Temporal Interaction of Emotional Prosody and Emotional Semantics: Evidence from ERPs

Silke Paulmann & Sonja A. Kotz

Max Planck Institute for Human Cognitive and Brain Sciences
Leipzig, Germany
paulmann@cbs.mpg.de

Abstract

Emotional prosody carries information about the inner state of a speaker and therefore helps us to understand how other people feel. However, emotions are also transferred verbally. In order to further substantiate the underlying mechanisms of emotional prosodic processing we investigated the interaction of both emotional prosody and emotional semantics with event-related brain potentials (ERPs) utilizing a prosodic and interactive (prosodic/semantic) violation paradigm. Results suggest that the time-course of emotional prosodic processing and emotional semantics differ. While a pure violation of a prosodic contour elicited a positivity between 450 ms and 600 ms, a violation of both emotional prosody and semantics elicited a negativity between 500 ms and 650 ms. These results suggest that emotional prosody and emotional semantics follow a different time-course. This holds true for all emotional prosodies (anger, disgust, fear, happy, pleasant surprise, sad) investigated. As the two conditions elicited two different electrophysiological components, the obtained results suggest that emotional prosody and semantics contribute differentially during the interaction of both information types. Furthermore, the data suggest that semantic information can override prosody when the two channels interact in time, that is, when the emotional prosodic contour agrees with the semantic content of a sentence.

1. Introduction

Disorders of emotional prosody -be it at the level of perception or production- influence social interactions. However, the underlying mechanisms of emotional prosody perception are not well understood. The present work investigates the perception of emotional prosody, with a particular emphasis on the potential relatedness between emotional prosody and emotional-semantics. The main question pursued is in how far the temporal integration of emotional semantics and emotional prosody can be specified, i.e. at which time do these two channels maximally interact? So far, little is known on how and when emotional prosody and emotional-semantics interact at the sentence level. However, there is previous evidence [1, 2, 3] which suggests that the time course of emotional prosody and semantics differ. The aim of the current ERP-experiment was to further investigate the interaction of emotional prosody and emotional-semantics by isolating the respective contribution of each emotional channel to this interaction by using a cross-splicing procedure. With the help of the cross-splicing method we violated an expectation specific to emotional prosody. Here we used pseudosentences as well as lexical sentences. Comparable to filtered speech, morphologically marked pseudosentences spoken with the varying emotional intonation patterns allow to eliminate lexical content while preserving emotional prosody.

We know that linguistic information is transferred not only through speech sounds but also via properties of the speaker, i.e. speaker identity (e.g. female/male; young/old) as well as the speaker's emotional state (e.g. happy/sad). Information about the speaker is typically thought to be encoded by pitch, intensity, and duration (or tempo) of the utterance. To explore the possible influence of speaker identity on the interaction of emotional prosody and emotional semantics, stimuli recorded from two speakers (male/female) were tested. In particular, we investigated voice identity to shed more light on the issue of gender voice specific emotional prosody processing. For instance, within the literature on emotional *expression* sex differences have been observed, i.e. female and male differ in their emotional expressiveness. Interestingly, these differences were noticed to arise already during development [4]. It is thus also important to further specify if there are also gender voice specific differences in emotional prosody *perception* [5].

2. Methods

2.1. Participants

Thirty native speakers of German (fifteen female) participated in the experiment. Female participants had a mean age of 24.13 (SD 1.96) and male participants had a mean age of 24.67 (SD 2.02). Participants received financial compensation for their participation.

2.2. Stimulus Material

The material consisted of semantically and prosodically matching stimuli (lexical/pseudo) for each of the six basic emotions (anger, fear, disgust, happiness, pleasant surprise, sadness) and a neutral baseline. For each emotion and sentence type, 30 sentences were presented, adding up to 210 matching lexical sentences and 210 matching pseudosentences. In addition, the same sentences were cross-spliced in two ways: a) in the combined semantic/prosodic violaten condition, a semantically and prosodically neutral start of a sentence was cross-spliced to an emotional semantically and prosodically matching end of the sentence, and b) in the pure emotional prosodic violation condition, a prosodically neutral start of a pseudosentence was cross-spliced to an emotional-prosodically end of the pseudosentence, resulting in 180 cross-spliced lexical sentences and 180 cross-spliced pseudosentences. As all sentences were spoken by a female and male speaker, a total of 1560 trials were presented in two sessions. Emotional prosodic valence was obtained in two earlier rating studies (one for the lexical sentences and one for the pseudosentences). All sentences were taped with a videocamcorder and later digitized at 16-bit/44.1 kHz sampling rate. The stimulus material was prosodically analyzed (i.e. pitch, intensity and duration of the sentences were extracted) using *Praat*. Results revealed comparable acoustical parameters across the two sentence types.

