



Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Neural processing of vocal emotion and identity

Katja N. Spreckelmeyer^{a,*}, Marta Kutas^{b,c}, Thomas Urbach^b, Eckart Altenmüller^d, Thomas F. Münte^e

^a Department of Psychiatry and Psychotherapy, School of Medicine, RWTH Aachen University, Pauwelsstr. 30, 52074 Aachen, Germany

^b Department of Cognitive Science, University of San Diego-California, La Jolla, CA 90021-0515, USA

^c Department of Neurosciences, University of San Diego-California, La Jolla, CA 90021-0515, USA

^d Institute of Music Physiology and Musicians' Medicine, University of Music and Drama-Hannover, Hohenzollernstr. 47, D-30161 Hannover, Germany

^e Department of Neuropsychology, University of Magdeburg, Universitätsplatz 2, D-39106 Magdeburg, Germany

ARTICLE INFO

Article history:

Accepted 8 June 2008

Available online xxx

Keywords:

Emotion

Identity

Voice

Prosody

Timbre

Auditory perception

Event-related potentials (ERP)

ABSTRACT

The voice is a marker of a person's identity which allows individual recognition even if the person is not in sight. Listening to a voice also affords inferences about the speaker's emotional state. Both these types of personal information are encoded in characteristic acoustic feature patterns analyzed within the auditory cortex. In the present study 16 volunteers listened to pairs of non-verbal voice stimuli with happy or sad valence in two different task conditions while event-related brain potentials (ERPs) were recorded. In an emotion matching task, participants indicated whether the expressed emotion of a target voice was congruent or incongruent with that of a (preceding) prime voice. In an identity matching task, participants indicated whether or not the prime and target voice belonged to the same person. Effects based on emotion expressed occurred earlier than those based on voice identity. Specifically, P2 (~200 ms)-amplitudes were reduced for happy voices when primed by happy voices. Identity match effects, by contrast, did not start until around 300 ms. These results show an early task-specific emotion-based influence on the early stages of auditory sensory processing.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Voices provide listeners with much more than the semantic information conveyed by the linguistic content of any given utterance—i.e., its meaning. Additionally, it may also afford inferences about the speaker's gender, age, health, or emotional state; these inferences are based on certain paralinguistic sound features of the voice (Traunmüller, 1997). Although these features seem to be analyzed automatically and effortlessly, very little is known about the brain mechanisms involved. The question is not trivial, since the encoding of different types of paralinguistic information, e.g. speaker-specific voice features and emotional prosody rely on relatively similar acoustic cues such as vocal quality and pitch contour.

How then does the brain extract the relevant cues? Recent models of voice perception (Belin, Fecteau, & Bedard, 2004; Ethofer, Anders, Erb et al., 2006; Schirmer & Kotz, 2006) assume that the recognition of paralinguistic information of this type is a multi-stage process. The first stage is assumed to consist of low-level sensory processing (e.g. frequency, intensity) of the acoustic input, relatively similar for all paralinguistic aspects (Belin et al., 2004). Consistent with this, direct comparison of speaker identification and emotion identification tasks were indeed found to activate

some of the same brain regions in the anterior temporal lobes (Imaizumi et al., 1997). Imaging studies of emotion identification or speaker identification in isolation implicate temporal lobe structures, especially bilateral regions of the middle STS close to A1, in both (Belin & Zatorre, 2003; Ethofer, Anders, Erb et al., 2006; Grandjean et al., 2005; von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003). However, each function/task alone also activates additional brain regions that the other does not, consistent with the proposal of additional processing steps unfolding in distinct task-specific cognitive systems (Belin et al., 2004). Because of the low temporal resolution of functional imaging the question of when in the auditory processing stream the analysis becomes task-specific remains unanswered. Although it has been suggested (Belin et al., 2004) that the task-specific steps involved in the perception of prosody take place subsequent to sensory processing, there is no empirical evidence for this assumption. Imaging data (Sander et al., 2005) suggest instead, that emotional significance may modulate the early sensory processing of vocal input, analogous to vision (Vuilleumier, Armony, Driver, & Dolan, 2001).

To get a finer delineation of the time courses of the processing of the different paralinguistic aspects of auditory input, we recorded event-related brain potentials (ERPs) as participants evaluated either emotional valence or voice identity. ERPs have proven to be an excellent means for the on-line monitoring of early auditory perception (Johnstone, Barry, Anderson, & Coyle, 1996; Näätänen et al., 1988; Winkler, Takegata, & Sussman, 2005;

* Corresponding author. Fax: +49 241 8082401.

E-mail address: kspreckelmeyer@ukaachen.de (K.N. Spreckelmeyer).

