sentences in the analysis for the bimodal condition, but that similar results were obtained when running analyses with the control bimodal condition, that is with lexical sentences.

unimodal and bimodal stimuli, $F(1, 22) = 8.91$, $p < .01$, $\Omega^2 = 0.147$, elicited significantly different amplitudes, with unimodal stimuli eliciting a larger amplitude than bimodal stimuli. Also, unimodal stimuli differed from multimodal stimuli, $F(1, 22) = 13.80$, $p < .01$, $\Omega^2 = 0.218$. Again unimodal stimuli elicited a larger amplitude than multimodal stimuli. For happy stimuli, unimodal stimuli also differed from multimodal stimuli, $F(1, 22) = 7.13$, $p < .03$, $\Omega^2 = 0.118$, in that one-channel stimuli elicited a larger P200 amplitude than three-channel stimuli. Neutral stimuli were processed similar to emotional stimuli. Here again, unimodal, $F(1, 22) = 11.41$, $p < .01$, $\Omega^2 = 0.185$, and bimodal, $F(1, 22) = 14.39$, $p < .01$, $\Omega^2 = 0.225$, stimuli elicited stronger amplitudes than multimodal stimuli.

There was also a COND × HEMI interaction, $F(2, 44) = 5.84$, $p < .01$, $\Omega^2 = 0.066$, suggesting lateralization differences for the COND effect. For left hemisphere electrode sites, $F(2, 44) = 11.41$, $p < .01$, $\Omega^2 = 0.312$, both unimodal, $F(1, 22) = 26.42$, $p < .0001$, $\Omega^2 = 0.356$, and bimodal stimuli, $F(1, 22) = 7.63$, $p < .05$, $\Omega^2 = 0.130$, differed from multimodal stimuli. Also, unimodal stimuli differed from bimodal stimuli, $F(1, 22) = 10.76$, $p < .01$, $\Omega^2 = 0.175$. For right hemisphere electrode sites, $F(2, 44) = 9.41$, $p < .01$, $\Omega^2 = 0.268$, both unimodal, $F(1, 22) = 21.67$, $p < .0001$, $\Omega^2 = 0.310$, and bimodal, $F(1, 22) = 14.12$, $p < .01$, $\Omega^2 = 0.221$, stimuli elicited stronger P200 amplitudes than multimodal stimuli.

Last, there was also a significant interaction between COND and REG, $F(2, 44) = 4.81$, $p < .05$, $\Omega^2 = 0.052$. At both anterior, $F(2, 44) = 4.65$, $p < .05$, $\Omega^2 = 0.137$, and posterior electrode sites, $F(2, 44) = 5.81$, $p < .05$, $\Omega^2 = 0.173$, we observed a significant COND effect. At anterior sites, both unimodal, $F(1, 22) = 5.95$, $p < .03$, $\Omega^2 = 0.097$, and bimodal stimuli, $F(1, 22) = 9.82$, $p < .01$, $\Omega^2 = 0.160$, differed from multimodal stimuli. Again, multimodal stimuli elicited the weakest amplitudes. At posterior electrode sites unimodal, $F(1, 22) = 50.41$, $p < .0001$, $\Omega^2 = 0.518$, and bimodal stimuli, $F(1, 22) = 11.94$, $p < .01$, $\Omega^2 = 0.238$, differed from multimodal stimuli. Also, unimodal stimuli elicited a stronger amplitude than bimodal stimuli, $F(1, 22) = 15.05$, $p < .001$, $\Omega^2 = 0.234$.

In summary, we report early emotional and neutral differentiation, which is specifically pronounced at anterior electrode sites. In addition, we observed a processing advantage for multimodal over bimodal and unimodal stimuli. The results suggest that this condition effect varies as a function of the distribution for specific contrasts (that is unimodal vs. bimodal and multimodal stimuli as well as bimodal vs. multimodal stimuli). These distributional differences seem unaffected by the valence of the stimuli. All effects are displayed in Figure 3, Figure 4, and Figure 5.

