Available online at www.sciencedirect.com
ScienceDirect
Brain and Language 105 (2008) 59–69
Brain and Language
www.elsevier.com/locate/b&l

An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical-sentence context

Silke Paulmann [a,*], Sonja A. Kotz [a,b]

[a] *Max Planck Institute for Human Cognitive and Brain Sciences, Neuropsychology of Language, P.O. Box 500 355, 04303 Leipzig, Germany*
[b] *Day Care Clinic of Cognitive Neurology, University of Leipzig, Germany*

Accepted 12 November 2007
Available online 3 January 2008

Abstract

Previous evidence supports differential event-related brain potential (ERP) responses for emotional prosodic processing and integrative emotional prosodic/semantic processing. While latter process elicits a negativity similar to the well-known N400 component, transitions in emotional prosodic processing elicit a positivity. To further substantiate this evidence, the current investigation utilized lexical-sentences and sentences without lexical content (pseudo-sentences) spoken in six basic emotions by a female and a male speaker. Results indicate that emotional prosodic expectancy violations elicit a right-lateralized positive-going ERP component independent of basic emotional prosodies and speaker voice. In addition, expectancy violations of integrative emotional prosody/semantics elicit a negativity with a whole-head distribution. The current results nicely complement previous evidence, and extend the results by showing the respective effects for a wider range of emotional prosodies independent of lexical content and speaker voice.

Keywords: Emotional prosody; Emotional semantics; ERPs; Integration

1. Introduction

When we speak we can use a variety of emotional intonations to give a specific meaning to a verbal expression. Dependent on the rise and fall pattern of the voice the interpretation of an utterance can change. Thus, emotional speech relies on psychoacoustic parameters such as fundamental frequency (F_0) and intensity, or loudness. Together with speech rate, or duration, and rhythm, these parameters constitute emotional prosody. Banse and Scherer (1996) defined acoustical profiles in vocal emotion expression with the help of several acoustic parameters. One of their main findings was that each emotion seems to have its own acoustic profile. For example, the vocalization of anger reveals a higher F_0 than the vocalization of sadness. Furthermore, intensity measures reveal louder vocalizations of happy than of sad utterances. To properly communicate emotional meaning, emotional prosody needs to be integrated with emotional semantics (Kotz & Paulmann, 2007) next to other non-verbal functions such as mimicry and gestures (e.g., Pourtois, de Gelder, Vroomen, Rossion, & Crommelinck, 2000). Even though such a cognitive process is crucial in every day communication, little is known about the underlying mechanism and the time-course of such an integration process. Therefore, the main aim pursued in the current research was to further substantiate the relative contribution of emotional channels, and the time-course of both emotional prosody and of emotional prosody with emotional semantics.

One general finding in the emotion literature is that emotional stimuli seem to be processed differently from neutral stimuli. One explanation that has been put forward for this processing difference is the evolutionary significance of emotions, i.e., emotional stimuli can lead to prior-

* Corresponding author. Fax: +1 514 398 8123.
E-mail address: Silke.Paulmann@mail.mcgill.ca (S. Paulmann).

0093-934X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandl.2007.11.005

itized processing strategies (see also e.g., Schupp, Junghöfer, Weike, & Hamm, 2004). These prioritized processing strategies may enforce attentional orienting or faster processing of emotional stimuli. Indeed, there is accumulating electrophysiological evidence that emotional differentiation occurs as early as 200 ms after stimulus onset during both *visual* emotional word processing (e.g., Begleiter, Porjesz, & Garozzo, 1979; Schapkin, Gusev, & Kuhl, 2000), and *auditory* emotional sentence processing (e.g., Paulmann & Kotz, in press). More specifically, we have shown that this first differentiation occurs independent of attention, i.e., under implicit processing demands, for both lexical- (Paulmann & Kotz, in press) and non-lexical vocal emotional stimuli (Paulmann, 2006).

However, not only the processing of vocal emotions *per se*, but also the integration of emotional channels seems to occur outside the focus of attention (e.g., Kotz & Paulmann, 2007; Pourtois et al., 2000). For instance, Pourtouis and colleagues investigated the integration of emotional auditory and visual stimuli, and report a larger N100 amplitude for congruent than incongruent visual and auditory emotional stimuli pairs (Pourtois et al., 2000). Moreover, Bostanov and Kotchoubey (2004) studied affective prosodic recognition in a context violation paradigm using emotional exclamations (e.g., "Wow", "Oooh", etc.). They reported a N300 to contextually incongruous exclamation, and interpret this component as an indicator of expectancy violation comparable to the well-known N400. If early negativities reported in the literature (e.g., Bostanov & Kotchoubey, 2004; Kotz & Paulmann, 2007) are indeed a form of the classical N400, this would provide evidence that emotional context may speed up speech processing, leading to an earlier onset of well established event-related brain potential (ERP) components (e.g., Bostanov & Kotchoubey, 2004; Kotz & Paulmann, 2007). Last, in a recent study by Wambacq and Jerger (2004), a standard oddball paradigm was applied to investigate emotional prosodic and lexical-semantic information processing of single words. The authors report a larger N400 amplitude in response to semantic than to prosodic stimuli. In contrast, they found a larger P3a amplitude for prosodic than for semantic stimuli. Interestingly, for stimuli carrying both emotional prosodic and semantic information a reduced N400 and an increased P3a was observed. Arguing along similar lines as Besson and colleagues (Besson, Magne, & Schön, 2002), Wambacq and Jerger (2004) proposed that even irrelevant semantics cannot be ignored. In short, there is accumulating evidence that processing vocal emotions (with or without the integration of further emotional information, such as semantics) is a very rapid and non-attentional process. Moreover, some emotional channels may be more important than others during emotional processing, e.g., a predominant processing of emotional semantic over emotional prosodic information has been reported.

