Neuropsychologia 50 (2012) 1609–1620

Contents lists available at SciVerse ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

It's special the way you say it: An ERP investigation on the temporal dynamics of two types of prosody

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

[a] Department of Psychology & Centre for Brain Science, University of Essex, Colchester, UK
[b] Minerva Research Group "Neurocognition of Rhythm in Communication", Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

ARTICLE INFO

Article history:
Received 13 September 2011
Received in revised form 7 March 2012
Accepted 12 March 2012
Available online 20 March 2012

Keywords:
Prosody
ERPs
Emotion
Linguistic
Prosodic expectancy positivity

ABSTRACT

Sentence prosody is long known to serve both linguistic functions (e.g. to differentiate between questions and statements) and emotional functions (e.g. to detect the emotional state of a speaker). These different functions of prosodic information need to be encoded rapidly during sentence comprehension to ensure successful speech communication. However, systematic investigations of the comparative nature of these two functions, i.e. are the two functions of prosody independent or interdependent, are sparse. The question at hand is whether the two prosodic functions engage a similar neural network and run a similar time-course or not. To this aim we investigated whether emotional and linguistic prosody are processed independently or dependently in an event-related brain potential (ERP) experiment. We merged a prosodically neutral head of a sentence to a second half of a sentence that differed in emotional and/or linguistic prosody. In a within-subjects design, two tasks were administered: in the "emotion task", participants judged whether the sentence that they had just heard was spoken in a neutral tone of voice or not (emotional task); in the "linguistic task", participants decided whether the sentence was a declarative sentence or not. As predicted, the previously reported prosodic expectancy positivity (PEP) was elicited by linguistic and emotional prosodic expectancy violations. However, the latency and distribution of the ERP component differed: whilst responses to emotional prosodic expectancy violations were elicited shortly after an expectancy violation (∼470 ms post splicing-point) and most prominently at posterior electrode-sites, the positivity in response to linguistic prosody had a later onset (∼620 ms post splicing-point) with a more frontal distribution. Interestingly, responses to combined (linguistic and emotional) expectancy violations resulted in a broadly distributed positivity with an onset of ∼170 ms post expectancy violation. These effects were found irrespective of the task setting. Given the differences in latency and distribution, we conclude that the processing of emotional and linguistic prosody relies at least partly on differing neural mechanisms and that emotional prosodic aspects of language are processed in a prioritized processing stream.

1. Introduction

Successful vocal communication requires to correctly use and interpret melodical and rhythmical properties of speech. Such speech modulation is achieved by varying psychoacoustic parameters such as fundamental frequency (F0), voice quality, loudness, speech rate, and rhythm, and it is subsumed under the term prosody. Prosody can serve various communicative functions. For instance, dependent on the rise and fall pattern of the voice, we can differentiate between different speech acts such as questions or declarative sentences. At the same time speech prosody can also be used to infer how others feel. For instance, when listening to a loud and high pitched voice, we are more likely to think the speaker is angry than sad. These different functions of prosodic information need to be decoded rapidly. However, the neural mechanisms underlying these functions are still not fully understood. Moreover, given that prosody can fulfil different communicative functions (sometimes simultaneously), the question arises whether these functions are independent or interdependent. In the current study, we looked at two functions of prosody, namely the distinction between different types of speech acts (from here on referred to as linguistic prosody) and the ability to extract information about the emotional state of the speaker (from here on referred to as emotional prosody). Specifically, we investigated the time-course underlying on-line processing of emotional and linguistic prosody and explored the comparative nature of these two functions using the event-related brain potential (ERP) method.

* Corresponding author at: University of Essex, Department of Psychology, Wivenhoe Park, Colchester, Essex CO1 2ZD, UK. Tel.: +44 01206 873422.
E-mail address: paulmann@essex.ac.uk (S. Paulmann).

0028-3932/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2012.03.014

Over the past decades, numerous attempts have been made to specify which brain regions are implicated in emotional and linguistic prosodic processing (for reviews see e.g. Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006; Wildgruber et al., 2004). Specifically, it has been asked whether the two prosodic functions are modulated by a shared or differentiated brain network. In an fMRI study, Wildgruber et al. (2004) reported right lateralized activation of the dorso-lateral frontal cortex, as well as bilateral activation of thalamic and temporal regions for both emotional and linguistic prosodic processing, thus suggesting a shared neural network; however, when participants had to focus on linguistic aspects of a stimulus, the authors reported activation of the left inferior frontal gyrus as opposed to activation of bilateral orbito-frontal areas when the task focus was on emotional stimulus aspects. This suggests that the two functions are subserved by different brain regions when emotional or linguistic properties of speech have to be evaluated, respectively. The imaging results supporting lateralization of prosody processing go well with earlier neuropsychological data gathered from left- and right-hemisphere-damaged patients. These data suggest that lateralization for prosody processing is driven by different acoustic cues in different hemispheres: It was reported that the two patient groups used acoustic cues differently when judging emotional prosody (Van Lancker & Sidtis, 1992), i.e. pitch information was used by patients with an unaffected right hemisphere, whilst durational cues were used by patients with an intact left hemisphere. Data like these were taken to support the functional lateralization hypothesis of prosody (e.g. Van Lancker & Sidtis, 1992), i.e. depending on the task at hand and/or cue processed, prosody processing is oriented more strongly to either the left or right hemisphere.

The time-course of emotional and linguistic prosodic processing has also been addressed in some rare comparative studies (e.g. Pihan, 2006; Pihan, Tabert, Assuras, & Borod, 2008). As for emotional prosodic processing, research has revealed that it is a highly rapid process that does not require the listener to pay explicit attention to the emotionality of a sentence. Specifically, it has been shown that neutral sentences can be distinguished from emotional sentences within 200 ms after sentence onset (Paulmann & Kotz, 2008a). In addition, it has also been shown that the detection of emotional category changes in an auditory input stream (e.g. happy vs. neutral, sad vs. neutral sounds) elicit a mismatch negativity (MMN) in a similar time-window (e.g. Schirmer, Striano, & Friederici, 2005; Thoennessen et al., 2010). Moreover, we have previously provided evidence that suggests rapid detection of emotional prosodic expectancy violations, i.e. detection of abrupt emotional speaker tone change, as indicated by the so-called prosodic expectancy positivity (PEP), a positive-going ERP component with an onset of ∼350 ms post violation onset (e.g. Kotz & Paulmann, 2007; Paulmann & Kotz, 2008b; Paulmann, Pell, & Kotz, 2008).

Speech comprehension is linked to regularly initiating expectancies about upcoming events (e.g. which word, sound, stress pattern, etc. is likely to come next), a process that can ensure efficient processing by narrowing down lexical-search, guiding syntactic parsing, and/or quickly reacting to emotional events (c.f., for example, Winkler, Denham, & Nelken, 2009). For emotional prosodic expectancy violations, we have argued that the quick detection of deviance is necessary to assign emotional significance to an utterance in order to respond appropriately (e.g. fight/flight). In particular, we have proposed that emotional prosodic deviance is detected with the help of specific acoustic cues that can carry emotional meaning such as pitch, voice quality, and intensity. The PEP has been reported in response to German sentences with a violation of emotional prosody irrespective of speaker voice, emotional category, or task and for both semantically neutral and pseudo-sentences (Kotz & Paulmann, 2007; Paulmann & Kotz, 2008b). Recently, the PEP has also been reported for emotional prosodic expectancy violations in a tonal language such as Mandarin (Chen, Zhao, Jiang, & Yang, 2011).