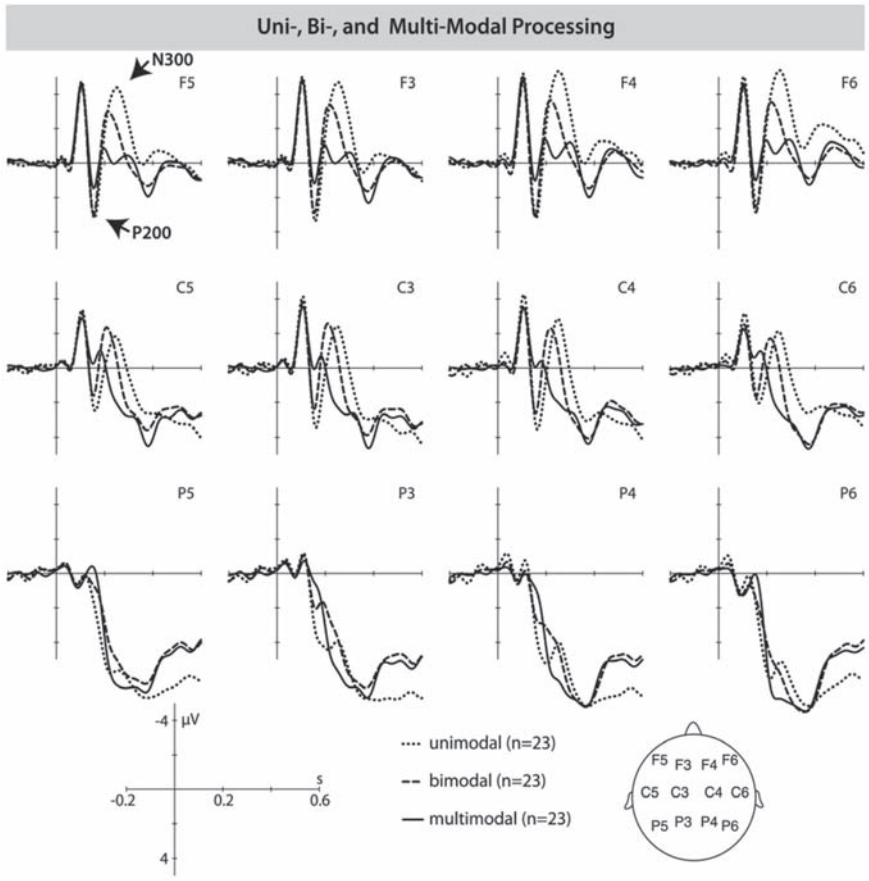

Figure 3. The image displays the facilitation effects as reflected in the P200 and N300 amplitudes at selected electrode sites.

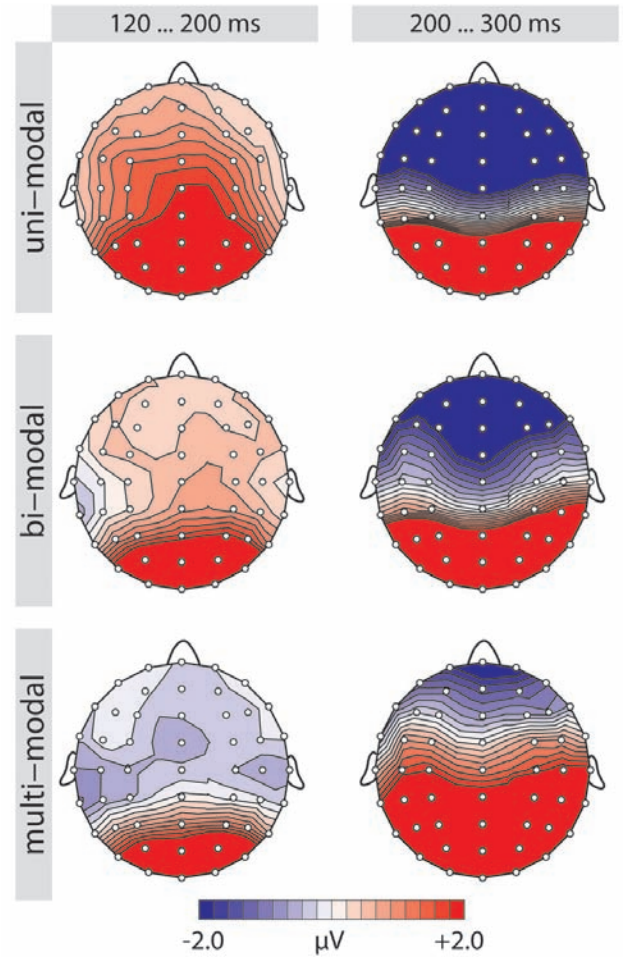

Figure 4. The scalp maps show the topographic distribution of the facilitation effect.

N300 Mean Amplitudes

In the time window for the N300, a highly significant effect of COND, $F(2, 44) = 34.34$, $p < .0001$, $\Omega^2 = 0.491$, was found. Posthoc contrasts revealed a significant difference between unimodal and bimodal stimuli, $F(1, 22) = 11.82$, $p < .01$, $\Omega^2 = 0.191$, bimodal and multimodal stimuli, $F(1, 22) = 29.50$, $p < .0001$, $\Omega^2 = 0.383$, as well as between unimodal and multimodal stimuli, $F(1, 22) = 59.94$, $p < .0001$, $\Omega^2 = 0.561$. As was the case for the P200 amplitudes, unimodal stimuli elicited the strongest N300 amplitude.