In one of our previous studies (Kotz & Paulmann, 2007), we investigated the integrative time-course of emotional prosody with neutral semantics and of emotional prosody with emotional semantics by means of a cross-splicing procedure. Cross-splicing offers a unique possibility to investigate the temporal dynamics of integrative processes as it allows for temporal synchronization of information. Due to the time-locking of critical deviation points in an on-going speech signal with respect to two experimental conditions, the cross-splicing technique allows to directly compare the respective brain responses of two critical conditions. In particular, this technique allows manipulating expectations specific to emotional prosody or to emotional prosody and emotional semantics (see Section 2 for detailed explanation), the two emotional information channels investigated in the present study. Based on the observation that semantically (e.g., Van Petten & Kutas, 1988) and emotionally (e.g., Schirmer, Kotz, & Friederici, 2002; Schirmer, Kotz, & Friederici, 2005) mismatching context information leads to integration difficulties of mismatching information, we argue the following. If expectancy deviation leads to increased integration difficulty, such integration difficulty should result in different event-related potential (ERP) modulations as a function of information specific expectancy deviation. Thus, emotional prosodic expectancy deviation should elicit a different ERP response than expectancy deviation in context of emotional prosody and emotional semantics. While emotional prosodic integration is referred to acoustic correlates such as perceived pitch, duration, and intensity, and their online integration in a speech stream, combined emotional prosodic/semantic integration requires the integration of acoustical information and lexical information. Our previous data clearly revealed qualitatively different ERP components elicited as a function of integration during emotional prosodic processing and combined emotional prosodic/semantic processing. More specifically, we showed that expectancy violations of emotional (angry and happy) prosody elicited a right-lateralized positivity, while combined emotional prosodic/semantic (angry and happy) expectancy violations elicited an early negativity in the ERP (Kotz & Paulmann, 2007). These results are also in line with studies that showed differential processing for linguistic prosodic processing and combined prosodic/semantic processing (Astésano, Besson, & Alter, 2004).

The present study follows up our previous work by extending these different processing dynamics in three dimensions. First, the minimal use of different emotional prosodies (i.e., angry, happy) was extended. Building on the assumption that human emotions engage distinct neural networks (e.g., Adolphs, 2002) that relate to emotion-specific processing mechanisms, we investigated six basic emotional prosodies (anger, disgust, fear, happiness, pleasant surprise, and sadness), even though some previous evidence argues for emotion-independent integration of emotional prosody and semantics (e.g., Kotz & Paulmann, 2007; Schirmer et al., 2002; Schirmer et al., 2005). Thus, one critical research question at stake was to investigate if the integration of emotional prosody and semantics is truly emotion-independent or not. Second, we included different speaker voices (female/male). This is a critical manipulation, as emotional and linguistic information are encoded via the properties of a speaker voice. We put this to test as our previous evidence indicated encoding differences for female and male speaker voices in a cross-language perception study (Pell, Kotz, Paulmann, & Alasseri, 2005). Furthermore, this manipulation is crucial as natural differences between female and male voices are reported, i.e., female voices are usually higher pitched than male voices (e.g., Lattner, Meyer, & Friederici, 2005). We hypothesize that if previously reported effects were related to prosodic change detection mechanism (in emotional context) only, they should operate relatively independently of speaker voice. Last, to investigate the nature of emotional prosody processing, stimuli with no semantic content (e.g., pseudo-stimuli) should lead to clear prosodic effects independent of lexical-semantic information. Comparable to filtered speech, morphologically marked pseudo-sentences spoken with the varying emotional intonation patterns allow eliminating lexical content while preserving emotional prosody. In particular, to understand the relative contribution and the underlying mechanisms of emotional prosody it seems necessary to test this information type without confounding information such as emotional semantics. To this end, we tested six basic emotional intonations in emotionally and morphologically marked pseudo-sentences (emotional prosodic condition) and in lexical-sentences (combined emotional prosodic/semantic condition). This approach allows comparing the unfolding of emotional prosody independent of lexical content with the online integration of emotional prosody with emotional semantics in lexical content.

As our previous evidence showed a right-lateralized positive ERP component for the processing of 'pure' emotional prosodic mismatches and additional evidence (e.g., Vingerhoets, Berckmoes, & Stroobant, 2003) suggests emotional prosodic lateralization, a replication and extension of such a lateralized effect should add to the existing debate on the lateralization of emotional prosody.

To summarize, the current experiment aims to further substantiate and to specify the (temporal) integration of prosody and semantics in emotional context. Moreover, we aim to further specify the relative contribution of emotional prosody and emotional prosody with semantics in an emotional utterance by establishing an ERP correlate for prosodic and prosodic/semantic deviance detection in emotional context. Based on sparse previous evidence (Kotz & Paulmann, 2007), we hypothesize that if registration of expectancy violation is truly valence-independent, i.e., the same response is elicited for positive and negative emotions, emotional prosodic expectancy violations of all six basic emotions should lead to an early right-lateralized positive ERP effect, while violations of emotional prosody/semantic expectancies should lead to an early bilaterally distributed negative ERP effect (see also Astésano et al., 2004; Kotz & Paulmann, 2007). In particular, we asked whether previously reported ERPs elicited in response to prosodic expectancy violations embedded in neutral semantic could be compared to ERPs elicited in response to expectancy violations when no lexical content is present. Theoretically, there should be no difference between the two approaches, i.e., we expect to replicate earlier results that is, prosodic expectancy violations elicit a right-lateralized positive ERP component shortly after the splicing point. Also, if the two hypothesized expectancy violation effects are solely related to emotional prosodic (and semantic) integration processing, they should be elicited irrespective of speaker type.

2. Materials and methods

2.1. Material

We presented semantically and prosodically matching stimuli (210 lexical- and 210 pseudo-sentences) for each of six basic emotions (anger, fear, disgust, happiness, pleasant surprise, sadness) and neutral (30 sentences each). In addition, the original recordings (i.e., the lexical- and pseudo-sentences) were cross-spliced in two ways:

2.1.1. Prosodic expectancy violation condition

We merged the acoustical signals of the first part of a prosodically neutral start of a pseudo-sentence (e.g., "Hung set/Mon set") and the second part of an emotional-prosodically end of a pseudo-sentence (e.g., "den Nestol verbarsicht ind gekobelt").

2.1.2. Combined semantic-prosodic expectancy violation condition

Here, we merged the acoustical signal of the first part of a semantically and prosodically neutral start of a sentence (e.g., "Er hat/Sie hat"; translation: "He has/She has") and the second part of an emotional semantically and prosodically matching end of a sentence (e.g., "den Gewinn verdoppelt und verdreifacht"; literal translation: "doubled and tripled the prize") (see Fig. 1 for graphical display of splicing procedure).

The splicing procedure resulted in 180 cross-spliced lexical- and 180 cross-spliced pseudo-sentences. As all sentences were spoken by a female and a male speaker, a total of 1560 trials were presented in two sessions. The mean splicing point was calculated by measuring the mean duration of the neutral start of the sentences that were used as splicing templates ("Er hat/Sie hat" vs. "Hung set/Mon set" articulated by two speakers). The mean splicing point occurred at 350 ms after sentence onset for emotional prosodically incongruent sentences and at 338 ms after sentence onset for combined incongruent sentences.