Woldorff & Hillyard, 1991). Early ERP components are especially sensitive to physical features of auditory inputs (Antinoro, Skinner, & Jones, 1969; Picton, Goodman, & Bryce, 1970). To disentangle task-specific modulations from possible confounds due to different acoustic parameters, we employed a priming paradigm with non-verbal voice stimuli, which were either congruent or incongruent with respect to the emotion expressed or the identity of the speaker (voice). By employing the exact same stimuli in both congruity conditions and in two different tasks (emotion and identity), we held the physical parameters of the target voices constant across all conditions (and associated comparisons). Thus, any ERP differences can be attributed solely to the processing of the emotional or voice characteristics of the auditory inputs. More specifically, we expected the priming of task-specific features in the congruent condition to be reflected in the ERPs when compared to the incongruent condition. We could thus use the relative timing of the priming effects for the two task conditions to infer the relative time course of the analysis of the emotional (prosodic) and identity aspects of voice input.

2. Materials and methods

2.1. Stimulus material

Stimulus pairs were created from individual musical notes sung on the syllable 'ha' by five professional female singers. Sung tones were used in order to avoid the linguistic and structural influences that a verbal utterance might have on the decoding process (Banse & Scherer, 1996). Singers were instructed to imbue each individual note with a happy or a sad quality. To select the notes that could be categorized on affect despite their brevity, we asked 10 participants (5 women, 21–30 yrs) naive to the purpose of the experiment and not involved in the main experiment to rate them on a 7-point-scale ranging from 1 (very sad) to 7 (very happy). For the EEG experiment only those tones were used that had consistently been categorized as happy (5, 6 or 7 on the rating-scale) or sad (1, 2 or 3 on the rating-scale) by at least 7 of 10 raters. From this pre-selection two sets of 10 stimuli were created, one consisting of tones categorized as happy, the other one consisting of tones categorized as sad. These sets were matched for arousal [$\text{mean} = 2.73$ ($SD = 0.39$) for happy, $\text{mean} = 2.39$ ($SD = 0.28$) for sad on a 1- to 7-scale, n.s.]. The intensity of valence was 5.33 ($SD = 0.45$) for happy, and 2.88 ($SD = 0.48$) for sad tones. The mean length was 373 ms ($SD = 63$ ms) for happy tones and 414 ms ($SD = 53$ ms) for the sad tones. The happy tones ranged between C4 (~262 Hz) and A#4 (~466 Hz) in pitch, the sad tones between B3 (~247 Hz) and A#4 (~466 Hz).

2.2. Participants

Sixteen undergraduates of the University of California, San Diego (UCSD) (8 women, 18–25 yrs) participated in the ERP-experiment for cash or course credit. The study was approved by the UCSD Human Subjects Committee and participants gave informed, written consent. All participants were right-handed. None of the students was enrolled in music classes although some reported that they had learned to play a musical instrument during childhood, which they did not play anymore. For the final analysis, one subject was excluded due to an excessive number of blink artifacts and a second due to performance at chance level.

2.3. Design

The two tasks were presented in alternating blocks. In task A (emotion matching) participants were asked to indicate whether

the emotional expression in a pair of tones was the same or different. For task B (identity matching) participants were asked to indicate whether the identity of the singer of a pair of tones was the same or different. Each task consisted of two blocks of 54 tone pairs. The order of blocks was counterbalanced across participants (either ABAB or BABA). Of the 108 total pairs 50% were congruent in emotional valence and 50% incongruent; there were an equal number of happy and sad tones. The same stimuli occurred as targets in both tasks, and in the congruent and incongruent conditions. In the emotion matching task, the first and the second tones were always sung by different singers regardless of the expressed emotion. Likewise, in the identity matching task, the emotion expressed for the two tones was identical, even if the singers were not. The same stimulus pairs were presented to all participants although in different orders for each participant.

2.4. Experimental procedure

Participants of the ERP-experiment were seated 127 cm in front of a 21-inch-monitor in a soundproof, electrically shielded chamber. Prior to the experiment each participant's individual auditory threshold was determined and stimulus volume was adjusted via an attenuator (Hewlett-Packard 350 D) to guarantee the same relative loudness for all participants.

The experimental session started with a short practice trial to familiarize the participants with the procedure. Participants were asked to fixate a cross at the center of the screen and refrain from blinking throughout the trial. Stimuli were presented via loudspeakers suspended from the ceiling of the testing chamber approximately 2 m in front of a participant, 0.5 m above and 1.5 m apart. The inter stimulus interval (ISI) was 2050 ms. After the offset of the second tone, the fixation cross remained on the screen for an additional 1200 ms. The screen went black for 400 ms before a prompt to respond appeared, at which point participants were asked to press one of two buttons signaling 'same' and 'different' decisions; assignment of hands to the response buttons was counterbalanced across participants. The button press was followed by a black screen for 1500 ms before the next trial started. Small pauses between blocks allowed the participants to stretch and to rest their eyes.