In addition, a significant interaction between EMO and COND was found, $F(4,88) = 2.78$, $p < .05$, $\Omega^2 = 0.027$. Posthoc tests revealed that for angry stimuli, unimodal, $F(1, 22) = 38.37$, $p < .0001$, $\Omega^2 = 0.448$, and bimodal, $F(1, 22) = 13.94$, $p < .01$, $\Omega^2 = 0.220$, stimuli differed from multimodal stimuli, with multimodal stimuli eliciting the weakest N300. For happy stimuli, unimodal stimuli also differed from bimodal stimuli, $F(1, 22) = 17.01$, $p < .001$, $\Omega^2 = 0.258$, and multimodal stimuli, $F(1, 22) = 41.21$, $p < .0001$, $\Omega^2 = 0.466$, in that unimodal channel stimuli elicited stronger P200 amplitudes than three-channel stimuli. For neutral stimuli one-modal, $F(1, 22) = 25.00$, $p < .0001$, $\Omega^2 = 0.343$, and bimodal, $F(1, 22) = 16.16$, $p < .001$, $\Omega^2 = 0.248$, stimuli elicited stronger amplitudes than multimodal stimuli.

Moreover, there was a COND × HEMI interaction, $F(2, 44) = 10.29$, $p < .0001$, $\Omega^2 = 0.119$, suggesting lateralization differences for the COND effect. For left hemisphere electrode sites, $F(2, 44) = 27.01$, $p < .01$, $\Omega^2 = 0.430$, both one-modal, $F(1, 22) = 43.31$, $p < .0001$, $\Omega^2 = 0.480$, and bimodal stimuli, $F(1, 22) = 33.16$, $p < .0001$, $\Omega^2 = 0.411$, differed from multimodal stimuli. Also, unimodal stimuli differed from bimodal stimuli,

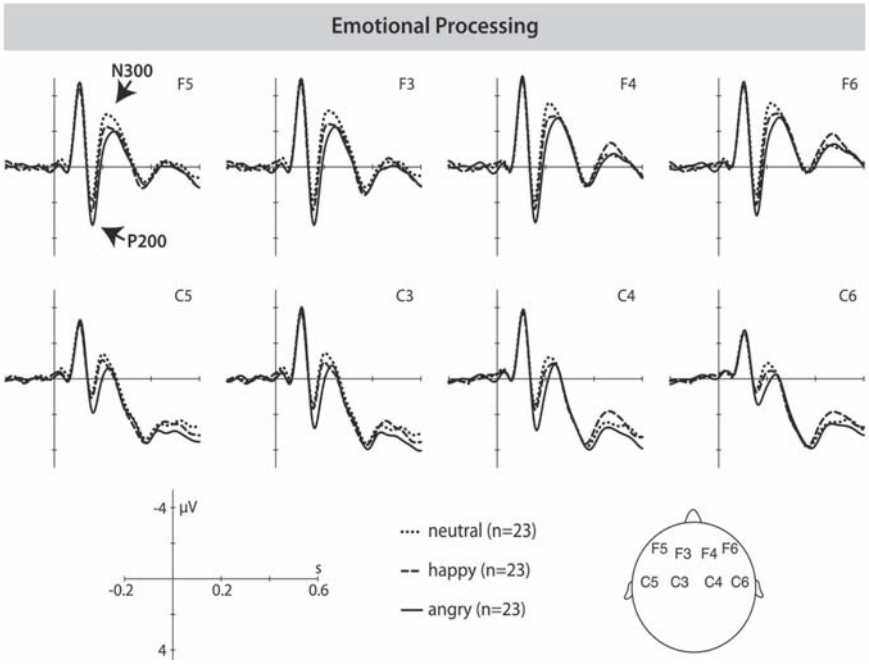

Figure 5. The image displays the emotion effect as reflected in the P200 and N300 amplitudes at selected electrode sites.

$F(1, 22) = 5.96$, $p < .05$, $\Omega^2 = 0.097$. Comparable to the effects observed for the P200 component, the N300 amplitude was weakest for multimodal stimuli. For right hemisphere electrode sites, $F(2, 44) = 36.55$, $p < .0001$, $\Omega^2 = 0.507$, the same picture emerged. Both unimodal, $F(1, 22) = 67.35$, $p < .0001$, $\Omega^2 = 0.590$, and bimodal, $F(1, 22) = 21.36$, $p < .0001$, $\Omega^2 = 0.307$, stimuli elicited stronger N300 amplitudes than multimodal stimuli. Also, bimodal stimuli elicited weaker amplitudes than unimodal stimuli, $F(1, 22) = 17.39$, $p < .001$, $\Omega^2 = 0.262$.