Emotional prosodic valence ratings were obtained in two earlier rating studies (one for the lexical and one for the pseudo-sentences; see Paulmann, Pell, & Kotz, 2008, for details on the lexical rating study). In both studies, participants rated the emotional category of each sentence in a forced-choice task (seven response alternatives). In addi-

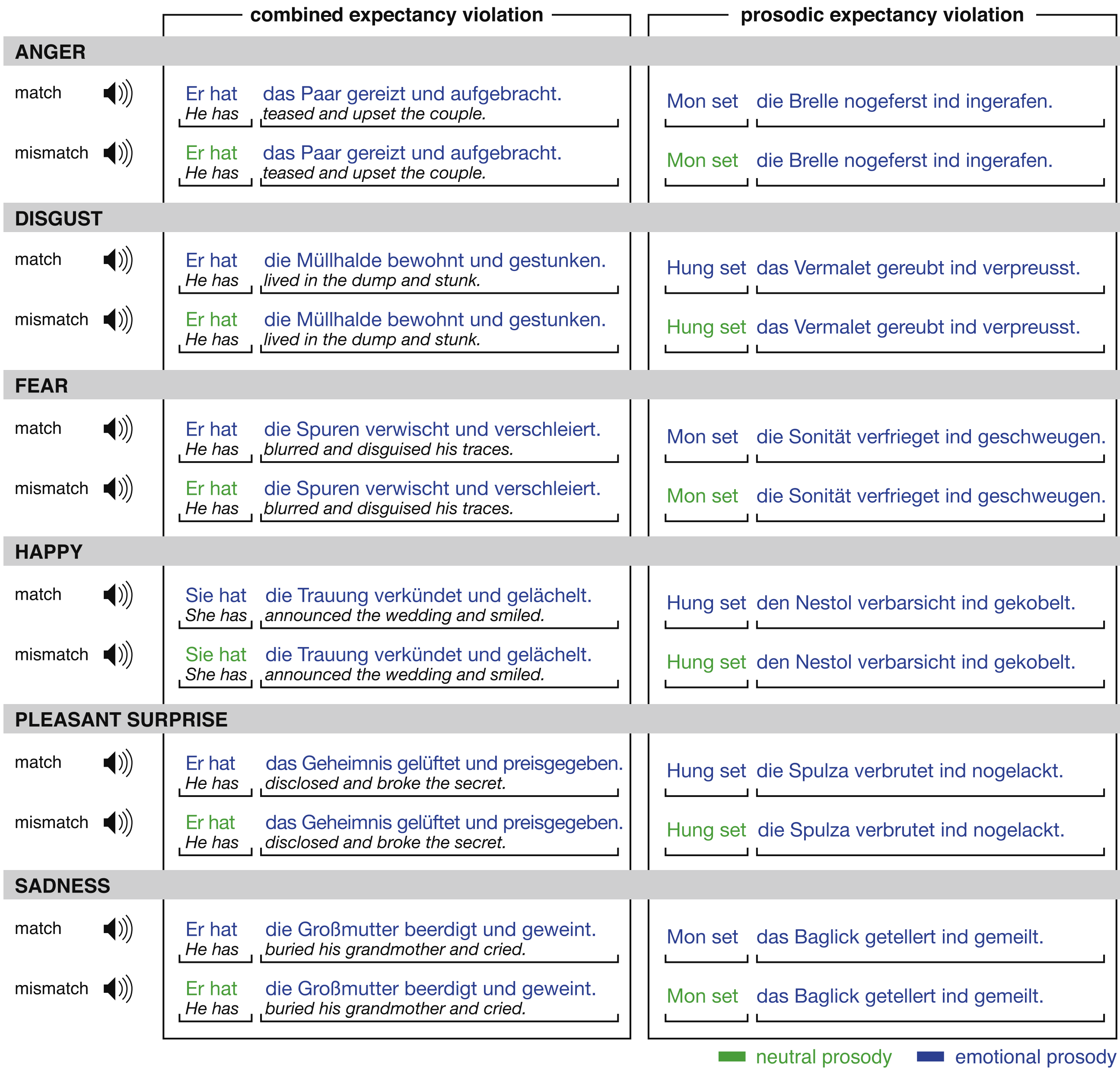

Fig. 1. The illustration explains the splicing procedure. The left hand column displays sentence examples from the combined prosodic/semantic expectancy violation, whereas the right hand column displays sentence examples from the prosodic expectancy violation condition.

tion, the sentences were rated according to their intensity on a 5-point scale (ranging from −2 to +2). In the following, the percentage correct (%) and respective standard deviations (*SD*) for each emotional category and sentence type are listed. Lexical-sentences: anger: 96% (*SD*: 3.9), disgust; 95% (*SD*: 6.9), fear: 72% (*SD*: 14.2), happy: 75% (*SD*: 9.4), neutral: 95% (*SD*: 5.0), pleasant surprise: 49% (*SD*: 12.0), sad: 81% (*SD*: 12.7). Pseudo-sentences: anger: 91% (*SD*: 3.4), disgust; 70% (*SD*: 9.4), fear: 64% (*SD*: 12.3), happy: 58% (*SD*: 8.3), neutral: 94% (*SD*: 3.1), pleasant surprise: 55% (*SD*: 8.8), sad: 71% (*SD*: 7.5).

Nouns and verbs in lexical-sentences were controlled for word frequency so that there was no difference between the emotional categories (Baayen, Piepenbrock, & van Rijn, 1995). All sentences were taped with a video camcorder attached to a high-quality clip-on microphone and later digitized at 16-bit/44.1 kHz sampling rate. The stimulus material was presented in a mono format and was prosodically analyzed (i.e. pitch, intensity, and duration) using *Praat* (Boersma & Weenink, 2003; see Table 1 for results).

2.2. Participants

Fifteen healthy female (mean age: 24 years (*SD* 1.96)), and 15 healthy male (mean age: 25 years (*SD* 2.02)) right-handed volunteers that had not participated in the rating studies took part in two counterbalanced experimen-

Table 1
The table lists results from the acoustical analyses of the unspliced (matching) sentences that were included in the current study