Linguistic prosodic processing has also been explored using ERPs. For instance, Li, Chen, and Yang (2011) recently reported a fronto-centrally distributed negative ERP response with a latency of 270–510 ms when prosodic prominence was manipulated in question–answer dialogues, whilst a broader and longer negative ERP effect (270–510 ms and 510–660 ms) was elicited in response to prosodic boundary violations. In addition, Steinhauer, Alter, and Friederici (1999) first reported that the marker of prosodic phrasing/structuring, as indicated by the closure positive shift (CPS), is elicited quickly after prosodic phrase boundaries. The CPS is most visible at midline electrodes and is usually long lasting (∼500 ms, sometimes even longer, see Steinhauer et al., 1999). Given its rapid elicitation, it has been argued that the CPS could be elicited by "pre-final constituent lengthening" (Steinhauer, 2003). A different positive ERP component with a slightly later onset latency, the so-called P800, has been reported in linguistic prosodic deviance detection (Astésano, Besson, & Alter, 2004). For instance, Astésano and colleagues merged declarative sentence beginnings with endings from questions and vice versa. The P800 is elicited in response to these linguistic prosodic expectancy violations and is argued to reflect prosodic re-analyses processes (Astésano et al., 2004) comparable to the PEP elicited in response to emotional prosodic expectancy violations. Similarly, Eckstein and Friederici (2005) report a late positivity in response to prosodic incongruity of a sentence final word (e.g. the prosodic pattern signalled another upcoming word instead of the end of sentence). Interestingly, in the study by Astésano et al. (2004), the P800 was reported to be task-dependent, i.e. it only emerged when prosody was in task focus, whilst the late positivity found by Eckstein and Friederici (2005) was reported even though participants did not explicitly have to focus on prosodic aspects of stimuli. This finding has prompted Eckstein and Friederici (2005) to interpret their elicited positivity as a delayed P600; however, it could well be that the authors report a similar prosodic expectancy positivity and that ERP differences related to task focus (explicit vs. implicit) stem from differences in stimulus materials between the different studies, such as prominence of deviations. In fact, we have previously suggested that the later onset and longer duration of the P800 when compared to the PEP may be due to the fact that detecting linguistic prosodic violations is not of primary evolutionary significance. Alternatively, linguistic prosodic violations may be more difficult to detect as the acoustic deviations are less salient. Hence, it could be that the two components may belong to the same family of positive ERP components linked to prosody processing (Kotz & Paulmann, 2007; Paulmann & Kotz, 2008b). The suggestion that saliency is a primary force in observed latency differences between positivities is supported by findings of Marques, Moreno, Castro, and Besson (2007). The authors asked musicians and non-musicians to detect pitch violations, i.e. participants had to decide whether sentence final words sounded normal or not (pitch manipulated). Not only did musicians show better behavioural performance in the experiment, but also they showed an earlier latency onset of a positive ERP component in response to deviant stimuli as opposed to non-deviant stimuli. Whilst in the non-musician group, a positivity was reported for the time-span of 600–800 ms after the onset of the mismatching endings, a positivity with a similar distribution was elicited 400 ms earlier for musicians. As the saliency of deviance affects neural mechanisms underlying expectancy violation processing, it can be argued that emotional stimuli will always be more salient than neutral stimuli (due to specific acoustic cue configurations which are recognized across cultures, see Pell, Paulmann, Dara, Alasseri, & Kotz, 2009), which should lead to an earlier onset of the PEP.

Still, systematic investigations of the comparative nature of these two functions are still rare. However, in a direct current (DC) potential study, Pihan et al. (2008) explored the discrimination of linguistic prosody in emotional (happy/fearful) as well as neutral sentences. The authors reported fewer errors in the discrimination of statements and questions in neutral as opposed to emotional sentences. They suggest this effect to be related to increased pitch variability in emotional as opposed to neutral sentences. According to the authors, explicit processing of pitch variability was not task relevant, whilst processing of pitch direction (fall or rise) was. Thus, they argue that increased pitch variability in emotional sentences as opposed to neutral sentences could have interfered with the perception of pitch direction. This interference would then lead to reduced accuracy in the discrimination of statements and questions in emotional sentences. Next to differences in behavioural data, Pihan and colleagues also report an increase in cortical activation for happy stimuli when compared to neutral, though no such difference was found for the comparison of stimuli expressing fear and neutral stimuli. Interestingly, the authors report bilateral cortical activation with a maximum over anterior electrode-sites. They interpret the bilateral cortical activation in the EEG data to complement the behavioural data: interference between pitch variability (emotion relevant) and pitch direction (linguistically relevant) could lead to emotion and linguistically relevant processing mechanisms even in a linguistically oriented task. Why this effect would particularly manifest when processing happy (and not fearful) stimuli awaits further clarification; however, one assumption could be that pitch variability was bigger in happy as opposed to fearful stimuli (Pihan et al., 2008).

Whilst these investigations suggest that emotional and linguistic prosodic processing are subserved by shared as well as differentiated neural sources, they do not allow to directly compare the time-course of processing these two functions. However, to arrive at a more precise model of language comprehension that takes into account the role of *both emotional and linguistic* prosody during speech processing, it is vital to shed more light on the time-course and processing nature of the different functions of prosody both individually and in an interactive manner. This is best achieved by studying the functional properties of prosody whilst they unfold in time. As previously demonstrated, cross-splicing is one way to do this as it allows violating prosodic expectancy (emotional or linguistic) in a temporally and acoustically controlled manner. Specifically, cross-splicing creates prosodic transitions that are similar across functions (emotional/linguistic prosody) and at the same time allows to temporally synchronize the interaction between these functions (see Kotz & Paulmann, 2007 for further details). Moreover, as opposed to other deviance-related experimental manipulations, an abrupt change of prosodic contour is arguably not too unnatural. Consider, for example, the following discourse:

X: "Yesterday, the property management asked your neighbour Tom to move out and today he is nowhere to be seen".

a) Y: "He has left the building and moved on".
b) Y: "He has left the building and moved on?"
c) Y: "He has left the building and moved on?!"

Version (a) could be intoned with either a neutral or an angry tone of voice. A questioning voice with a rise in pitch could be used to intone version (b), and a shocked and at the same time questioning high pitched, loud tone of voice with a pitch rise towards the end could be used to express version (c). Now assume that it took Y a short while to let X's statement sink in. In this case, it would not be unnatural for Y to start out his reply in a neutral tone of voice indicating that he is simply answering with a statement as well, but as soon as he realizes what this message actually means (e.g. that Tom has left without returning the Ipod that Y lent him), he could suddenly switch to version (c), creating a transition from a neutral statement to an angry question. Hence, it can be argued that dynamic and often sudden deviations in prosody can occur in daily-life (Paulmann & Kotz, 2008b) and it seems crucial for any neural model of speech comprehension to take into account the role prosody plays in dynamic speech comprehension.

The aim of the current study was to further define the time-course underlying emotional and linguistic prosodic processing and to investigate whether the two functions of prosody are in- or interdependent. To explore this, we created stimuli that followed the example elaborated above. We chose to use angry prosody to represent emotional prosody as our own findings suggest that the PEP is elicited irrespective of valence and emotional category (e.g. Paulmann & Kotz, 2008b). Moreover, this allows us to compare our findings directly to findings from Chen et al.'s (2011) and Kotz and Paulmann's (2007) expectancy violation studies as they also used angry prosody to represent emotional prosody. Based on imaging data (Wildgruber et al., 2006, 2004), it can be hypothesized that processing differences between emotional and linguistic prosody may be found in the spatial domain with differing task demands (focus on emotion or on linguistic aspect of the sentence). Although ERPs have a low spatial resolution, they nevertheless offer the opportunity to describe gross pattern changes: the underlying assumption is that if ERP effects differ in their scalp distributions, the generating neural source is likely to be at least partially different. Based on neurophysiological evidence (e.g. Astésano et al., 2004; Chen et al., 2011; Kotz & Paulmann, 2007; Paulmann & Kotz, 2008b), we further hypothesize that emotional and linguistic prosodic deviances will elicit positive ERP components, which should differ in their onset latency if true that the PEP latency is dependent on cue saliency. Although an abrupt transition will be signalled by similar acoustic properties (e.g. pitch change) across the two conditions, it is assumed that emotional prosodic deviance can be detected more quickly, as specific acoustic parameters (e.g. parameters related to voice quality) may signal their emotional relevance in addition to changes in pitch contours. This rapid detection of some over other acoustical properties could result from prioritized processing of emotional stimuli (e.g. Vuilleumier, 2005). Alternatively, differences in latencies of previously reported positivities may result from varying experimental factors such as stimuli, speaker type, and language of investigation. If this was indeed the case, the present investigation should fail to find differences across conditions. Finally, we also test whether similar ERP components are elicited when both emotional and linguistic prosodic expectation is violated. Previous evidence (e.g. Astésano et al., 2004; Kotz & Paulmann, 2007) that tested other combined deviance detection (e.g. prosody and semantics) has reported an earlier onset of deviance-related ERP components. Latter observation is in line with the assumption that language processing is facilitated when provided in multiple channels (e.g. voice, semantics; Paulmann, Jessen, & Kotz, 2009). The question remains whether a similar pattern can be found when different information is provided through the same channel. This would suggest that the amount of information available can influence processing speed rather than the amount of channels the information is provided in. If true, we can hypothesize an earlier onset of ERP components in response to double violations.