Finally, there was also a significant interaction between COND and REG, $F(2, 44) = 5.81$, $p < .05$, $\Omega^2 = 0.065$. At both anterior, $F(2, 44) = 34.78$, $p < .05$, $\Omega^2 = 0.329$, and posterior electrode sites, $F(2, 44) = 17.25$, $p < .0001$, $\Omega^2 = 0.190$, we observed a significant COND effect. At anterior sites, both one-modal, $F(1, 22) = 50.91$, $p < .0001$, $\Omega^2 = 0.520$, and bimodal stimuli, $F(1, 22) = 15.13$, $p < .001$, $\Omega^2 = 0.235$, differed from multimodal stimuli. Also, unimodal differed from bimodal stimuli processing, $F(1, 22) = 26.26$, $p < .0001$, $\Omega^2 = 0.354$. Again, multimodal stimuli always elicited the weakest amplitudes. At posterior electrode sites one-modal, $F(1, 22) = 30.18$, $p < .0001$, $\Omega^2 = 0.433$, and bimodal stimuli, $F(1, 22) = 32.73$, $p < .0001$, $\Omega^2 = 0.408$, differed from multimodal stimuli, again less channels of information elicited stronger amplitudes than more channels of information.

Finally, there was also an interaction between EMO and REG, $F(2, 44) = 6.67$, $p < .01$, but none of the step-down analyses reached significance.

In summary, the N300 component differentiates between multimodal, bimodal, and unimodal stimuli processing. Unimodal stimuli always elicit stronger amplitudes than bimodal or multimodal stimuli. The effect seems to be slightly more robust at anterior electrode sites as the N300 is typically anteriorly distributed. Again, valence effects do not seem to play a major role in this distinction process (see Figures 3–5 for graphical display of effects).

P200 Peak to N300 Peak Latencies

In this analysis, only one highly significant effect of COND, $F(2, 44) = 29.26$, $p < .0001$, $\Omega^2 = 0.450$, was found, suggesting different peak-to-peak latencies for the different stimulus modalities. Posthoc contrasts revealed a significant latency difference between unimodal and bimodal stimuli, $F(1, 22) = 47.81$, $p < .0001$, $\Omega^2 = 0.50$, bimodal and multimodal stimuli, $F(1, 22) = 6.63$, $p < .03$, $\Omega^2 = 0.109$, as well as between unimodal and multimodal stimuli, $F(1, 22) = 38.78$, $p < .0001$, $\Omega^2 = 0.450$. Peak-to-peak latency was longest for unimodal stimuli (82 ms), followed by bimodal (66 ms) and multimodal (57 ms) stimuli, suggesting facilitation effects for multimodal presentation independent of valence.

Discussion

The present study investigated the nature of multimodal communication processing with unimodal (facial expression), bimodal (facial expression and prosody), and multimodal (facial expression, prosody, and lexical) levels of information in a nonconflict paradigm. We hypothesized that step-wise addition of information should facilitate processing and in turn modulate the amplitude size and latency of two early ERP components, namely, the P200 and N300. These two components were of particular interest as they are commonly associated with the detection of stimulus salience and content evaluation, respectively. We report a gradual and systematic decrease of the P200 and N300 amplitudes from unimodal to multimodal information processing. Similarly, we observe the same pattern in a peak-to-peak latency analysis of the P200/N300 complex. In line with previous studies (Pourtois et al., 2000, 2002; de Gelder et al., 1999; Lebib et al., 2004; Wu & Coulson, 2007), the current results confirm facilitation of multiple channels processing in early ERP components, but extend these findings to nonconflict paradigms by adding emotional and neutral coherent information in a step-wise manner.