2.3. Procedure

Each subject was seated comfortably at a distance of 115 cm from a computer monitor in a sound-attenuating room equipped with a three-button response panel. Half of the subjects pressed the yes-button with their right hand and the no-button with their left hand. The sentences were presented via loudspeaker. Instructions with examples asked subjects to listen to the presented sentence, read a following word (flashed on the screen for 300 ms) and to verify a word or pseudoword probe as accurately and as quickly as possible. Participants were asked to avoid eye movements during sentence presentation. The intertrial interval was 2000 ms.

2.4. ERP Recording and Data Analysis

The electroencephalogram (EEG) was recorded with 59 Ag-AgCl electrodes mounted in an elastic cap according to the 10-20 system each referred to the nose (NZ). Bipolar horizontal and vertical EOGs were recorded for artifact rejection purposes. Electrode resistance was kept under 5k Ω. Data was rereferenced offline to linked mastoids. The 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, but all statistical analyses were computed on non-filtered data.

ERP components of interest were determined by visual inspection. For statistical analysis electrodes were grouped into four *Scalp Regions of Interest*. Each following *SROI* defined a critical region of four scalp sites: left frontal (LF): FP1 AF7 AF3 F9 F7 F5 F3 FT9 FT7 FC5 FC3; right frontal (RF): FP2 AF4 AF8 F4 F6 F8 F10 FT10 FT8 FC6 FC4; left posterior (LP): TP9 TP7 CP5 CP3 P9 P7 P5 P3 PO7 PO3 O1; and right posterior (RP): CP4 CP6 CP6 TP8 TP10 P4 P6 P8 P10 PO4 PO8 O2. To keep the number of electrodes constant for each *SROI*, midline electrodes were excluded from the analysis. The null-hypothesis was rejected for *p*-values smaller than 0.05. The Geissser-Greenhouse correction was applied to all repeated measures with greater than one degree of freedom in the numerator. The p-values for break-down comparisons were corrected using a modified Bonferroni procedure. Due to space limitations only relevant results will be reported.

3. Results

For the ERP analysis, separate ANOVAs for the two different violation types were conducted in two time windows (pure violation: 450-600 ms & combined violation: 500-650 ms). The two different conditions (combined semantically/prosodically cross-spliced material vs. prosodically cross-spliced material) were calculated with separate ANOVAs, treating *M* (Match: prosodically and semantically matching stimuli vs. Mismatch or spliced stimuli), *P* (emotional prosodies of anger, disgust, fear, happiness, neutral, pleasant surprise, sadness), and *Speaker* (female vs. male voice) as repeated-measures factors and *Sex* (female/male) as a between-subject factor. To control for possible session effects, we also included the factor *Session* (first vs. second) in the ERP analysis. In addition to the factors listed above, the factors *HEMI* (Left vs. Right Hemisphere) and *REG* (Frontal vs. Parietal Region) were included.

3.1. ERP results

450 ms to 600 ms: **Emotional Prosodic Violation:** In the time window of 450 ms to 600 ms for the emotional prosodic violation, a significant effect of *Speaker* was found (F(1,28)=10.01, $p<.01$), with a less positive-going waveform for the male speaker than for the female speaker. Also, the critical main effect for *P* was significant (F(6,168)=2.27, $p<.05$), indicating more positive-going ERP waveforms for the prosodically cross-spliced sentences than for the unspliced sentences. Post-hoc comparisons are listed in the following: 1) neutral vs. angry sentences (*P* effect: F(1,28)=8.04, $p<.01$); 2) neutral vs. disgust sentences (*P* effect: F(1,28)=4.73, $p<.05$); 3) neutral vs. fearful sentences (*P* effect: F(1,28)=10.27, $p<.01$); 4) neutral vs. happy sentences (*P* effect: F(1,28)=8.3, $p<.01$); 5) and neutral vs. pleasant surprise sentences (*P* effect: F(1,28)=5.54, $p<.05$); with all comparisons showing more positive ERP waveforms for the violated (cross-spliced) prosodic expectancy than for the unviolated sentences.