In short, first, we hypothesize that both emotional and linguistics prosodic expectancy violations will elicit a PEP-like effect. The PEP for emotional prosodic expectancy violations should have an earlier onset than the positivity elicited for linguistic prosodic expectancy violations given stronger cue saliency information. Double violations should elicit an even earlier onset if true that the PEP is dependent on cue saliency strength. Finally, although the PEP has been reported to be elicited under different task instructions (Kotz & Paulmann, 2007), task demand differences could manifest in spatial distribution differences (c.f. Wildgruber et al., 2006,

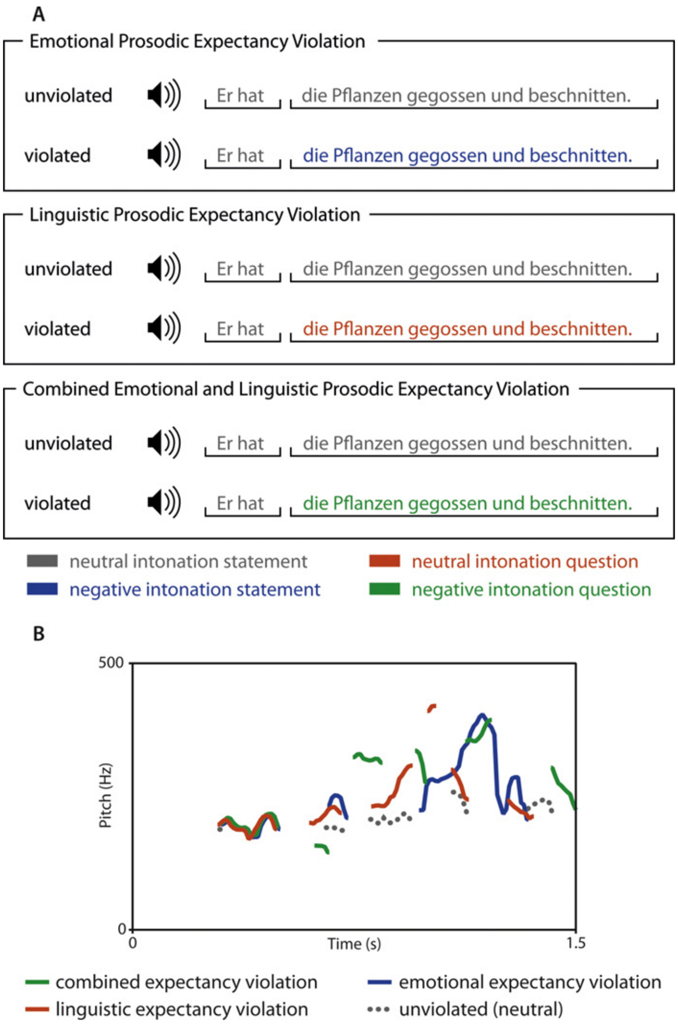

Fig. 1. (A and B) Part A of the illustration explains the splicing procedure for emotional prosodic, linguistic prosodic, and combined prosodic expectancy violations. The sentence context is always neutral (literal translation: "She has watered the plants and cut [them]"). Part B of the illustration shows example pitch contours for prosodically violated (red, blue, green lines) sentences in comparison to unviolated neutral sentences (grey; upper box). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

2004). However, given the low spatial resolution of ERPs, the exact distribution differences could not be predicted with certainty.

2. Materials and methods

2.1. Stimulus material

We created 50 syntactically identical and semantically neutral sentences as base stimuli (e.g. "Er hat die Buecher gelesen und verstanden" [*He has read and understood the books*]. Each of these sentences was spoken in different intonations to create different experimental conditions: (1) spoken in a neutral prosody, intoned as a declarative sentence; (2) spoken in an angry prosody, intoned as a declarative sentence; (3) spoken in a neutral prosody intoned as a question; (4) spoken in an angry prosody intoned as a question. All sentences were spoken by a German speaking actress and digitally recorded in a sound-attenuated booth using a high quality microphone (16-bit, 44.1 kHz sampling rate). After the recording, we created cross-spliced versions of these sentences by merging a neutral head (declarative sentence, neutral prosody of a sentence to an end of sentence that differed in linguistic and/or emotional prosody; see Fig. 1A and B). In total, 150 different stimuli were produced this way: 50 sentences containing an emotional prosodic expectancy violation, 50 sentences containing a linguistic prosodic expectancy violation, and 50 sentences containing a double violation of linguistic and emotional prosodic expectancy. The average duration of sentences was approx. 3 s. The mean splicing point as calculated from the mean duration of the two possible beginnings that were used as templates was 380 ms after the sentence onset (345 ms for "Er hat" and 417 ms for "Sie hat").[1] In addition to the 50 base stimuli and 150 cross-spliced stimuli, 300 sentences (50 spliced, 250 unspliced; neutral and emotional) were used as fillers resulting in a total of 500 sentences (200 spliced/300 unspliced). Acoustical analyses for sentence parts post-splicing point can be found in Table 1a and b.

2.2. Participants

20 healthy volunteers (10 male) with no reported hearing impairment and normal or corrected to normal vision participated in the experiment. All participants were right-handed, native speakers of German, with a mean age of 24 years (SD = 2.9). They were paid 7€/h for their participation. This study was approved by the ethics committee of the Max Planck Institute for Human Cognitive and Brain Sciences.

2.3. Procedure

Each participant was tested individually in a sound-attenuated room, seated comfortably at a distance of 115 cm from a computer monitor with a response-button panel placed in their lap. Stimuli were presented via loudspeakers placed at both sides of the monitor. The Experimental Run Time System (ERTS; Beringer, 1993) was used to run the experiment. All 500 stimuli were pseudo-randomized and split up into 10 blocks containing 50 sentences each. After each block, participants were given a short self-determined break to relax. Participants were given two different tasks one of which had to be performed after each presentation of a stimulus. For half of the blocks, participants judged whether the sentence that they had just heard sounded prosodically neutral or not (emotional task), for the other half, they judged whether the sentence was a declarative sentence or not (linguistic task). The different tasks where given in an alternating order. Participants responded by pressing the left or the right button on a three-button response panel. Half of the participants pressed the left button for judging the sentence as emotional or as a declarative sentence, respectively, and the right button for judging a sentence as neutral or as a question. The other half followed a reversed button assignment. An experimental trial was as follows: A white fixation cross that helped participants to avoid unnecessary eye movements was presented in the middle of a black screen during sentence presentation. At the offset of the sentence presentation, this fixation cross remained on screen for 1000 ms until it was replaced by a question mark, which was presented for 1500 ms. The question mark prompted participants to respond yes/no (i.e. was the utterance intoned as a statement/was the utterance spoken in an emotional tone of voice) on a response panel before an inter-trial interval of 2000 ms followed before the next trial began.

2.3.1. EEG recording procedure

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, 1991), each referenced on-line to the nose and re-referenced offline to the averaged mastoids. Bipolar horizontal and vertical EOGs were recorded for artefact rejection purposes. Trials contaminated by EOG or movement artefacts were rejected with a threshold of 30 μV for each participant and were omitted from the ERP analysis. In addition, individual data sets were scanned for extreme drifts and other artefacts not detected by the automatic rejection procedure. Electrode resistance was kept below 10 kΩ. EEG signals were recorded continuously with a band pass filter between DC and 250 Hz and were sampled at a rate of 500 Hz. For graphical display, ERPs were filtered off-line with a 7 Hz low pass filter. All statistical analyses were carried out on otherwise unfiltered data and were computed including correctly answered trials only. Data processing was carried out with the EEProbe software (MPI-CBS, Leipzig). On average, approx. 17% of correctly answered trials were rejected. This can be broken down for each individual condition as follows: 15% of neutral, 18% of emotional prosodic violated, 17.5% of linguistic prosodic violated, and 18% of combined prosodic violated sentences were rejected. Data were epoched starting 200 ms before stimulus onset and lasting for 1500 ms. After visual inspection, a time-line analysis was conducted to determine the different onset latencies of ERP components of interest. Based on this analysis, the following ERP time windows were included in the statistical analysis: 550–850 ms, 850–1000 ms, and 1000–1400 ms after sentence onset (i.e. ∼170, ∼470, and ∼620 ms, respectively, after cross-splicing point).[2]

[1] Previous cross-splicing evidence (e.g. Kotz & Paulmann, 2007) suggests that the first two syllables of a sentence provide enough information to build up a prosodic expectancy. In fact, ERP evidence exploring the time-course of emotional prosody suggests that neutral prosody can be distinguished from emotional prosody within 200 ms after sentence onset (e.g. Paulmann & Kotz, 2008a).

[2] Note that we follow previous research (e.g. Kotz & Paulmann, 2007; Paulmann & Kotz, 2008a, 2008b; Paulmann et al. 2008) in calculating ERPs epoched from sentence onset rather than epoched from cross-splicing point).

Table 1
a (top) and b (bottom). These tables list results from acoustical analyses comparing all presented conditions. Analyses were conducted for materials starting after splicing point (a) and for the first vowel of the definite article (b) occurring shortly after splicing point. Note: mean/range F0 = mean/range pitch; mean/range db = mean intensity as measured in dB; syll/s = speech rate as measured in syllables per second; F1 = first formant; jit_ddp = jitter (ddp); jit_local = jitter (local); HNR_mean = Harmonic to noise ratio (mean). All acoustical analyses were carried out with praat (Boersma & Weenink, 2009).