Our results go hand in hand with several findings on audiovisual integration in the literature. Stekelenburg and Vroomen (2007) report a decrease in P200 amplitude and a latency reduction for audiovisual integration compared to unimodal processing. Van Wassenhove and colleagues (2005) describe an amplitude reduction on the N1/P2 complex for audio-visual speech processing compared to audio-only processing. Pourtois and colleagues (2002) state a latency effect on the P2b component, a positive deflection directly following the P200. Audiovisual presentation of congruent emotional facial expressions and words intoned in a happy or fearful voice elicited a shorter P2b latency compared to incongruent audiovisual presentation (Pourtois et al., 2002). However, our data extend these findings in several ways. We clearly show that the given number of channels directly modulates facilitation in form of amplitude and latency reduction with the three-channel presentation resulting in highest facilitation. Furthermore, we suggest that the respective component modulations reflect reduced and speeded processing effort. In short, there seems to be a direct link between amplitude size and number of channels present, a finding previously observed in other paradigms (Kotz & Paulmann, 2007; Paulmann & Kotz, 2008b). The results raise the question why multimodal (emotional) processing is more "successful," that is, quicker and of less effort than unimodal processing – which is already quick and accurate (see Borod et al., 2000; Balconi & Pozzoli, 2003; Batty & Taylor, 2003; Ekman, 1992; Elfenbein & Ambady, 2002; Paulmann & Kotz, 2008a; Paulmann et al., 2008).

There are several potential answers to this question. First, congruent information presented via multiple channels may be less error-prone than information presented via

just one modality. Consider a situation in which vision is hampered. For instance, during a foggy night we may have to rely on vision as well as auditory perception to judge if someone is in danger and screaming for help or just play-fighting with a partner. To rely on both visual and auditory information results in a more accurate interpretation of a given situation. Although sometimes the (overlapping) information gathered from different sources may be redundant, it can also help to reduce a processing effort. If the same information is perceived via multiple channels, it is more likely to be correct than when relying on only one modality. Hence, less information from each channel is necessary to detect a stimulus and elicit a response, reflected in shorter latencies. Similarly, less processing is required because less information is sufficient for the detection of, say, emotionality, leading to a decrease in amplitude.

There are claims that facilitation can be related to the early integration of different information or communication channels, which would imply that redundant information is used strategically rather than processed side by side. ERP evidence indeed suggests that integration already takes place during early stages of emotional and neutral information processing (within the first 200 ms), arguably even before each individual channel has been adequately processed (de Gelder et al., 1999). Our current investigation cannot say much about the nature of channel integration, as it is still possible that there are distinct processing mechanisms for unimodal, bimodal, and multimodal information processing which lead to facilitation of the latter. Future studies are called for to directly investigate the nature of stepwise channel *integration*. For now, based on previous evidence (e.g., de Gelder et al., 1999) we can presume that the facilitation observed here is triggered by early channel integration, though we cannot completely rule out the existence of a distinct processing mechanism for unimodal, bimodal, or multimodal stimuli. Either way, the current data provide evidence that redundancy and overlapping information lead to a processing advantage reflected in shorter latencies and decreased amplitudes.

It may also be meaningful to take a closer look at the processes underlying human communication processing. For instance, it has been suggested that populations of neurons in channel-specific (e.g., auditory, visual) input systems are activated after an encounter with an (emotional) stimulus. Although these systems can each act individually, they are also highly conjoined, thus licensing for information fusion (e.g., Niedenthal, 2007). It is assumed that sensory input invokes so-called "nodes," or network units, which by means of spreading activation distribute activity to other network parts. One may argue that each information channel (visual, lexical, prosodic) activates both channel-specific knowledge and channel-unspecific, or more comprehensive, knowledge by spreading activation. The more knowledge is activated, the more optimized early encoding can be, which in turn may trigger privileged treatment, such as quicker processing routes, leading to facilitative perception of a multimodal stimulus in comparison to a stimulus with less information.

Interestingly, the current investigation reports facilitation effects for multimodal over bimodal or unimodal stimuli not only for emotional, but also for neutral stimuli. This suggests that multimodal processing is always advantageous irrespective of the stimulus quality. In fact, given the reported evidence on more general audio-visual speech processing, it is not too surprising that neutral multimodal stimulus processing shows a similar facilitation effect as emotional multimodal processing (see Campanella & Belin, 2007; van Wassenhove et al., 2005). Yet, previous evidence continuously reports prioritized (e.g., rapid analysis, attentional bias) processing of emotional information (e.g., Anderson & Phelps, 2001; Paulmann & Kotz, 2008a; Eimer & Holmes, 2002; Fox et al., 2000). Based on this evidence we would have expected time-course differences between emotional and neutral information processing with different numbers of channels. However, the current results suggest that the "emotional advantage" does not apply to all multimodal processing situations. In fact, the emotional quality of a stimulus does not enhance multimodal processing. This is not to say that the stimulus characteristics (emotional vs. neutral) were not adequately processed in the current study as we report a main effect for emotion in the P200 component (e.g., Ashley, Vuilleumier, & Swick, 2004; Paulmann & Kotz, 2008a) in an emotion evaluation-unrelated task. Furthermore, the behavioral results reveal faster response times to emotional than to neutral stimuli. Therefore, we suggest that the missing differential effect of emotional and neutral multimodal processing could be task-related. In particular, one might argue that any emotional effect would be most pronounced when the situation calls for emotional behavior (e.g., running for help in the example situation described above). This may not be the case when attention is not directly focused on an emotion encoded in a stimulus. This suggestion is supported by our own data, as we report differential condition effects for angry, happy, and neutral stimuli. However, the P200 in response to happy unimodal stimuli differed only from happy multimodal stimuli, but not from happy bimodal stimuli – nor were happy bimodal stimuli different from multimodal happy stimuli. Obviously, this claim is speculative, and future studies will have to explore the *emotional* facilitative effect in tasks that require the thorough evaluation of the emotionality of a stimulus.