Taken together, the results revealed a significant speaker effect indicating a more positive going ERP amplitude for the female speaker than for the male speaker. Also, and critical for our research question, the results revealed a positive-going ERP component for all emotional prosodically cross-spliced sentences in comparison to unviolated sentences, except for the emotional category of sadness. This emotional prosodic violation effect was not qualified by *Speaker* or by SROI.

500 ms to 650 ms: **Combined Semantic/Prosodic Violation:** In the time window of 500 ms to 650 ms, the ERP analysis revealed a main effect for *Speaker* (F(1,28)=13.90, $p<.001$), with amplitudes generally being more negative for the male speaker. A critical main effect of *M* reached also significance (F(1,28)=4.29, p=.05), revealing a more negative-going component for the violated, i.e. cross-spliced sentences.

A critical interaction between *M* and *REG* (F(1,28)=4.27, $p<.05$) allowed for a step-down analysis by *REG*, This analysis revealed that the negativity for the violation was more pronounced at parietal electrodes (F(1,28)=10.77, $p<.01$).

Also, due to an interaction between *M* and *HEMI* and *REG* and *Sex* (F(1,28)=5.79, $p<.01$), step-down anlyses were carried out by *Sex* and then by *Sex*, *REG*, and *HEMI* which indicated that the negativity found for the violated sentences in male participants had a more frontally left distribution (F(1,14)=3.94, $p<.01$).

All in all, results show a negative-going ERP component for the combined prosodically and semantically violated sentences. It seems as if this negativity is more pronounced at parietal electrodes and for male participants this component is left frontally distributed.

4. Discussion

The current experiment substantiates ERP evidence on the underlying mechanisms of emotional prosody and emotional-semantics. The obtained ERP effects are comparable to previous results [1, 3] but extend these to pseudosentences for the purely prosodic effects. The fact that both types of incongruencies elicit varying brain responses points to the fact that the splicing procedure does not just reflect an acoustic artefact, but induces information specific incongruency responses in the ERP. These results clearly show that different brain responses

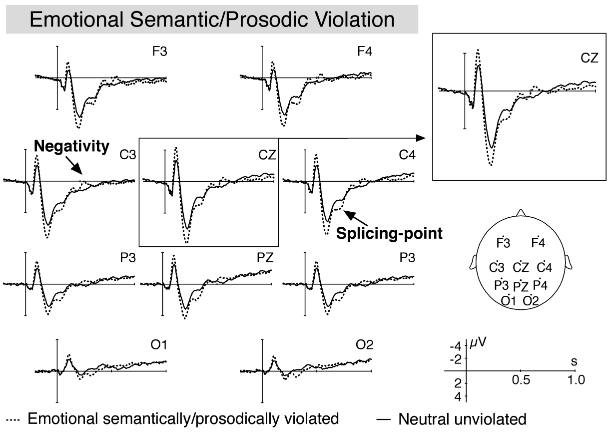

Figure 1: *This illustration shows ERPs elicited by emotional semantically and prosodically violated and unviolated sentences articulated by both speaker at selected electrode sites. Waveforms show the average for combined violated (black, dotted) and neutral unviolated (black) sentences from 200 ms prior to stimulus onset up to 1000 ms post-stimulus onset.*

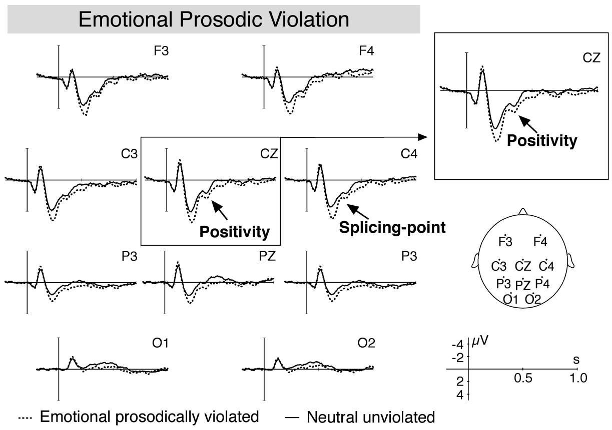

Figure 2: *This illustration shows ERPs elicited by emotional prosodically violated and unviolated pseudosentences articulated by both speaker at selected electrode sites. Waveforms show the average for emotional prosodically violated (black, dotted) and neutral unviolated (black) pseudosentences from 200 ms prior to stimulus onset up to 1000 ms post-stimulus onset.*

are elicited when a) emotional semantics and emotional prosody mismatch, or b) when emotional prosody is violated. In addition, speaker differences were observed, i.e. the ERP amplitude in response to sentences articulated by the female speaker were always more postive going than ERP amplitudes in response to sentences produced by the male speaker. In the following these effects will be discussed separately.