Type	EMOTION	Mean F_0 (Hz)	Mean dB	Duration (s)
Lexical-sentences	Anger	268.26 ± 32.86	66.54 ± 2.35	2.58 ± 0.24
	Disgust	185.97 ± 67.81	66.24 ± 2.56	2.94 ± 0.56
	Fear	184.57 ± 68.25	65.49 ± 2.89	3.37 ± 1.03
	Happy	199.01 ± 65.83	65.03 ± 2.48	2.47 ± 0.22
	Neutral	159.16 ± 40.08	65.93 ± 2.51	2.72 ± 0.34
	Pleasant surprise	290.28 ± 73.32	66.80 ± 3.11	2.53 ± 0.28
	Sadness	159.57 ± 33.64	65.88 ± 2.41	2.71 ± 0.35
Pseudo-sentences	Anger	259.85 ± 30.73	71.03 ± 1.93	2.85 ± 0.33
	Disgust	190.46 ± 44.39	69.45 ± 2.53	3.60 ± 0.42
	Fear	189.99 ± 61.85	68.26 ± 2.87	3.62 ± 0.78
	Happy	260.51 ± 80.77	71.28 ± 1.94	2.86 ± 0.34
	Neutral	169.46 ± 34.53	69.04 ± 3.00	3.40 ± 0.50
	Pleasant surprise	313.87 ± 66.32	71.41 ± 2.47	3.15 ± 0.34
	Sadness	181.11 ± 26.70	69.33 ± 2.29	3.07 ± 0.41

Means displayed are collapsed across speaker voices. The table lists mean pitch (measured in Hz), mean intensity (measured in dB), and duration of sentence (measured in seconds) as well as respective standard deviations ($\pm$).

tal sessions separated by at least 3 days. All participants were university students and native speakers of German with no reported hearing or neurological problems. All participants had normal or corrected-to-normal vision and were paid for their participation.

2.3. Procedure

Each participant was seated comfortably at a distance of 115 cm from a computer monitor in a sound-attenuating chamber. Sentences were presented via loudspeakers located to the left and right side of the computer. After sentence offset a visual probe was presented for 300 ms. Participants had to respond as quickly and accurately as possible if the visual probe word had occurred in the previous sentence or not. Half of the participants pressed the yes-button with their right hand and the no-button with their left hand (vice versa) on a two-button response panel. The *Experimental Run Time System* (ERTS; Beringer, 1993) was used to run the experiment. Participants were asked to avoid eye movements during sentence presentation. The inter-trial-interval was 1500 ms. Four experimental lists contained all sentences in pseudo-randomized order. In each list, half of the sentences were spoken by the female speaker, the other half by the male speaker. One version of one of the four lists was presented in the first session and the other version was presented in the second session. Half of the participants began with the first version and the remaining with the second version.

The electroencephalogram (EEG) was recorded with 59 Ag–Ag Cl electrodes mounted in an elastic cap according to the modified expanded 10–20 system (American Electroencephalographic Society), each referenced to the nose (NZ). Bipolar horizontal and vertical EOGs were recorded for artifact rejection purposes. Artifacts caused by eye or muscular movements were rejected off-line on a subject-by-subject basis and were omitted from the ERP analysis. Trials contaminated by EOG or movement artifacts were rejected with a threshold of 30.00 μV. Electrode resistance was kept below 5 k-Ω. Data were re-referenced offline to linked mastoids. EEG signals were recorded continuously with a band pass between DC and 70 Hz and digitized at a rate of 250 Hz. ERPs were filtered off-line with a 7 Hz low pass for graphical display only.

Based on previous evidence and additional time-line analyses, ERP time windows were defined (see Kotz & Paulmann, 2007).[1] Thus, ERPs were averaged with a 200 ms pre-stimulus baseline for the following time windows: between 450 and 650 ms (i.e. for the combined prosodic/semantic expectancy violation, and between 700 and 1000 ms for the prosodic expectancy violation. All statistical analyses were computed including correct trials only.

2.4. Data analysis

Behavioral results were not analyzed because no prior hypotheses were made with regard to the effect expectancy deviance detection may have on probe verification. Please note that the task was only administered to ensure that participants listened to the sentences. This was clearly the case as accuracy scores for the two different conditions revealed very good sentence comprehension (matching pseudo-sentences: 93.2% *SD*: 5.8; mismatching pseudo-sentences: 92.8% *SD*: 5.9; matching lexical-sentences: 97.5% *SD*: 3.7; mismatching lexical-sentences: 96.8% *SD*: 4.0).

ERPs for the two different conditions were calculated with separate analyses of variance (ANOVAs) to analyze the two different time windows.[2] Mean amplitudes were calculated treating *M* (match: prosodically (and semantically) matching stimuli vs. mismatch: mismatching stimuli), *P* (emotional prosodies of anger, disgust, fear, happiness, pleasant surprise, sadness), and *Speaker* (female vs. male voice) as repeated-measures factors. In addition, we included the factor *Session* (first vs. second) to control for possible session effects. Distributional differences were calculated by the factors *HEMI* (left vs. right hemisphere) and *REG* (anterior, central, and posterior region; for list of electrodes in each region see Fig. 2). We also included *Sex*

[1] As our previous research revealed ERP effects 90 ms after the splicing point for the combined violation and 340 ms after the splicing point for prosodic expectancy violations, we aimed to replicate these effects in comparable time-windows in the current study. Based on this evidence and additional time-line analyses, we thus ran statistical analyses for ERP effects 110 ms after the splicing point for the combined violation and 350 ms after the splicing point for prosodic expectancy violations.

[2] Even though the current experimental design would have allowed to include a factor time-window in the statistical analysis, we opted to stay consistent with our previous approach to data analysis in order to be able to compare results (i.e., test-retest reliability; see Kotz & Paulmann, 2007). Thus, the two conditions were analyzed separately in two different time windows.