Condition	(a)Acoustical analysis post splicing-point				
	Mean F0	Range F0	Mean dB	Range dB	Syll/s
Combined	299.37	222.57	62.87	51.96	4.65
	(19.22)	(53.44)	(1.36)	(5.11)	(.79)
Emotional	232.87	209.20	62.89	50.97	4.79
	(14.84)	(63.25)	(1.54)	(5.16)	(.65)
Linguistic	241.43	275.23	61.36	45.99	4.79
	(22.09)	(63.33)	(1.34)	(2.95)	(.89)
Neutral	198.73	203.31	61.50	46.23	4.52
	(12.02)	(93.28)	(1.43)	(4.84)	(.47)

Condition	(b)Acoustical analysis for the first vowel post splicing-point				
	Mean F0	F1	jit_ddp	jit_loc	HNR_mean
Combined	295.04	580.47	.0053	.0051	12.29
	(50.09)	(193.84)	(.0041)	(.0026)	(3.27)
Emotional	235.34	494.63	.0108	.0098	14.10
	(27.3)	(180.46)	(.0171)	(.0099)	(3.13)
Linguistic	205.30	370.06	.0042	.0060	19.20
	(25.26)	(81.34)	(.0022)	(.0035)	(1.83)
Neutral	173.90	371.02	.0081	.0103	18.48
	(33.76)	(58.59)	(.0069)	(.0048)	(2.68)

2.3.2. Data analysis

For each time-window of interest, separate ERPs for each condition at each electrode site were averaged for each participant with a 200 ms pre-stimulus baseline. Mean ERPs were then analysed using the general linear model (GLM) with repeated measures by SAS (SAS 9.2). Mean amplitudes were calculated treating time-window (550–850 ms, 850–1000 ms, and 1000–1400 ms), condition (unviolated, linguistic prosodically spliced, emotional prosodically spliced, doubly-spliced) and task (emotional, linguistic) as repeated-measures factors. In addition, distributional differences were accounted for by the factors HEMI (left vs. right hemisphere) and REG (anterior and posterior region; for list of electrodes in each region see Fig. 2). Only significant interactions including the critical factor condition will be reported in the following.

3. Results

3.1. Behavioural findings

We only report accuracy rates as our paradigm (delayed button response) could not be used to accurately calculate reaction times. One participant was excluded from the behavioural analysis as he accidentally switched the answer buttons on the response panel for some part of the experiment (2 blocks in total). The analysis for accuracy rates revealed a main effect of task ($F(1,18) = 10.25$,

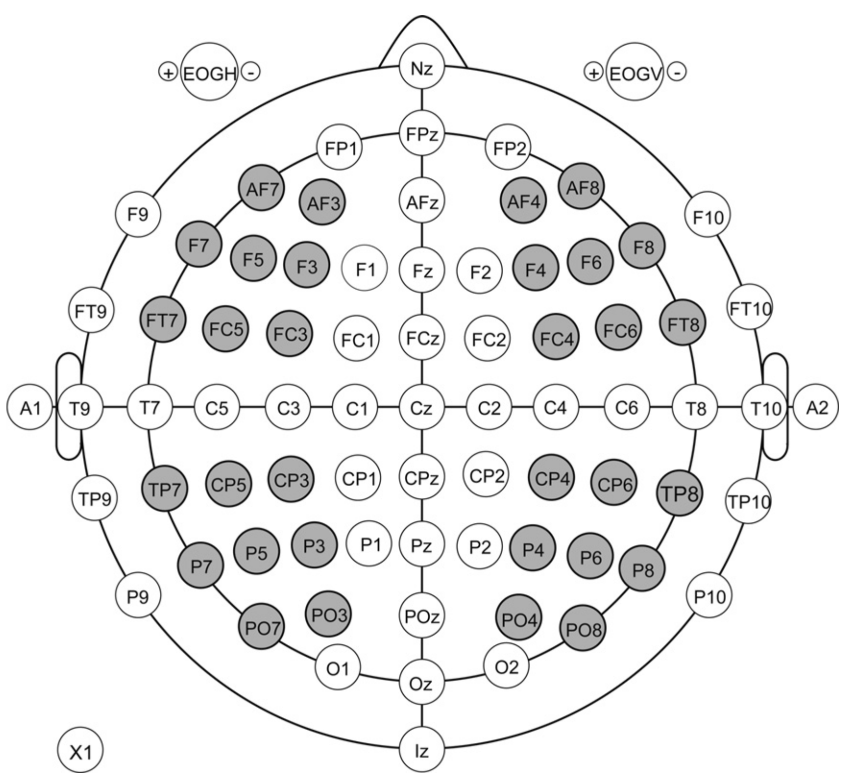

Fig. 2. The head shows how electrodes were grouped into different regions of interest (ROIs) for the statistical analysis.

$p < .01$) with participants responding more accurately during the linguistic evaluation task (92% vs. 84% correct). The main effect of condition was also significant ($F(3,54) = 3.67$, $p < .05$) revealing slightly higher accuracy rates for non-spliced (92%) as opposed to linguistic and emotionally spliced sentences (84% each). Responses to combined spliced materials (92%) were not found to be different to non-spliced sentences. Finally, the interaction between task × condition turned out to be significant ($F(3,54) = 6.89$, $p < .05$). Whilst no effects on accuracy rates of task were found for emotionally spliced and neutral sentences, results showed that the linguistic evaluation task was easier for participants ($F(1,18) = 14.06$, $p < .01$) when listening to combined spliced sentences (94% vs. 76%), whilst the opposite, i.e. better performance in the emotional evaluation task ($F(1,18) = 5.43$, $p < .05$) was true for the linguistic prosodic expectancy violation (75% vs. 90%).

Taken together, results for accuracy rates suggest that the difficulty to perform the task correctly depends on the sentence type one is listening to, i.e. in general, it does not seem as if one task was more difficult to perform than the other.

3.2. ERPs: prosodic expectancy positivity

The analysis of ERP mean amplitudes for all three time-windows revealed a significant condition effect ($F(3,57) = 7.25$, $p < .001$) suggesting different ERP mean amplitudes for the different conditions in all time-windows of interest. Post hoc contrasts confirmed more positive-going ERPs for doubly violated sentences when compared to neutral sentences ($F(1,19) = 15.05$, $p < .01$), but this effect was not confirmed for single violation conditions compared to neutral ($p > .19$). In addition, there was an interaction between condition × REG ($F(3,57) = 3.48$, $p < .05$). Post hoc contrasts revealed a significant difference between doubly violated sentences and neutral sentences at frontal ($F(1,19) = 9.16$, $p < .01$) and parietal ($F(1,19) = 14.59$, $p < .01$) electrode-sites. At parietal electrode sites, the contrast between emotionally prosodically violated and neutral sentences was also significant $F(1,19) = 4.37$, $p = .05$). In both instances, violations elicited a more positive-going ERP waveform than neutral sentences.

We also found significant interactions between time-window × condition × task ($F(6,114) = 2.34$, $p = .05$) and time-window × REG × condition × task ($F(6,114) = 2.28$, $p = .05$), however, follow-up analyses failed to support significant differences.

Critically, the three-way interaction of condition × time-window × REG ($F(6,114) = 11.40$, $p < .0001$) turned out to be significant, suggesting differently distributed condition effects in the three different time-windows. In the early time window (550–850 ms), post hoc contrasts confirmed a more positive-going ERP response to doubly violated sentences when compared to neutral sentences at frontal ($F(1,19) = 11.92$, $p < .01$) and parietal ($F(1,19) = 5.48$, $p < .05$) electrode-sites. In the time-window of 850–1000 ms, post hoc contrasts showed that doubly-violated and neutral sentences tendentially differed ($F(1,19) = 3.63$, $p = .07$) at frontal electrode-sites. At parietal electrode-sites, the analysis revealed more positive-going ERPs for doubly-violated sentences ($F(1,19) = 17.36$, $p < .001$) and emotional prosodically violated sentences ($F(1,19) = 4.95$, $p < .05$), but not for linguistic prosodically violated sentences. Finally, post hoc contrasts in a time window from 1000 to 1400 ms after sentences onset revealed a more positive-going ERP response to doubly violated sentences when compared to neutral sentences at both frontal ($F(1,19) = 7.84$, $p < .05$) and parietal sites ($F(1,19) = 15.90$, $p < .001$). In addition, emotionally prosodically violated sentences were also found to elicit a more positive-going ERP than neutral sentences at posterior locations ($F(1,19) = 9.56$, $p < .01$). A similar effect was found for linguistically violated sentences at frontal electrode-sites in this time-window ($F(1,19) = 5.24$, $p < .05$). To sum up, violating both emotional and linguistic prosodic expectancy results in a positive ERP component with a whole head distribution, whilst similar effects are found at slightly later points in time at posterior sites for emotional and at anterior sites for prosodic linguistic expectancy violations. The effects are visualized in Fig. 3.