Emotional Prosodic Violation: As hypothesized, the violation of an emotional prosodic intonation contour, i.e. when a neutral start of a pseudosentence was cross-spliced to an emotional end of a pseudosentence, elicited a positivity between 450 and 600 ms after sentence onset. The obtained ERP effect replicates results from previous studies [1, 3] but extends these in several ways. First of all, the effect was elicited by pseudosentences, suggesting that the effect is truly elicited by an emotional prosodic violation and is not manipulated by semantic content of the cross-spliced sentence. Second, as this positivity was similar for all *emotional prosodic* contour violations this effect seems to be prosodic in nature.

To better understand this positivity and its underlying function it is helpful to take other language-related ERP components into account. For example, [2] have reported a P800 in response to linguistic prosody contour violations. In a similar cross-splicing study, the authors manipulated the intonation contour of statements and questions. They suggested that the P800 found was closely linked to F0 contour violations. A similar conclusion can be drawn from the present study. However, the P800 in the Astésano et al. (2004) experiment was elicited only when task instructions focussed on prosodic processing. Here, we report a positivity under implicit task instruction. Additionally, the onset of the two components differed, suggesting that the two components may be related but may not reflect the same process(es).

A second positive ERP component that has been linked to prosodic processing is the Closure Positive Shift (CPS) [6]. However, whereas the CPS has been shown to be elicited by prosodic phrase boundaries, i.e. during/after segmentation processes of long and syntactically complex sentences, the current material did not require linguistic phrase segmentation due to its syntactically simple construction. Thus, it seems unlikely that the current positivity and the CPS reflect the same language process.

The third well-known positive component, namely the P600, has been argued to reflect syntactic reanalysis evidence, but also more general reanalyses and integration processes of linguistic and non-linguistic information [6, 7, 8, 9]. However, an important difference between the P600 and the component reported here is that the P600 is always elicited by structural errors, whereas the current positive ERP component was elicited by violating emotional prosodic expectancy. This could still imply a function that reflects reanalysis of an expected context continuation. Following this line of argumentation, the positivity elicited by emotional prosodic intonation contour violations could reflect reanalysis of emotional prosodic aspects of a stimulus based on an F0 manipulation (see [2]).

One last aspect should be considered. The positivity was elicited irrespective of valence. This leads to the assumption that a general prosodic contour violation might have elicited similar effects. Taking the P800 component observed by [2] into account, it seems reasonable to generalize that prosodic contour violations may always elicit a positive ERP component. It is thus suggested that several positive ERP components (e.g. P800, CPS) may belong to the same family of ERP components, all reflecting responses to prosodic violations, but differing in extent and distribution depending on the nature of the prosodic information at state. For instance, emotional prosodic contour violations may be responded to faster than linguistic prosodic contour violations due to the evolutionary significance of emotional stimuli. Indeed, the current result fits nicely to reports in the literature [10] that it may be evolutionary advantageous to rapidly detect unexpected emotional events. In sum, there is considerable evidence that suggests that emotional prosody processing is a highly automatized process that does not seem to be influenced by valence. It is therefore suggested that the positivity obtained in the current experiment may be closely related to other positivities elicited in prosodic content. However, due to the reasons elaborated above, different task instructions and different linguistic aspects may affect the time course of these positivities.

Combined Emotional Prosodic and Semantic Violation: The present study aimed to further specify the interaction between emotional prosody and emotional-semantics at the sentence level, and to underline that this interaction is probably not dependent on emotional valence, i.e. there is no processing difference between violations of emotional prosodies belonging to different emotional valences. Indeed, a parietally distributed negative-going ERP component was elicited due to the double violation. The current results suggest that responses to combined emotional prosodic and emotional-semantic violations are valence-independent and thereby replicate previous results [3]. The fact that the two experimental conditions tested here elicited two different ERP responses suggests that emotional semantics (negativity) can rule over emotional prosody (positivity). The current approach materializes the contribution of verbal and prosodic emotional information.