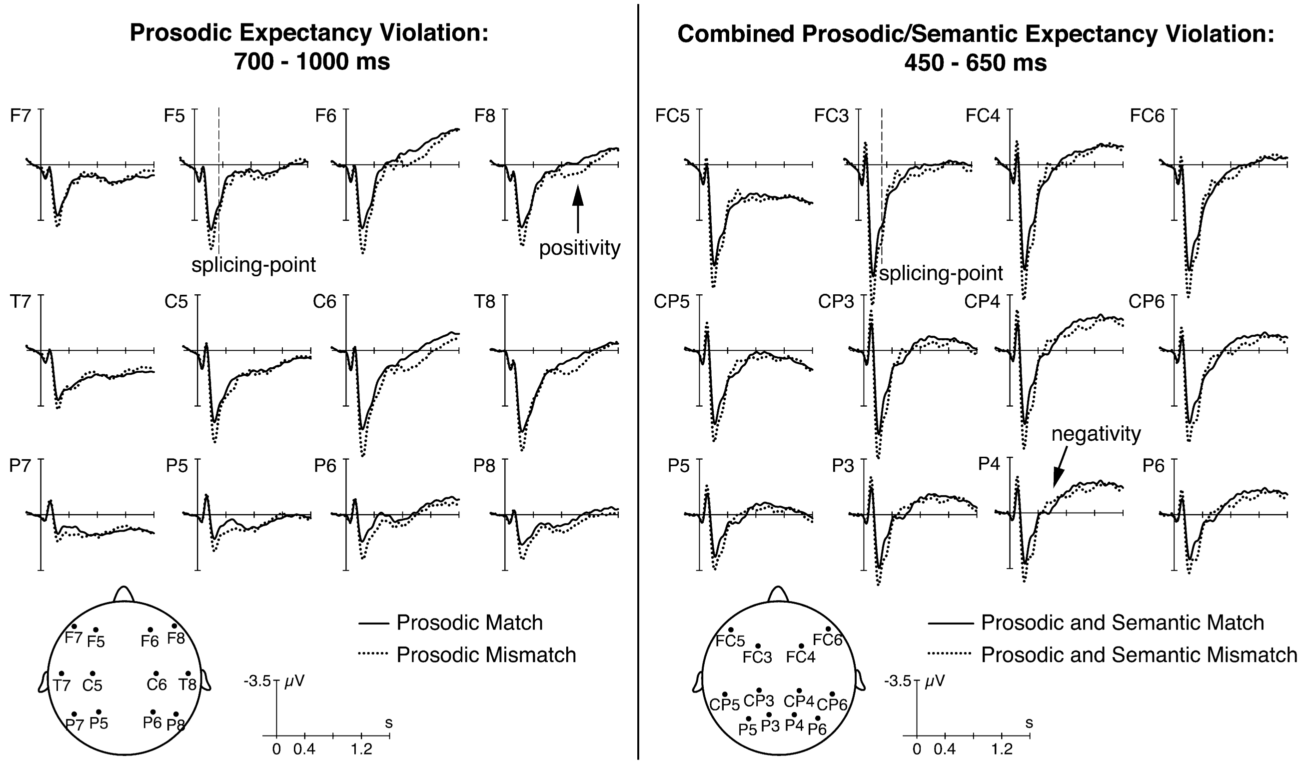

Fig. 2. The illustration shows ERPs elicited by spliced and unspliced sentences at selected electrode-sites. For both conditions, waveforms show the average for emotional incongruent sentences (dotted) and emotional congruent sentences (solid) from 200 ms prior to stimulus onset up to 1200 ms post-stimulus onset.

(female/male) as a between-subject factor.[3] The Geissser–Greenhouse correction (Geisser & Greenhouse, 1959) was applied to all repeated-measures with greater than one degree of freedom in the numerator. Only significant interactions with critical factors (P, M) are reported in step-down analyses. Effect size was estimated by omega-square (Ω^2), i.e., the coefficient of determination that represents the proportion of variance in the dependent variable accounted for by the independent variable. For between-subject designs, Ω^2 effect sizes greater than 0.138 are considered large effects, indices between 0.048 and 0.138 are considered medium effects, and values between 0.0009 and 0.048 are considered small effects (c.f., Olejnik & Algina, 2003).

3. ERP Results

3.1. Prosodic expectancy violation: 700 ms to 1000 ms

In the time window between 700 and 1000 ms a significant effect of M was found ($F(1,28) = 4.13$, $p = .05$, $\Omega^2 = 0.05$) with more positive-going waveforms for emotional prosodically mismatching sentences than for matching sentences. Also, the critical main effect for M interacted significantly with *HEMI* ($F(1,28) = 11.99$, $p < .01$, $\Omega^2 = 0.08$). Step-down analyses by *HEMI* revealed a significant M effect ($F(1,28) = 8.43$, $p < .01$, $\Omega^2 = 0.11$) at right hemisphere electrode-sites, indicating more positive-going ERP waveforms for emotional prosodically mismatching than for matching sentences. Moreover, there was a significant interaction between $Speaker \times P \times M$ ($F(5,140) = 4.31$, $p < .01$, $\Omega^2 = 0.02$). Step-down analyses were carried out by *Speaker* and then by *Speaker* and P which revealed a significant M effect for sentences spoken by the male speaker with an angry prosody ($F(1,28) = 4.18$, $p = .05$, $\Omega^2 = 0.05$), showing more positive-going waveforms for mismatching than matching sentences. In addition, we found an M effect for fearful prosody ($F(1,28) = 20.00$, $p < .001$, $\Omega^2 = 0.24$). Here, mismatching sentences showed more negative-going ERP waveforms than matching sentences.

Taken together, the results revealed a positive-going ERP effect at right hemisphere electrode-sites for all emotional prosodic mismatching sentences in comparison to matching sentences. Results also indicate that only mismatching angry sentences spoken by a male speaker elicit a bilateral positivity for mismatching sentences, while fearful mismatching sentences elicit a more negative-going ERP waveform than matching fearful sentences. (See Fig. 2 for a graphical display of ERPs for all prosodically matching and mismatching sentences.)

[3] The present study investigated the time-course of emotional prosody and combined emotional prosody with semantics. As previous research has suggested that listeners' sex may influence implicit emotional prosody processing, *Sex* was included as a between-subject factor.

3.2. Combined semantic/prosodic expectancy violation: 450 ms to 650 ms

In the time window of 450–650 ms, the ERP analysis revealed a main effect for *Speaker* ($F(1,28) = 16.29$, $p < .001$, $\Omega^2 = 0.20$), with amplitudes being more negative-going for the male than the female speaker. A critical main effect of *M* was also significant ($F(1,28) = 8.54$, $p < .01$, $\Omega^2 = 0.11$) revealing a more negative-going component for mismatching than matching sentences. This effect was not qualified by region. An interaction between *Speaker* and *P* ($F(5,140) = 2.7$, $p < .05$, $\Omega^2 = 0.01$) allowed for a step-down analysis by *Speaker*. The *P* effect was only significant for sentences articulated by the female speaker ($F(5,140) = 3.03$, $p < .05$, $\Omega^2 = 0.03$) revealing varying negative-going amplitudes for emotional sentences; however, since *P* did not interact with the critical factor *M*, post-hoc comparisons were not carried out.

Again, results revealed a significant difference between ERP amplitudes elicited by the male and the female speaker. Moreover, results showed a whole-head distributed negative-going ERP effect for combined prosodically and semantically mismatching sentences compared to matching sentences (see Fig. 2 for a graphical display of ERPs for all prosodically and semantically matching and mismatching sentences and see Table 2 for summary of ERP results).