4. Discussion

The current investigation set out to further define the temporal dynamics of emotional and linguistic prosodic processing. Specifically, we violated prosodic expectancy in listeners to investigate how and when specific functions of prosody are used and integrated during online auditory sentence processing. Irrespective of task demands, we report positive ERP responses with different onsets in emotional (~470 ms post violation onset), linguistic (~620 ms post violation onset) and combined (~170 ms post violation onset) emotional and linguistic prosodic expectancy violations. These positivities do not only differ with regard to latency and duration, but also with regard to their distribution. A positive ERP response to emotional prosodic expectancy violations is found at posterior electrode sites, whilst the positivity in response to linguistic prosodic expectancy violations is found at anterior electrode sites. ERPs in response to combined violations are distributed across the whole scalp. Thus, results are comparable to previous electrophysiological results (Astésano et al., 2004; Kotz & Paulmann, 2007; Paulmann & Kotz, 2008b; Pihan et al., 2008), but extend these findings as the present results allow to comment on the in- and inter-dependence of two types of prosody.

4.1. On the absence of task effects

The absence of task effects in the ERP data are somewhat surprising given that the analysis of behavioural results revealed accuracy rate differences. Specifically, participants were better at linguistic as opposed to emotional evaluation of combined violated sentences, whereas the opposite was true when listening to linguistic prosodic expectancy violations. In general, this seems to suggest rather weak task effects (i.e. it is not the case that one task was truly more difficult than the other). Also, given that ERPs measure processes while they unfold in time and behavioural data reflect measurements from a discrete point in time, the two measurements do not necessarily have to go hand in hand. However, without further studies, it will be difficult to say why ERPs failed to support the behavioural task effects. For now, we suggest that task demands do not play an important role during early stages of prosodic re-analyses processes. In fact, this hypothesis goes well with previous data, in which we found task effects (implicit vs. explicit prosody processing) only after successful deviance detection (Kotz & Paulmann, 2007). Interestingly, in this particular study, task demands did not impact the processing of emotional prosodic expectancy violations, but did so when processing combined semantic and prosodic expectancy violations. However, this was only the case when emotional prosody categorization was mandatory. Again, this suggests that task instructions play only a minimal role during on-line processing of expectancy violations. Still, future studies should explore this issue further as imaging studies (e.g. Wildgruber et al., 2006, 2004), which also have a low temporal resolution, have reported differences when task demands focussed on emotional rather than linguistic aspects of stimuli.

4.2. Prosodic positivities – one big family or simply variations of established positivities?

All three expectancy violations elicit a positive ERP component in contrast to non-violated sentences. Even though the positivities have different latencies, durations, and distributions, we suggest

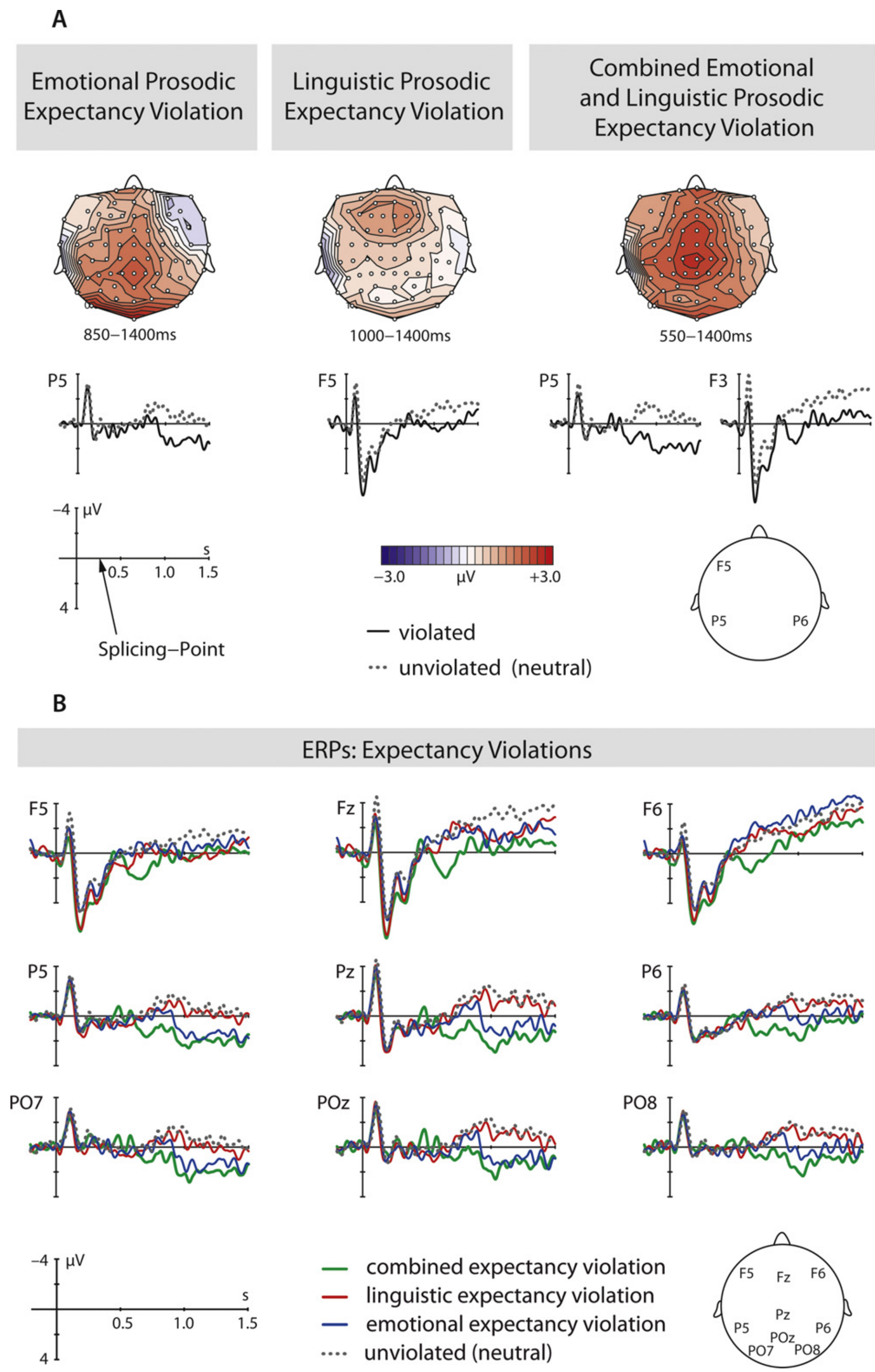

Fig. 3. The illustration shows significant PEP effects at selected electrode-sites from 200 ms before the start of the sentence up to 1500 ms into the sentence. Negativity is plotted upwards. The maps show the distribution of the effect for the time-windows reported in Section 3.

that they belong to the family of prosodic expectancy positivities (PEPs) and are linked to prosodic re-analysis. The question arises as to whether positivities with different latencies, durations, and distribution can be considered to belong to the same family. Moreover, in the past, other positive ERP components such as the P600, linked to syntactic and semantic re-analysis processes, and the P300, linked to mismatch detection, have been reported (e.g. Osterhout & Holcomb, 1992; van Herten, Kolk, & Chwilla, 2005; Picton, 1992). In how far are these positivities different from the positivities found in the present study?

Considering the variability in latency, duration, and distribution, it seems unlikely that the current positivities belong to the P300 family. The P300 is elicited robustly in response to mismatch detection, typically between 250 and 400 ms after deviant onset, often in an odd-ball paradigm (see Polich & Kok, 1995 for review). That is, the P300 is elicited in response to infrequent deviants, particularly when participants are asked to discriminate between standard and deviants (e.g. count deviants). However, the present study does not have a typical mismatch/oddball paradigm: half our stimuli were cross-spliced, whilst the other half was not. Moreover, P300 amplitudes usually range between 10 and 20 μV (c.f. Polich & Kok, 1995), whilst the present positive amplitudes are much smaller. Finally, the morphology of the positivites elicited in response to single violations does not seem to match the typical morphology of either a P3a or P3b component given their duration and rather shift-like appearance. Only the positivity elicited in the combined violation could be argued to reflect a P300-like morphology at least at frontal electrode-sites (but with a long duration). However, if the positivity in response to combined violations is a P300, how can the differences for the single expectancy violations be explained? That is, why should the combined condition elicit a P300 and the single expectancy violations do not? Also, why should they elicit a very different ERP component although arguably similar prosodic re-analysis processes need to be carried out (c.f. paragraphs below)?