On a critical level, it needs to be acknowledged that the amplitude reductions observed here could also reflect interference or competition effects rather than facilitation as proposed. In particular, it can be argued that 200–300 ms do not provide enough *content* information in the multimodal condition to cause a facilitation effect at this point of time. Two observations make this interpretation less likely though. First of all, we see a reduction in amplitude for both bimodal and multimodal stimuli. While it is true that emotional content is not yet avail-

able for the multimodal stimuli 200–300 ms after sentence onset, previous research clearly showed that emotionally relevant details about a stimulus, such as valence (Paulmann & Kotz, 2008a), arousal (Paulmann & Kotz, 2006), and emotional knowledge in the mental store (Paulmann & Pell, 2009), can be inferred from vocal expressions within the first 200 ms of stimulus presentation. Based on this evidence, reduced amplitudes for bimodal stimuli can be interpreted as reflecting facilitation of information processing because both channels provide enough (emotional) category information at this point in time. Thus, since amplitude reductions are interpreted to reflect a facilitation effect for bimodal stimuli in contrast to unimodal stimuli, it would be incoherent to assume that amplitude reductions for multimodal in contrast to bimodal and unimodal stimuli would reflect competition among the channels. Moreover, given that neutral stimuli elicit the same kind of pattern as emotional stimuli, it seems reasonable to argue that it is not of primary relevance at this point in time whether the channel provides *enough* information (e.g., emotional content); rather, it seems to be important that the channel is *at all* present and providing *any kind* of information (e.g., lexical status of the stimulus). That is, although the emotional content may not yet be readily available in the latter stimuli, listeners have clearly processed the lexical channel information at this point in time (see Paulmann, 2006). Second and more critical, we additionally report a shorter latency onset for multimodal stimuli in contrast to bimodal and unimodal stimuli. This speaks in favor of a facilitation interpretation since – to our current knowledge – there is no evidence that either competition or interference effects would lead to *quicker* online processing of information.

Finally, it is worth mentioning that the two ERP components related to the facilitation effect (P200 and N300) may in fact reflect two different processing stages. As highlighted in the introduction, the P200 component is related primarily to the emotional salience detection of a stimulus (e.g., Paulmann 2006; Paulmann & Kotz, 2008a; Paulmann et al., 2008). In contrast, the N300 component has been related to a more in-depth evaluation of a stimulus that may include semantic and conceptual processing. As we observe the facilitation effect in both components, we suggest that facilitation effects can be found in two different processing stages. First, the early integration of different channels that may lead to faster and more accurate processing is primarily triggered through physical attributes of the stimuli. In a second step, processing requires a more thorough analysis of the stimuli, which may include conceptual knowledge processing or processing of category-specific knowledge/concept representation[2]. As previously stated, this conceptual activation may simply be stronger for stimuli that elicit this activation from three different sources than for single-channel stimuli with less activation strength. In turn, multimodal stimuli are processed quicker and more efficiently (shorter latencies) as well as with less effort (reduced amplitudes).

Conclusion

The processing of multimodal information tends to be quicker and easier than processing the information from less channels. Presumably, information from different channels are encoded and integrated at an early point in time during processing as reflected in the P200 and N300 components. The combined perception (that is, information from different sources) may help to form a more unified, holistic percept of a stimulus and may be advantageous in real-life situations. We propose that such facilitation may be explained by an increased activation of network units that allow for preferential processing as reflected in the observed facilitation effect. Moreover, we suggest that this facilitation effect is not limited to emotional stimuli (since we find a similar effect for neutral stimuli). Nevertheless, under certain circumstances (e.g., attentional focus on the emotional characteristics of a stimulus) emotional multimodal processing may be even more advantageous than multimodal processing of neutral events. Future studies will need to follow to fully specify the nature of multimodal emotional processing.