4. Discussion

Previous investigations on the integration of emotional prosody and emotional semantics only allowed looking into this phenomenon by means of cross-modal manipulations (e.g., Schirmer et al., 2002; Schirmer et al., 2005). However, this approach does not reflect natural online integration of different information types. By means of cross-splicing, the integration and time-course development of different information types can be time-locked and give a more direct measure of online integration (Kotz & Paulmann, 2007). The present study aimed to substantiate and to specify the integrative time-course of emotional prosody without semantics and of emotional prosody with emotional semantics. In this context, we also tested whether expectancy violations of either information type varied as a function of emotional valence or speaker gender. The current results substantiate previous ERP effects (Astésano et al., 2004; Kotz & Paulmann, 2007), but extend these to pseudo-sentences spoken in six basic emotions by both female and male speakers. The results clearly show that specific brain responses occur with a different time-course when (1) emotional prosodic expectancy is not fulfilled, or (2) when expectancy of both emotional semantics and emotional prosody is not fulfilled. The ERP effects elicited in the two conditions thus show a different time-course, topography, and polarity. In the following, these effects will be discussed in relation to previous evidence; however, as the combined expectancy violation effect was mainly a replication and extension of our previous result (Kotz & Paulmann, 2007), specific emphasis is given to the emotional prosodic expectancy violation condition out of semantic context.

Table 2
Summary of ERP F- and *p*-values for both conditions in both time windows for all main effects

Source	*df*	*F* Value	*p* Value
Prosodic expectancy violation: 700–1100 ms			
Sex	1.28	0.01	.9396
Session	1.28	2.35	.1368
Speaker	1.28	3.57	.0691
P	5.140	1.19	.3181
M	1.28	4.13	.0518
HEMI	1.28	9.04	.0055
REG	2.56	1.26	.2799
Combined semantic/prosodic expectancy violation: 450–650 ms			
Sex	1.28	1.30	.2633
Session	1.28	4.46	.0437
Speaker	1.28	16.29	.0004
P	5.140	1.61	.1761
M	1.28	8.54	.0068
HEMI	1.28	2.37	.1349
REG	2.56	25.16	<.0001

4.1. Expectancy violations in emotional context: A speaker-independent effect?

Shortly after the splicing point (approx. 350 ms), emotional prosodic expectancy violations elicit a right-lateralized positive-going ERP effect while combined semantic/prosodic expectancy violations elicited an early negative ERP effect (approx. 110 ms after the splicing point). As the transfer of prosodic (emotional or linguistic) information occurs via the properties of the speaker voice, we asked if expectancy violations are detected independent of speaker voice. This hypothesis was confirmed: prosodic expectancy violation elicited a positivity, and combined expectancy violation elicited a negativity irrespective of speaker voice. Even though the current data reveal a general speaker effect in the combined violation (not discussed here), no significant interaction between expectancy violation and speaker voice was observed. This strengthens the interpretation that processing of prosodic expectancy violations in an emotional context is not dependent on the gender of the speaker voice, but may solely be related to the emotional attributes of a stimulus (see Kotz & Paulmann, 2007, and also discussion below).

4.2. Expectancy violations in emotional context: A valence-independent effect?

Most existing electrophysiological studies on the integration of emotional channels have concentrated on one or two specific emotions (e.g., Kotz & Paulmann, 2007; Pourtois et al., 2000; Schirmer et al., 2002; Schirmer et al., 2005). However, there is evidence that processing different basic human emotions may engage distinct neural networks (e.g., Adolphs, 2002). This observation leads to the assumption of emotion-specific processing mechanisms that we investigated here. If the detection and the analysis of emotion-specific expectancy violations are modulated by distinct processing mechanisms, this should be mirrored in systematic, diverse ERP effects for different emotional prosodies. The present results speak against an emotion-specific detection mechanism of expectancy violations as all emotional prosodic contour deviations with no lexical content elicited a comparable positivity shortly after the splicing point, and all emotional prosodic contour deviations with lexical content present elicited a similar negativity shortly after the splicing point. This replicates and extends our previous evidence, and supports valence-independent processing of emotional prosodic (and semantic) expectancy violations (Kotz & Paulmann, 2007).

Indeed, all prosodic contour deviations, except for the emotional category of fear uttered by the male speaker, elicited a similar positivity. Therefore, the response to prosodic expectancy violations in emotional context seems to be purely prosodic. Since no robust emotion effect is observed, one may interpret this result as a general prosodic mismatch effect. Considering the late positive ERP component (P800) observed by Astésano et al. (2004) that was elicited by linguistic prosodic mismatches and was closely linked to F_0 contour violations, it seems reasonable to conclude that prosodic contour mismatches may always elicit a positive ERP component. However, as discussed elsewhere, emotional prosody is conveyed not only via pitch modulation but also via more fine-grained acoustical parameters (see e.g., Banse & Scherer, 1996). We therefore propose that the positivity elicited by expectancy violations reflects an on-line registration of a transition from a neutral to an emotional significant stimulus.

Interestingly, the present results suggest that prosodic deviance detection is not directly related to one particular acoustical parameter configuration as, for example, both, sentences going from neutral into high pitched pleasant surprise or low pitched sad sentences elicited a similar positive ERP component (see Table 1 for acoustical information on matching sentences). Furthermore, the emotional significance of our stimuli seems to promote registering the mismatch irrespective of the focus of attention as the effects were elicited under implicit emotional prosody processing task instruction. Additionally, the onsets of the two ERP effects differ. This suggests that the two effects may be related, but may not reflect the same process(es). More specifically, prosodic contour mismatches in emotional context may be responded to faster than prosodic contour mismatches in linguistic context due to the evolutionary significance of emotional stimuli. Even though this assumption is highly speculative, the current result fits nicely with reports in the literature that claim it may be evolutionary advantageous to rapidly detect unexpected emotional events (Vuilleumier, 2005). However, this needs to be further investigated. For instance, a study including linguistic prosodic deviations *and* emotional prosodic deviations compared to the same neutral control sentence could help to specify if the same positive ERP component is elicited. However, careful temporal manipulation needs to be assured to ensure correct temporal comparisons. In particular, the possible difference in temporal aspects of emotional prosody and linguistic prosody processing could be specified in such a paradigm. Moreover, future research will have to specify if the positivity observed here and the positivities elicited by other prosodic deviances (such as CPS, or P800) belong to the same family or not.