Alternatively, the positivity in response to combined expectancy violations comprises two parts (a P300 effect and a PEP) at least at frontal electrode-sites. But, if true that only the combined violation is considered as a “deviant” and the remaining stimuli are considered to be “standard”, the question as to why the single violations (i.e. the stimuli considered standard) still elicit a PEP would nevertheless have to be explained. More importantly, given that non-violated emotional questions were presented in the experiment as fillers, it is unlikely that the double violation was considered the odd one out. Therefore, based on the differences with regard to amplitude, morphology, task, and design between the current and past research, we suggest that the present positive ERP components are unlikely to be P300 responses, i.e. it seems as if the detection of the expectancy violation is not just an ‘automatic’ deviance detection process based on a simple pitch manipulation.

Considering that variability in latency and distribution of the current positivities, as well as the deviation from the standard oddball paradigm speak against the interpretation of the current positivities as P300s, it remains to be addressed whether the components could be interpreted as P600s. The classical P600 has a centro-parietal distribution and generally peaks approx. 600 ms after the onset of a syntactic violation. It is usually elicited in response to syntactic ambiguities and syntactic complexities (e.g. Kaan & Swaab, 2003). More recently it has also been linked to thematic re-analysis (see e.g. Bornkessel-Schlesewsky & Schlesewsky, 2008) processes. In the present study, no such syntactic “anomaly” was present, rendering it unlikely that the underlying processes can easily be linked back to the P600. Moreover, the earliest positivity observed here has an onset of approx. 170 ms after violation occurrence – this seems far too early for a classical P600. In addition, at least two of the positive ERP components observed here follow a different distribution than the classical P600 (though note that some studies have reported a frontally distributed P600; c.f. Kaan & Swaab, 2003). Taken together, the paradigm differences (syntactic violations vs. prosodic violations), as well as the latency and distribution differences, suggest that the present positivities are unlikely to be P600s.

Another positive ERP component reported in the literature is the CPS. The CPS has been argued to be a marker of prosodic processes (e.g. Steinhauer, 2003) and (next to the PEP) seems to be the most promising candidate when trying to link the current positive ERPs to previously reported positivities. The CPS is argued to reflect processes related to prosodic phrasing, i.e. when listeners group segments/words together. Specifically, it has been suggested that the CPS is triggered by a “pre-final constituent lengthening” (Steinhauer, 2003, p. 162). The present paradigm did not require prosodic phrasing, which makes it hard to argue that the positivities observed here are a sub-type of the CPS. Critically, the same neutral head of a sentence was cross-spliced to different sentence endings, thus, no pre-final constituent lengthening occurred in the stimuli (also, using the same head, should then have resulted in the same positivities for all conditions). In sum, this suggests that the observed effects can also not easily be linked to the CPS.

Lastly, the P800 has been reported to be elicited in response to linguistic prosodic expectancy violations (Astésano et al., 2004). We previously hypothesized that the P800 and the PEP in response to emotional expectancy violations reflect similar prosodic re-analysis processes; however, latency and duration of the positivity may depend on cue salience. Here, we would like to argue that the P800 is part of the PEP family, i.e. latency of positivities in response to prosodic expectancy violations will depend on prominence of deviations (see detailed discussion further below). This could then indicate that the positivity reported by Eckstein and Friederici (2005) would also fall under the umbrella term PEP, i.e. unexpected prosodic deviances (e.g. absence of an expected upcoming word, abrupt tone of voice change, abrupt speech act change) will elicit positive ERP components that reflect repair or re-analysis of prosodic information.

Based on these temporal and spatial (e.g. onset latency, duration, scalp distribution) differences as well as functional differences between the current and previously established ERPs, we suggest that the present positivities can be best linked to the previously reported PEP. Specifically, previous data show that the PEP can vary in onset latency (see e.g. Chen et al., 2011 or Paulmann et al., 2008 for earlier onsets of PEP than reported in Kotz & Paulmann, 2007) and distribution (see e.g. distributions reported in Paulmann & Kotz, 2008b; Paulmann et al., 2008; Kotz & Paulmann, 2007) depending on stimuli manipulations. We have further reported that it can be elicited task independently (Kotz & Paulmann, 2007). What all paradigms that elicit a PEP have in common is that listeners detect an abrupt change in prosodic contour. In the following paragraphs, we will outline why we think the current effects should be considered to belong to the PEP family.

4.3. Emotional vs. linguistic prosodic processing: time-course differences

Emotional and linguistic prosody are expressed by overlapping acoustic cues such as changes in pitch, loudness, or speech rate suggesting that the positivities observed here are closely linked to expected acoustical parameter violations. With regard to the functional significance of the ERPs, it has been suggested that they reflect acoustically motivated re-analysis processes (Astésano et al., 2004; Kotz & Paulmann, 2007; Paulmann & Kotz, 2008b). In particular, it has been proposed that with regard to emotional prosodic expectancy violations, the PEP reflects processes that compute or draw on emotion-related acoustic features to determine emotional significance during on-line speech comprehension. Given that linguistic prosody expectancy violations elicit a similar though slightly later and differently distributed positive ERP, we suggest that prosodic re-analysis is triggered when violating prosodic expectancy. In the case of the present emotional expectancy violations, a transition from a neutral to an emotionally significant stimulus is registered, whilst in the case of the present linguistic violations, a transition from a statement to a question is registered. Given the differences in latency and distribution, it is likely that these re-analysis processes rely on partly different neural mechanisms for linguistic and emotional prosodic processing.

The current data support that emotional prosodic expectancy violations are detected more quickly than linguistic prosodic expectancy violations in a paradigm that used the same kind of splicing procedure (and the same sentence context) for both manipulations rendering it unlikely that previously reported latency differences between the P800 (Astésano et al., 2004) and the PEP

(Kotz & Paulmann, 2007; Paulmann & Kotz, 2008b) were due to differences in stimulus manipulation procedures. Interestingly, combined expectancy violations resulted in an even earlier onset of the PEP supporting the hypothesis that different language relevant information streams are rapidly integrated by the listener in emotional contexts (c.f. Kotz & Paulmann, 2007; Paulmann et al., 2009). We previously hypothesized that this speeded processing of combined expectancy violations may be linked to the time-point in sentence processing at which the input information (e.g. acoustic input) first differs from the expected input (e.g. emotional and linguistic information). Combined deviations of emotional and linguistic expectancy may be detected more quickly than single deviations as the deviance is more prominent when it is signalled through multiple, at times possibly redundant, cues.

With regard to the earlier onset of the emotional prosodic expectancy violation as opposed to the linguistic prosodic expectancy violation, we postulate that specific acoustic cue configuration patterns (c.f. Paulmann & Kotz, 2008a) can be used to detect emotionality very early during sentence processing. Deviancy can be detected more quickly when information is emotionally relevant as it could be advantageous to appropriately react to the stimulus (c.f. Vuilleumier, 2005). The question is, of course, which acoustic cues are used by the system to rapidly determine emotionality. As mentioned previously, it has been shown that specific combinations of acoustic cues (e.g. high pitch, quick tempo, loud voice) can signal specific emotions (e.g. Banse & Scherer, 1996) and that these patterns may be similar across languages (Pell et al., 2009). In fact, we have shown that stimuli can accurately be classified with regard to their emotional category based on three acoustic cues (pitch, intensity, duration) in a discriminant feature analysis (Paulmann et al., 2008). For the current set of stimuli, acoustical analyses suggest that the different conditions (neutral, combined violated, emotionally/linguistically violated) differ from one another with regard to several acoustical parameters post-splicing point (c.f. Table 1a and b). In particular, linguistic prosody was expressed with a large F0 range (typical rising pattern for questions) and few loudness variations, whilst latter cue was more prominent for emotional stimuli. Furthermore, acoustic measurements from the first vowel occurring shortly after splicing-point (taken from the articles “der”, “die”, “das”, “den”), showed that mean pitch, the first formant, as well as jitter and harmonic to noise ratio measurements differed between neutral (i.e. the expected prosody) and emotional as well as combined expectancy violations, but not between neutral and linguistic violations (in this case, only the mean pitch contrast and one jitter measurement contrast turned out to be significant). This not only suggests that participants can detect prosodic expectancy violations quickly when hearing non-expected acoustic cues, but also it highlights the possibility that emotional and linguistic prosodic processing can indeed be based on different acoustic cues with emotional prosodic processing relying more on voice quality related measurements such as jitter and harmonic to noise ratio in addition to pitch (patterns) which can be particularly meaningful for linguistic processing. The latter suggestion is in line with previous observations that suggested that emotional prosodic deviance detection may not be tied to just one particular acoustical parameter configuration (Paulmann & Kotz, 2008b) when we showed that both sentences going from neutral to emotions that were expressed with high pitch (e.g. pleasant surprise, happiness) and sentences going from neutral to emotions that were expressed with low pitch (e.g. sadness) elicited the same kind of PEP.