Acknowledgments

Both first authors contributed equally to the paper. The authors would like to thank Kerstin Flake for help with graphical presentation, Stephen Hopkins for help with statistical analyses, and Ina Koch for help with data acquisition. The support of the German Academic Exchange Service (DAAD) to the first author is gratefully acknowledged. This study was funded by the German Research Foundation (DFG FOR-499 to S.A.K.).

References

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology, 12*, 69–177.

Anderson, A.K., & Phelps, E.A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature, 411,* 305–309.

American Electroencephalographic Society. (1991). Guidelines

[2] Note that recent models of emotional processing (e.g., Adolphs, 2002; Phillips, Drevets, Rauch, & Lane, 2003; Schirmer & Kotz, 2006) go hand in hand with the assumption of two distinct processing stages. In particular, recent models propose an early "low-level" analysis (feature extraction) followed by a more in depth evaluation of a stimulus.

for standard electrode position nomenclature. *Journal of Clinical Neurophysiology, 8,* 200–202.

Ashley, V., Vuilleumier, P., & Swick, D. (2004). Time course and specificity of event related potentials to emotional expressions. *Neuroreport, 15,* 211–216.

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

Batty, M., & Taylor, M.J. (2003). Early processing of the six basic facial emotional expressions. *Brain Research. Cognitive Brain Research, 17,* 613–620.

Borod, J.C., Pick, L.H., Hall, S., Sliwinski, M., Madigan, N., Obler, L.K. et al. (2000). Relationships among facial, prosodic, and lexical channels of emotional perceptual processing. *Cognition and Emotion, 14,* 193–211.

Bostanov, V., & Kotchoubey, B. (2004). Recognition of affective prosody: Continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology, 41,* 259–268.

Calvert, G.A., Brammer, M.J., & Iversen, S.D. (1998). Crossmodal identification. *Trends in Cognitive Science, 2,* 247–253.

Campanella, S., Gaspard, C., Bruyer, R., Debatisse, D., Crommelinck, M., & Guerit, J.M. (2002). Discrimination of emotional facial expressions in a visual oddball task: An ERP study. *Biological Psychiatry, 59*, 171–186.

Campbell, R. (2008). The processing of audio-visual speech: Empirical and neural bases. *Philosophical Transactions of the Royal Society London B Biological Science, 363,* 1001–1010.

Carretié, L., & Iglesias, J. (1995). An ERP study on the specificity of facial expression processing. *International Journal of Psychophysiology, 19*, 183–192.

Carretié, L., Iglesias, J., & García, T. (1997). A study on the emotional-processing of visual stimuli through event-related potentials. *Brain and Cognition, 34*, 207–217.

Carretié, L., Iglesias, J., García, T., & Ballesteros, M. (1997). N300, P300 and the emotional processing of visual stimuli. *Electroencephalography and Clinical Neurophysiology, 103*, 298–303.

Carretié, L., Mercado, F., Tapia, M., & Hinojosa, J.A. (2001). Emotion, attention, and the "negativity bias," studied through event-related potentials. *International Journal of Psychophysiology, 41*(1), 75–85.

de Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Science, 7,* 460–467.

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

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

de Gelder, B., Vrommen, J., & Pourtois, G. (1999). Seeing cries and hearing smiles: Crossmodal perception of emotional expressions. In G. Aschersleben, T. Bachmann, & J. Müsseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events*. *Advances in Psychology, 129* (pp. 425–438). Amsterdam: North-Holland/Elsevier.

Eimer, M., & Holmes, A. (2002). An ERP study on the time course of emotional face processing. *Neuroreport, 13*, 427–431.

Eimer, M., Holmes, A., & McGlone, F.P. (2003). The role of spatial attention in the processing of facial expression: An ERP study of rapid brain responses to six basic emotions. *Cognitive, Affective, and Behavioral Neuroscience, 3*, 97–110.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6,* 169–200.

Elfenbein, H.A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128,* 203–235.

Fox, E., Lester, V., Russo, R., Bowles, R.J., Pichler, A., & Dutton, K. (2000). Facial expressions of emotion: Are angry faces detected more efficiently? *Cognition and Emotion, 14,* 61–92.

Federmeier, K.D., & Kutas, M. (2002). Picture the difference: Electrophysiological investigations of picture processing in the two cerebral hemispheres. *Neuropsychologia, 40*, 730–747.

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

Greenhouse, S., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95–112.

Hamm, J.P., Johnson, B.W., & Kirk, I.J. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology, 113*, 1339–1350.

Keppel, G., 1991. *Design and analysis: A researcher's handbook.* Englewood Cliffs, NJ: Prentice Hall.