4.3. Expectancy violations in emotional context: Differences between lexical and pseudo-sentences?

In general, the present findings suggest that responses to prosodic and prosodic/semantic expectancy violations in emotional context are emotion- and speaker-independent and thereby replicate and extend previous results (Astésano et al., 2004; Kotz & Paulmann, 2007). As the two expectancy violations elicited two distinct ERP components following a different time-course, latency, and polarity, it is assumed that the integration of emotional prosody alone and of emotional prosody with semantics is subserved by different underlying mechanisms. While prosodic expectancy violations elicited a positivity, combined violations elicited a negative ERP component. The assumption that emotional prosody and emotional prosody with semantic information is processed by different underlying mechanisms gets some support from recent imaging studies. For instance, Vingerhoets et al. (2003) investigated blood flow velocity (BFV) in healthy participants during the identification of emotions conveyed by prosody or semantics. When attention was paid to emotional semantics, a left-hemispheric lateralization of BFV was reported while a right-hemispheric lateralization was reported when attention focused on emotional prosody. This result substantiates previous evidence suggesting a distinct emotional prosodic lateralization (Vingerhoets et al., 2003; see also Kotz, Meyer, & Paulmann, 2006, chap. 15, for factors influencing heterogeneous evidence on emotional prosody processing and lateralization). Functionally speaking, the present result suggests that irrespective of semantic valence, emotional semantic information seems to predominate emotional prosodic information. The results add evidence to the hypothesis that semantics cannot be ignored even if not attended to (e.g., Besson et al., 2002). However, it remains to be shown whether the brain response was triggered by the emotional semantic mismatch alone, since the effect did not occur independent of emotional prosody.

We propose that the negative ERP component elicited by combined expectancy violations reflects semantically driven expectancy violation detection and is comparable to the well-known N400 that has been shown to be larger for semantically incongruent words than for congruent ones. While it could be argued that the whole-head distribution speaks against an N400-like negativity, the fact that auditory studies often report more frontally distributed N400s in addition to its classical central posterior distribution makes this argument rather unlikely (e.g., Connoly & Phillips, 1994; Connoly, Phillips, Stewart, & Brake, 1992; Connoly, Stewart, & Phillips, 1990). In addition, there is some research that aimed to localize the source of the N400. These studies report several mainly, but not exclusively, left lateralized fronto-temporal structures (Halgren et al., 2002).

However, a striking result is the observation that the current negativity is elicited very shortly after the splicing point (approx. 100 ms). This early effect can be explained in three ways: first of all, the emotional significance of a stimulus may trigger faster speech processing. Indeed, our results suggest that both semantic and prosodic processing can be speeded up if occurring in an emotional context. Again, to verify this hypothesis future studies testing emotional and non-emotional semantic mismatches in one experiment are needed. Secondly, the early onset of the negativity may be linked to a moment when acoustic information signals a deviation from a semantic expectancy (see also Kotz & Paulmann, 2007, and Van Petten, Coulson, Rubin, Plante, & Parks, 1999). Alternatively, it could be argued that the current negativity is an N200-like component; however, the fact that the N200 is usually reported to be more frontally/centrally distributed (e.g., Pritcharsd, Shappell, & Brandt, 1991; Näätänen, 1990) makes this interpretation less plausible. Last, it could be argued that the task utilized here may have put the spotlight on to the semantic content and thus semantic processing dominates emotional prosody processing as a function of task. However, our previous evidence suggests that this is rather unlikely as we have observed a similar negative ERP effect under implicit *and* explicit task instructions (see Kotz & Paulmann, 2007, for detailed discussion). Moreover, there is growing evidence that the integration of different emotional channels, i.e., not only prosody and semantics, but also prosody and mimic, occurs particularly early during processing (e.g., Kotz & Paulmann, 2007; Pourtois et al., 2000). In fact, building on the observation that more emotional channels facilitate subtle transitions in an emotional sentence context, one can hypothesize that the more channels available, the faster emotional integration occurs (see Kotz & Paulmann, 2007 for similar argument).

Last, we asked if ERP responses to prosodic expectancy violations embedded in neutral semantic context would differ from responses to expectancy violations embedded in pseudo-sentences. As the current positive expectancy violation effect was triggered in pseudo-sentences, we suggest that the effect is truly elicited by an emotional prosodic expectancy deviation. This renders the conclusion that emotional prosodic expectancy violations linked to a right-lateralized positivity reported were not influenced by neutral semantic content. Several accounts for such a dominant right-lateralized processing of emotional prosodic stimuli have been presented in the literature (for reviews see e.g., Baum & Pell, 1999; Kotz et al., 2006, chap. 15; Lalande, Braun, Charlebois, & Whitaker, 1992). For instance, recent research suggests that the right hemisphere is dominantly involved in emotional prosodic attributes processing, while the left hemisphere is argued to be involved more strongly in linguistically based processing (e.g., Friederici & Alter, 2004; Pell, 2006; Van Lancker Sidtis, Pachana, Cummings, & Sidtis, 2006). The data we present here cannot rule out a left-hemispheric contribution when processing emotional prosody as we found a whole-head distributed negativity for combined prosodic/semantic expectancy violations; however, the data point to the fact that when emotional prosody is processed independent of (emotional) semantic content, the prosodic expectancy violation results in a right-lateralized effect. Moreover, the idea that the left hemisphere integrates verbal-semantic processes with emotion- or pitch-related right hemisphere processes is strengthened by our data (Pell, 2006). Clearly, the detection of prosodic expectancy violations in emotional context depends on emotional prosodic attributes processing, while the detection of a combined prosodic and semantic expectancy violation should depend on *both* emotional prosodic parameter processing and lexical-semantic processing. The current results thus suggest that the left hemisphere is engaged in prosodic *and* semantic information processing, while the right hemisphere may respond more distinctly when emotional prosody without semantic information is processed. As mentioned further above, our evidence is in line with a recent study by Vingerhoets et al. (2003) that reports a significant right-lateralized blood flow velocity (BFV) effect when identifying emotional prosody in contrast to a left-sided specialization when identifying an emotion conveyed by semantics. However, considering the multiplicity of acoustic parameters that signal emotional prosody it remains to be determined which of these parameters are most strongly supported by the right hemisphere.

Finally, as mentioned elsewhere (e.g., Kotz & Paulmann, 2007; Schirmer & Kotz, 2006), one needs to be cautious when drawing conclusions about underlying generators of ERPs and thus the surface distribution of ERP components as this surface realization component is a consequence of summation effects from underlying brain sources. Thus future brain imaging studies are needed to specify the engagement of the left and right hemisphere during expectancy violation processing in emotional and linguistic context. In particular, it will be stimulating to explore the underlying brain structures subserving emotional prosodic (and semantic) deviance detection. However, neuroimaging studies have low temporal resolution. Thus, in order to specifically link brain structures to one emotional function at a given point in time it may be helpful to study patients with specific brain lesions with ERPs. The current paradigm seems to be suitable to further specify impairments in patients who suffer from emotional speech comprehension difficulties (e.g., patients with lesions of the basal ganglia or orbito-frontal brain areas). By establishing different ERP correlates for emotional prosodic and emotional prosodic/semantics processing, the nature of a possible deficit can be specified i.e., is it prosodic or semantic. Such studies are currently underway.