Alternatively, the earlier onset of the PEP in response to combined and emotional expectancy violations could be due to emotional prosody diverging earlier than linguistic prosody from neutral prosody. However, acoustical analyses for F0 and F1 contours make it unlikely that the PEP differences observed are due to

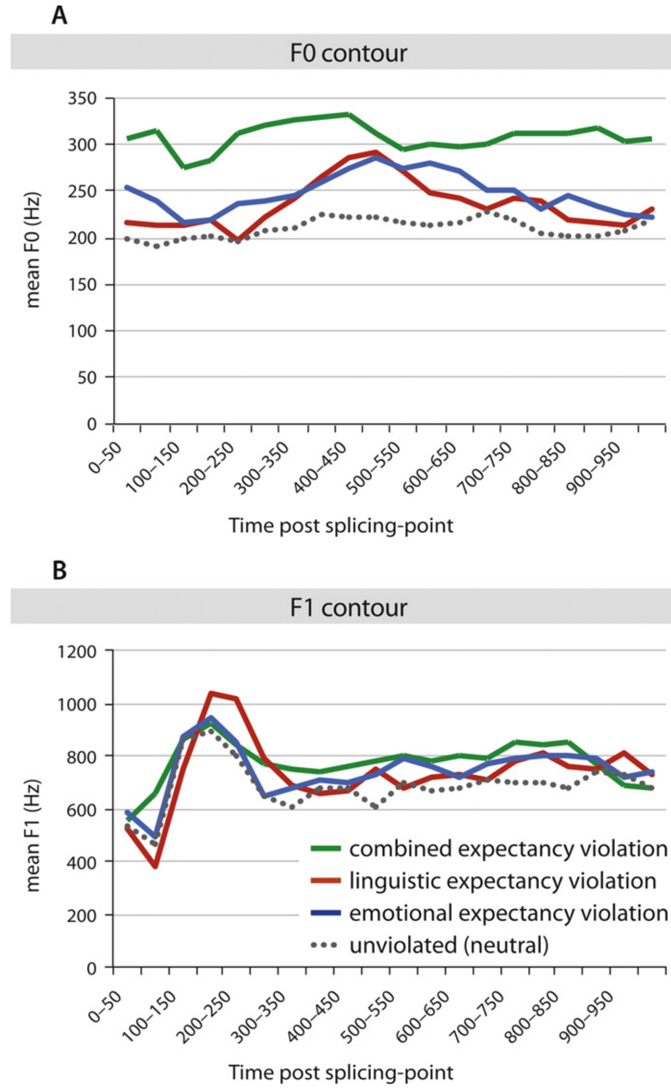

Fig. 4. (A and B) The graphs show results of acoustical analyses for F0 and F1 contours in 50 ms time bins following the splicing-point.

linguistic prosody differing from neutral prosody at a later point in time than emotional prosody differing from neutral prosody. Fig. 4A and B shows F0 and F1 values every 50 ms after the splicing-point. Statistical analyses confirmed that F0 differed significantly between non-violated sentences and emotional, linguistic, and combined expectancy violations in the first 100 ms after the splicing-point as well as between 200–250 ms and 400–450 ms after the splicing-point. Thus, linguistic expectancy violations and neutral non-violated sentences differed significantly right after the splicing-point (in addition to later time points). The time-line analysis also revealed that neutral sentences differed from combined expectancy violations in almost all time-windows analysed, suggesting that the prosodic contour of emotional questions is, in general, very different compared to the prosodic contour of neutral sentences. This difference may lead to the long lasting PEP effect observed in the present data. Interestingly, F1 differed significantly between neutral sentences and combined expectancy violations in the time-windows of 50–100 ms and between 200 and 250 ms for neutral sentences and linguistic prosodic expectancy violations but not all between neutral sentences and emotional violations. Taken together, this suggests that both F0 and F1 differed between linguistic prosody and neutral prosody shortly after the splicing-point. This makes it unlikely that the PEP is elicited later for linguistic violations because linguistic prosody diverges later from neutral prosody. Instead, we believe that this is further evidence that different cues (or cue combinations) signal different prosodic functions. For example, voice quality related cues could be particularly prominent for emotional prosodic stimuli (but never for linguistic prosodic stimuli). In short, dependent on the presence/absence of acoustic cues, the listener can detect the relevant information leading to different processing mechanisms as reflected in the current ERP differences (latency, duration, and topography) for the two functions. Going back to the hypothesis about the relationship between processing speed and the amount of information available to the listener, one can also assume that emotional and combined violations are detected more rapidly than linguistic violations simply because more emotionally relevant acoustic cues are available from an early point in time until the end of the signal.

Finally, it needs to be mentioned that the present study explored emotional prosodic deviance detection by presenting angry sounding stimuli only. The PEP has been reported to be speaker, task, valence, and emotional category independent (e.g. Kotz & Paulmann, 2007; Paulmann & Kotz, 2008b); however, we cannot completely rule out the possibility that the current findings cannot be generalized to other emotional prosodies. The same limitation holds true for our linguistic prosodic stimuli. Following previous designs (e.g. Pihan et al., 2008; Astésano et al., 2004), we presented participants with stimuli that signalled different speech acts (question/declarative sentences). Obviously, differentiating between different speech acts is not the only linguistic function that prosody serves (e.g. turn taking). Thus, future studies will have to explore whether emotional prosody processing always precedes linguistic prosody processing or whether the precedence is dependent on cue salience.

Still, we suggest that prosodic deviance detection is based on a conglomerate of several acoustical parameters that signal emotional or linguistic function of prosody. Emotional-relevant pattern changes are detected more rapidly than non-emotional pattern changes due to their salience (e.g. emotional significance, cues being available earlier). Pattern changes that indicate combined violations are even more quickly detected because the deviance is more prominent when more than one information channel signals a change in prosody.

4.4. Emotional vs. linguistic prosodic processing: PEP distribution differences

Violating both emotional and linguistic prosodic expectancy results in an ERP component that is distributed over the whole scalp as opposed to only posterior effects for emotional and only anterior effects for prosodic linguistic expectancy violations, suggesting that both processes interact in time when the two functions are time-locked. The distribution of effects is in contrast to previous reports that revealed a PEP with a right-lateralized maximum for emotional prosodic expectancy violations (Kotz & Paulmann, 2007; Paulmann & Kotz, 2008b) and a P800 with a left temporo-parietal maximum in response to linguistic prosodic expectancy violations (Astésano et al., 2004). However, other reports revealed a globally distributed or even left-lateralized PEP response (Chen et al., 2011; Experiments 1 and 2 respectively) depending on attention/task focus. In addition, Marques et al. (2007) reported a whole-head distributed positive ERP in response to prominent deviance (strong incongruity), whilst less prominent deviance (weak incongruity) resulted in a left-centro-parietally distributed positivity for musicians. For non-musicians, the effect had a later onset and a bilateral distribution for prominent deviances but was only visible in the midline region for less prominent deviance. If one accepts the possibility that linguistic prosodic expectancy violations are less prominent than emotional prosodic expectancy violations and even less prominent than combined violations, then the present results fit well with the current literature. Specifically, this would suggest that both task demands (e.g. Chen et al., 2011) and stimuli properties can influence latency and distribution of positive ERP components that are related to prosodic deviance detection. Moreover, it seems as if simple prosodic deviance detection (linguistically or emotionally motivated) can generally be linked to a positive ERP component, whilst combined prosodic and semantic expectancy violations are tied to components with negative polarity (e.g. Kotz & Paulmann, 2007; Paulmann & Kotz, 2008b). Moreover, paired with the observation that other language related deviance detection (e.g. semantic, syntactic) processes are sometimes reported to elicit components with different polarity, latency, and distribution (e.g. LAN, N400; and see Donchin et al., 1997 for review on deviant processing in general), we suggest that processes reflected in positivities from the "PEP-family" are specifically tied to prosodic evaluation processes.