We propose that the negative ERP component elicited by combined incongruencies reflects semantically driven incongruency detection and is comparable to the well known N400, as it has been shown that the N400 is larger for semantically incongruent items than for congruent ones. The current result implies that in the combined violation condition, emotional semantics reigns over emotional prosody. However, it remains speculative to suggest that the brain response was triggered by the emotional semantic violation alone, since it was elicited by a double violation and did not occur in isolation. The current results go hand in hand with results suggesting that semantics cannot be ignored even if not attended to [11].

Last, it could be argued that the task utilized here may have put the spotlight on to the semantic content. Therefore, it could be assumed that semantic processing predominates emotional prosody processing as a function of task. However, earlier studies suggest that this is unlikely. Similar negative ERP components in response to a combined emotional semantic and prosodic violation were been obtained under implicit and explicit task instructions [3].

Speaker effects: A last aim of the current study was to investigate to which extent speaker identity influences emotional prosody processing. The female voice elicited larger positive-going ERPs than the male voice. This suggests that the speaker difference may be driven by evolutionary factors [12]. In particular, female voices elicited stronger ERP effects comparable to observations from functional imaging where stronger activation patterns were found for female voices than for male voices [12]. It is assumed that the high-pitch female voices may be perceived as socially and biologically more salient than a low-pitch male voice. Interestingly, speaker identity does not seem to influence the processing of an emotional prosodic or combined emotional prosodic and semantic violation, i.e. the positivity and negativity observed in the current study were found irrespective of speaker gender. This suggests detection of an incongruency independent of speaker voice.

5. Conclusions

To summarize, the present findings are highly comparable to effects observed in previous studies [3]. Here, results were extended by introducing pseudosentences carrying different emotional prosodies, allowing to investigate emotional prosody in isolation.

Finally, it is believed that the current positive ERP component elicited by pure emotional prosodic violations is closely linked to F0 and intensity violations and is proposed to reflect re-analysis processes of F0 and intensity violations. Whether this positivity is prosody specific or comparable to the positivity elicited by syntactic violations remains a matter of debate. In contrast, the negative ERP component elicited by combined violations is assumed to reflect prosodic and semantic integration problems of sentential incongruencies. It is further suggested that this negativity is also influenced by emotional prosodic violation. Thus, it is concluded that the influence of emotional prosody enhances the propositional intent of an utterance, whether in semantic-prosodic congruent or incongruent presentation and this irrespective of speaker gender.

6. References

[1] Kotz, S.A. and Alter, K. and Besson, M. and Schirmer, A. and Friederici, A.D., 2000. The interface between prosodic and semantic processes: An ERP study. *Journal of Cognitive Neuroscience, Suppl.*, 83, 123.

[2] Astésano, C. and Besson, M. and Alter, K, 2004. Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research*, 18, 172-84.

[3] Paulmann, S. and Kotz, S.A., 2005. When Emotional Prosody and Semantics interact in Time: *Journal of Cognitive Neuroscience, Suppl.*, 63.

[4] Salminen, J.K. and Saarijärvi, S. and Äärelä, E. and Toikka, T. and Kauhanen, J. 1999. Prevalence of alexithymia and its association with sociodemographic variables in the general population of Finland. *J Psychosom Res*, 46, 75-82.

[5] Schirmer, A. and Kotz, S.A. and Friederici, A.D., 2002. Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, 14, 228-233.

[6] Steinhauer, K. and Friederici, A.D., 2001. Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, 20, 267-295.

[7] Friederici, A.D., 1998. The neurobiology of language comprehension. In: *Language Comprehension: A Biological Approach.*, pp. 263-301 Springer: New York.

[8] Hagoort, P. and Brown, C. and Groothusen, J., 1993. The syntactic positive shift as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439-484.

[9] Osterhout, L. and Holcomb, P.J., 1992. Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785-806.

[10] Vuilleumier, P., 2005. How brains beware: neural mechanisms of emotional attention. *Trends Cogn Sci*, 9, 585-594.

[11] Besson, M. and Magne, C. and Schön, D., 2002. Emotional prosody: sex differences in sensitivity to speech melody. *TRENDS in Cognitive Sciences*, 6, 405-407.

[12] Lattner, S. and Meyer, M.E. and Friederici, A.D., 2005. Voice Perception: Sex, Pitch, and the Right Hemisphere. *Human Brain Mapping*, 24, 11-20.