5. Conclusions

To summarize, the present findings replicated and extend previous results (Astésano et al., 2004; Kotz & Paulmann, 2007) by introducing pseudo-sentences spoken in different emotional prosodies and by different speakers. Speaker identity does not influence the processing of prosodic or combined prosodic/semantic expectancy violations in an emotional context. This suggests detection of expectancy violations independent of speaker voice. The current study gave rise to different ERP correlates elicited by prosodic or combined prosodic/semantic deviance detection in emotional context. These effects not only show a different polarity, but also a different time-course, i.e., deviance detection is faster if both emotional prosody and semantics signal this change. Together with previously reported results this suggests distinct brain signatures of emotional sentences processing. Finally, it is believed that the current positive ERP component elicited by pure prosodic expectancy violations is closely linked to acoustic parameter deviation detection that may reflect an on-line registration of a transition process during which an auditory stimulus takes on an obvious emotional prosodic significance. Whether this positivity is emotional prosody specific or comparable to positivities elicited by linguistic prosodic deviations remains a matter of future investigations. In contrast, the negative ERP effect elicited in the combined expectancy violation condition is assumed to reflect prosodic and semantic integration difficulty in sentence context. Thus, it is concluded that the influence of emotional prosody enhances the propositional intent of an utterance, whether in semantic-prosodic matching or mismatching presentation and occurs irrespective of the speaker gender.

Acknowledgments

We wish to thank Kristiane Werrmann for help with subject recruitment and data acquisition. This work was funded by the German Research Foundation (DFG FOR 499 to S.A. Kotz).

References

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology, 12*, 169–177.

American Electroencephalographic Society (1991). Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology, 8,* 200–202.

Astésano, C., Besson, M., & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research, 18*, 172–184.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1995). *The CELEX Database (CD-ROM). Center for lexical information*. Nijmegen: Max Planck Institute for Psycholinguistics.

Banse, R., & Scherer, K. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614–636.

Baum, S. R., & Pell, M. D. (1999). The neural basis of prosody: Insights from lesion studies and neuroimaging. *Aphasiology, 13*, 581–608.

Begleiter, H., Porjesz, B., & Garozzo, R. (1979). Visual evoked potentials and affective ratings of semantic stimuli. In H. Begleiter (Ed.), *Evoked Brain Potentials and Behavior* (pp. 127–141). New York: Plenum Publishing Corporation.

Beringer, J. (1993). *Experimental run time system (computer software)*. Frankfurt/Main, Germany: BeriSoft Corporation.

Besson, M., Magne, C., & Schön, D. (2002). Emotional prosody: Sex differences in sensitivity to speech melody. *Trends in Cognitive Science, 6*, 405–407.

Boersma, P. & Weenink, D. Praat: Doing phonetics by computer (Version 4.1.13), computer program, 2003. Retrieved July 15, 2003, from http://www.praat.org/.

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

Connoly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of terminal word of spoken sentences. *Journal of Cognitive Neuroscience, 6*, 256–266.

Connoly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language, 43*, 1–18.

Connoly, J. F., Stewart, S. H., & Phillips, N. A. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language, 39*, 302–318.

Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: a dynamic dual pathway model. *Brain and Language, 89*, 267–276.

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

Halgren, E., Dhond, P. R., Christensen, N., Van Petten, C., Marinkovic, K., Lewine, J. D., & Dale, A. M. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage, 17*, 1101–1116.

Kotz, S. A., Meyer, M., & Paulmann, S. (2006). Lateralization of emotional prosody in the brain: An overview and synopsis on the impact of study design. In Anders, Ende, Junghöfer, Kissler, Wildgruber (Eds.), *Progress in Brain Research, 156,* pp. 285–294.

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

Lalande, S., Braun, C. M. J., Charlebois, N., & Whitaker, H. A. (1992). Effects of right and left hemisphere cerebrovascular lesions on discrimination of prosodic and semantic aspects of affect in sentences. *Brain and Language, 42*, 165–186.

Lattner, S., Meyer, M., & Friederici, A. D. (2005). Voice perception: Sex, pitch, and the right hemisphere. *Human Brain Mapping, 24*, 11–20.

Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods, 8*, 434–447.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences, 13*, 201–288.

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

Paulmann, S., & Kotz, S. A. (in press). Early emotional prosody perception based on different speaker voices. Neuroreport.

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

Pell, M. (2006). Cerebral mechanisms for understanding emotional prosody in speech. *Brain and Language, 96*, 221–234.

Pell, M. D., Kotz, S. A., Paulmann, S., & Alasseri, A. (2005). Recognition of basic emotions from speech prosody as a function of language and sex. *Abstracts of the Psychonomic Society 46th Annual Meeting*, Toronto, Ontario.

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*(6), 1329–1333.

Pritcharsd, W. S., Shappell, S. A., & Brandt, M. E. (1991). Psychophysiology of N200/N400: A review and classification scheme. *Advances in Psychophysiology, 4*, 43–106.

Schapkin, S. A., Gusev, A. N., & Kuhl, J. (2000). Categorization of unilaterally presented emotional words: An ERP analysis. *Acta Neurobiologiae Experimentalis, 60*, 17–28.

Schirmer, A., & Kotz, S. A. (2006). Sex differences in the processing of verbal and nonverbal emotional expressions. *Trends in Cognitive Sciences, 10*, 24–30.

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research, 14*, 228–233.

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2005). On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research, 24*(3), 442–452.

Schupp, H. T., Junghöfer, M., Weike, A. I., & Hamm, A. O. (2004). The selective processing of briefly presented affective pictures: An ERP analysis. *Psychophysiology, 41*, 441–449.

Van Lancker Sidtis, D., Pachana, N., Cummings, J. L., & Sidtis, J. J. (2006). Dysprosodic speech following basal ganglia insult: Toward a conceptual framework for the study of the cerebral representation of prosody. *Brain and Language, 97*, 135–153.

Van Petten, C., & Kutas, M. (1988). The use of event-related potentials in the study of brain asymmetries. *International Journal of Neuroscience, 39*, 91–99.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning Memory and Cognition, 25*, 394–417.

Vingerhoets, G., Berckmoes, C., & Stroobant, N. (2003). Cerebral hemodynamics discrimination of prosodic and semantic emotion in speech studied by transcranial Doppler ultrasonography. *Neuropsychology, 17*, 93–99.

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Science, 9*, 585–594.

Wambacq, I. J., & Jerger, J. F. (2004). Processing of affective prosody and lexical–semantics in spoken utterances as differentiated by event-related potentials. *Cognitive Brain Research, 20*, 427–437.