4.5. Emotional vs. linguistic prosodic processing: in- or inter-dependent processing mechanisms

The current findings allow to directly compare the neural mechanisms underlying emotional and linguistic prosodic processing. Given differences in latency, duration and distribution of the PEP, it is suggested that the two functions of prosody may rely on partly different neural mechanisms. Whilst functionally speaking, the PEP for both conditions may be linked to acoustically motivated re-analyses processes, distributional and time-course differences also allow to speculate that these re-analyses processes are modulated by a differentiated brain network. This finding is partly in line with the hypothesis that the same acoustic cues can be processed by different neural networks depending on their function (e.g. Van Lancker, 1980; Van Lancker & Sidtis, 1992). Initially, the functional hypothesis has primarily been linked to hemisphere differences in that the more linguistically emphasized the task, the more the left hemisphere is involved, and, theoretically, the smaller the linguistic load, the stronger the right hemisphere involvement. The current experiment switched between emotional and linguistic task demands in every experimental block. Though we fail to report task differences, it is conceivable that switching between emotional and linguistic aspects of stimuli recruited both hemispheres equally. This is clearly speculative but could be tested in future studies with more participants (to exclude the possibility that missing task effects are due to a lack of power), or a between-subjects design (one group carrying out the linguistic task and the other engaging in the emotional task).

Whilst emotional and linguistic task demands varied between blocks in the present study, both tasks required the participants to focus on the prosody of the sentences presented. In the past, we have utilized tasks that require the listener to process prosody implicitly rather than explicitly (e.g. Paulmann et al., 2009; Paulmann & Kotz, 2008b) and have found a similar positive ERP component in response to emotional prosodic violations. Thus, whilst the PEP elicited to emotional prosody seems to be task-independent (see Kotz & Paulmann, 2007 for direct comparison), future studies could explore if the PEP effect for linguistic prosody can also be replicated when prosody is not in task focus.

Another prosody processing hypothesis, the so-called parameter dependence hypothesis suggests that specific acoustic cues are preferably processed in one hemisphere over the other (e.g. pitch is preferably processed in the right hemisphere whilst durational parameters and loudness engage the left hemisphere more strongly). Whilst our data fail to report hemisphere distribution differences, the different functions of prosody elicit ERP differences at anterior and posterior electrode sites. If one accepts the possibility that certain acoustic cues play a more dominant role during emotional as opposed to linguistic prosody processing (e.g. voice quality parameters) as suggested by our acoustical data, then our current data lend support to the notion that different parameters can engage different brain regions to a different extent (though note that ERPs are not suited to comment on which brain regions could modulate specific effects). That is, the current results suggest that the two functions of prosody recruit partially different brain structures, arguing for the possibility that prosody processing forms a continuum and is relative rather than absolute. This claim is supported by the finding that combined expectancy violations elicit a PEP at both anterior and posterior electrode sites. Thus, we provide evidence that when participants focus on emotional or linguistic aspects of stimuli, the neural network underlying emotional and prosody processing is not strictly lateralized (see Pihan et al., 2008 for similar finding in DC recordings).

4.6. Limitations of the current study

Whilst the present results allow addressing the in- and interdependence of emotional and linguistic prosody at a new level, some questions remain unanswered. In the following we raise a number of questions that should be followed up in future research:

First, the current study used the distinction of statements and questions to represent linguistic prosody and angry prosody to represent emotional prosody. Clearly, this investigation is only first evidence and future studies will have to confirm whether similar effects can be found when stimuli of different prosodic quality are used. It seems likely though that the PEP will not differ when using other emotional prosodies (c.f. Paulmann & Kotz, 2008b). However, it remains open whether the same holds true for linguistic prosody. Future studies should therefore explore whether other prosodic expectancy violations elicit similar results (i.e. violating speech act expectancies).

Secondly, one of the surprising findings is the absence of ERP task effects. One suggestion is that task effects as reported in imaging (e.g. Wildgruber et al., 2006, 2004) or behavioural data (e.g. present findings) reflect differences at later processing stages and do not influence early on-line processing of prosodic expectancies. Future studies should thus clarify if task effects come to play only at later processing stages by directly comparing early and late emotional prosody evaluation stages.

Thirdly, the variability of PEP latency and duration should be explored further. In particular, future studies should determine whether the earlier onset of the PEP in response to emotional prosodic and combined violations can be linked to a meaningful acoustic cue (or cues) or to partially redundant cues that make it easier to detect expectancy violations. One way to rule out either one possibility is to create stimuli that systematically vary in the availability of one or multiple cues that have proven to be meaningful in the past (e.g. F0, F1, HNR, high-frequency energy).

4.7. Conclusion

The present study set out to explore the comparative nature and time-course underlying two functions of sentence prosody, namely emotional and linguistic prosody processing. Results suggests that the two prosodic functions run a different time-course in that emotional prosodic aspects are processed more rapidly during on-line speech comprehension than linguistic prosodic aspects. We have suggested that this timing advantage might be due to the listener paying special attention to emotional stimuli very early during processing as they need to be integrated with other language functions at a later time point (c.f. Kotz & Paulmann, 2011; Schirmer & Kotz, 2006). Moreover, distribution and latency of reported ERP effects suggest that the two functions engage in a partly differentiated neural network when processed individually but use shared neural resources when processed in parallel, supporting the idea that the two functions interact in time during on-line speech comprehension. Future models of auditory speech comprehension should try to address the role of both emotional and prosodic prosody during speech processing.

Acknowledgements

The authors would like to thank Ina Koch for help with subject recruitment and data acquisition and Kerstin Flake and Andrea Gast-Sandmann for help with graphical presentation. This work was funded by the German Research Foundation (DFG FOR-499 to S.A.K.).

References

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.
Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636.
Beringer, J. (1993). *Experimental run time system*.
Boersma, P., & Weenink, D. (2009). *Praat: Doing phonetics by computer* (*Version 5.1.25*) [*Computer program*]. Retrieved from http://www.praat.org/.
Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on semantic P600 effects in language comprehension. *Brain Research Reviews*, *59*, 55–73.
Chen, X., Zhao, L., Jiang, A., & Yang, Y. (2011). Event-related potential correlates of the expectancy violation during emotional prosody processing. *Biological Psychology*, *86*, 158–167.
Donchin, E., Spencer, K. M., & Dien, J. (1997). The varieties of deviant experience: ERP manifestations of deviance processors. In G. J. M. van Boxtel, & K. B. E. Bocker (Eds.), *Brain and behavior: Past, present, and future* (pp. 67–91). Tilburg: Tilburg University Press.
Eckstein, K., & Friederici, A. D. (2005). Late interaction of syntactic and prosodic processes in sentence comprehension as revealed by ERPs. *Cognitive Brain Research*, *25*, 130–143.
Kaan, E., & Swaab, T. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of Cognitive Neuroscience*, *15*(1), 98–110.
Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, *1151*, 107–118.
Kotz, S. A., & Paulmann, S. (2011). Emotion, language and the brain. *Language and Linguistics Compass*, *5*(3), 108–125.
Li, X., Chen, Y., & Yang, Y. (2011). Immediate integration of different types of prosodic information during on-line spoken language comprehension: An ERP study. *Brain Research*, *1386*, 139–152.
Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violations in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, *19*(9), 1453–1463.
Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 785–806.
Paulmann, S., Jessen, S., & Kotz, S. A. (2009). Investigating the multimodal nature of human communication: Insights from ERPs. *Journal of Psychophysiology*, *23*, 63–76.
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., & Kotz, S. A. (2008). Functional contributions of the basal ganglia to emotional prosody: Evidence from ERPs. *Brain Research*, *1217*, 171–178.
Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, *37*, 417–435.
Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology*, *9*, 456–479.
Pihan, H. (2006). Affective and linguistic processing of speech prosody: DC potential studies. *Progress in Brain Research*, *156*, 269–284.
Pihan, H., Tabert, H., Assuras, S., & Borod, J. (2008). Unattended emotional intonations modulate linguistic prosody processing. *Brain and Language*, *105*, 141–147.
Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: An integrative review. *Biological Psychology*, *41*, 103–146.
Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*(1), 24–30.
Schirmer, A., Striano, T., & Friederici, A. (2005). Sex differences in the pre-attentive processing of vocal emotional expressions. *NeuroReport*, *16*, 635–639.
Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language*, *86*, 142–164.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*, 191–196.
Thoennessen, H., Boers, F., Dammers, J., Chen, Y. H., Norra, C., & Mathiak, K. (2010). Early sensory encoding of affective prosody: Neuromagnetic tomography of emotional category changes. *NeuroImage*, *50*, 250–259.
van Herten, M., Kolk, H. H., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, *22*, 241–255.
Van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal. *International Journal of Human Communication*, *13*, 101–109.
Van Lancker, D., & Sidtis, J. (1992). The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: All errors are not created equal. *Journal of Speech and Hearing Research*, *35*, 963–970.
Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Science*, *9*, 585–594.
Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research*, *156*, 249–268.
Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., et al. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cerebral Cortex*, *14*, 1384–1389.
Winkler, I., Denham, S., & Nelken, I. (2009). Modeling the auditory scene: Predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, *13*, 532–540.