Kotz, S.A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research, 1151,* 107–118.

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.

Lang, P.J. (1993). The network model of emotion: Motivational concerns. In R.S. Wyer & T.K. Scrull (Eds.), *Advances in social cognition* (pp. 109–133). Hillsdale, NJ: Erlbaum.

Lang, P.J., & Davis, M. (2006). Emotion, motivation, and the brain: Reflex foundations in animal and human research. *Progress in Brain Research, 156,* 3–29.

Lebib, R., Papo, D., Douiri, A., de Bode, S., Gillon Dowens, M., & Baudonniere, P.M. (2004). Modulations of late event-related brain potentials in human by dynamic audiovisual speech stimuli. *Neuroscience Letters, 372*(1–2), 74–79.

LeDoux, J.E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience, 23,* 155–184.

Luck, S.J., & Hillyard, S.A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology, 31,* 291–308.

Miller, L.M., & D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *The Journal of Neuroscience, 25,* 5884–5893.

Niedenthal, P.M. (2007). Embodying Emotion. *Science, 316,* 1002–1005.

Pantev, C., Elbert, T., Ross, B., Eulitz, C., & Terhardt, E. (1996). Binaural fusion and the representation of virtual pitch in the human auditory cortex. *Hearing Research, 100,* 164–170.

Paulmann, S. (2006). *Electrophysiological evidence on the processing of emotional prosody: Insights from healthy and patient populations (Max Planck Series in Human Cognitive and Brain Sciences*, Vol. 71). Leipzig, Germany: Max Planck Institute for Human Cognitive and Brain Sciences.

Paulmann, S., & Kotz, S.A. (2006, August). *Valence, arousal, and*

task effects on the P200 in emotional prosody processing. Poster presented at the 12th Annual Conference on Architectures and Mechanisms for Language Processing 2006 (AMLaP), Nijmegen, The Netherlands.

Paulmann, S., & Kotz, S.A. (2008a). Early emotional prosody perception based on different speaker voices. *Neuroreport, 19,* 209–213.

Paulmann, S., & Kotz, S.A. (2008b). An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical-sentence context. *Brain and Language, 105*, 59–69.

Paulmann, S., & Pell, M.D. (2009, April). *Rapid processing of vocal emotion expressions as revealed by ERPs.* Poster presented at the Evoked Potentials International Conference XV, Indiana University, IN, USA.

Paulmann, S., Pell, M.D., & Kotz, S.A. (2008). How aging affects the recognition of emotional speech. *Brain and Language, 104*, 262–269.

Paulmann, S., Schmidt, P., Pell, M.D., & Kotz, S.A. (2008, May). *Rapid processing of emotional and voice information as evidenced by ERPs.* In Speech Prosody 2008, Fourth Conference on Speech Prosody, Campinas, Brazil.

Pell, M.D., Monetta, L., Paulmann, S., & Kotz, S.A. (in press). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior.*

Phillips, M.L., Drevets, W.C., Rauch, S.L., & Lane, R. (2003). Neurobiology of emotion perception I: The neural basis of normal emotion perception. *Biological Psychiatry, 54,* 504–514.

Picton, T.W., Woods, D.L., Barribeau-Braun, J., & Healy, T.M.G. (1977). Evoked potential audiometry. *Journal of Otolaryngology, 6,* 90–119.

Pizzagalli, D., Regard, M., Lehmann, D. (1999). Rapid emotional face processing in the human right and left brain hemispheres: An ERP study. *Neuroreport, 10*, 2691–2698.

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.

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

Schirmer, A., & Kotz, S.A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotion processing. *Trends in Cognitive Science, 10,* 24–30.

Schutter, D.J.L.G., de Haan, E.H.F., & van Honk, J. (2004). Functionally dissociated aspects in anterior and posterior electrocortical processing of facial threat. *International Journal of Psychophysiology, 53*, 29–36.

Stekelenburg, J.J., & Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience, 19*, 1964–1973.

Steinhauer, K., Alter, K., & Friederici, A.D. (1999). Brain responses indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2,* 191–196.

van Wassenhove, V., Grant, K.W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences, 102*, 1181–1186.

Vogel, E.K., & Luck, S.J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology, 37*, 190–203.

Wu, Y.C., & Coulson, S. (2007). How iconic gestures enhance communication: An ERP study. *Brain & Language, 101,* 234–245.

Accepted for publication: May 11, 2009

Silke Paulmann

Department of Psychology
University of Essex
Wivenhoe Park
Colchester, C04 3SQ
UK
Tel. +44 1206 873802
Fax +44 1206 873801
E-mail paulmann@essex.ac.uk