3. Apparatus and recording

3.1. ERP-recording

The electroencephalogram (EEG) was recorded from 26 tin electrodes mounted in an elastic cap with reference electrodes placed at the left and right mastoid (see [supplementary material for montage](#)). Electrode impedances were kept below 5 k Ω . The EEG was digitized continuously at 250 Hz. The EEG activity was recorded against a left mastoid reference electrode and re-referenced off-line to the mean of the activity at right and left mastoid electrodes. Electrodes placed at the outer canthus of each eye were used to monitor horizontal eye movements. Vertical eye movements and blinks were monitored by an electrode below the right eye referenced to the right lateral prefrontal electrode. Averages were obtained for 2048 ms epochs (including a 500 ms pre-stimulus baseline period) separately for first and second tones of each pair. Trials contaminated by eye movements or amplifier blocking or other artifacts within the critical time window were rejected prior to averaging. ERPs were calculated by time domain averaging correct trials of each participant in different conditions.

The average ERPs were quantified by mean amplitude measures using the mean voltage of the 500 ms time-period preceding the onset of the stimulus as a baseline reference. Time windows to cal-

culate mean amplitudes for the statistical analyses were set as follows: 50–150 (N1), 150–250 (P2), 300–400 and 400–1000 ms. Electrode sites used for the analysis were midline prefrontal (MiPf), left and right lateral prefrontal (LLPf and RLPf) and medial prefrontal (LMPf and RMPf), left- and right-medial frontal (LMFr and RMFr), and medial central (LMCe and RMCe), midline central (MiCe), midline parietal (MiPa), left and right mediolateral parietal (LDPa and RDPa) and medial occipital (LMOc and RMOc).

The resulting data for the second tones of both tasks were entered into separate analyses of variance (ANOVAs) on repeated measures with factors ‘congruence’ (=task-specific congruence vs. incongruence between 1st and 2nd tone), ‘emotion’ (=emotional category of the 2nd tone [happy or sad]), ‘laterality’ (left-lateral, left-medial, midline, right-medial and right-lateral) and ‘caudality’ (prefrontal, fronto-central and parieto-occipital). Only correct trials were included in the analysis. Separate ANOVAs were performed on data of the 4 time-windows followed by comparisons between pairs of conditions. Only effects are reported reaching an alpha level of .05.

Whenever there were two or more degrees of freedom in the numerator, the Huynh-Feldt epsilon correction was employed. The original degrees of freedom and the corrected p -values are reported.

3.2. Reaction time experiment

For methodological reasons participants’ responses in the ERP-experiment were purposely delayed. Thus, no reaction times were recorded. However, a separate reaction time experiment was conducted using the same stimulus set with a different group of 10 participants (5 women, 26–32 yrs).

The emotion matching task and the identity matching task occurred in alternating blocks resulting in a total of eight blocks of 15 tone pairs per task. Order of pairs was randomized. After four blocks of each task, hand assignment to response keys was switched in all participants. The first 10 trials of the initial block and of the first block after the hand reassignment were not included in any analyses.

4. Results

4.1. ERP-experiment

4.1.1. Behavioral data

The mean percentage of correctly matched pairs in the emotion matching task was low but well above chance (62.9%; $SD = 9.1\%$). The performance in the identity matching task was slightly, although not significantly, better (68.5%; $SD = 13.1\%$; $p = 0.091$). There was no significant correlation within subjects of the performance scores in the two tasks ($r = -0.32$; $p = n.s.$).

4.1.2. ERP data

The grand average waveforms to the second tone in both tasks (see Figs. 1 and 2) are characterized by an N1–P2-complex, typical of auditory stimulation (Näätänen et al., 1988), followed by a long-duration component negative over frontal sites and increasingly more positive toward more posterior sites (main effect of caudality between 300 and 400 ($F(2,26) = 12.13$, $p < 0.001$) and 400 and 1000 ms ($F(2,26) = 11.59$, $p < 0.001$)).

In the emotion matching task no main effect of emotion or congruence was found. The earliest effect of congruence in the emotion matching task is seen in the time window of the P2 component (150–250 ms), but only for happy tones (interaction congruence \times emotion: $F(1,13) = 5.65$, $p < 0.05$; $p < 0.05$ for pairwise comparison of happy/congruent vs. happy/incongruent); the P2 is smaller for emotionally congruent than incongruent happy tones (Fig. 1). Subsequent to this there is a significant effect of emotional congruence between 300–400 ms post-stimulus onset, evident in the response to both happy and sad tones: regardless of the emotion, congruent responses are associated with relatively greater negativity. This later effect is seen at fronto-central and parieto-occipital but not at prefrontal electrodes (interaction congruence \times caudality, $F(2,26) = 3.66$, $p < 0.05$, $p < 0.05$ for pairwise comparisons).

In the identity matching task, too, no main effect of congruence or emotion was found. Here, the earliest effect of congruence (see Fig. 2) is seen in the 300–400 ms time window ($F(1,13) = 4.72$,

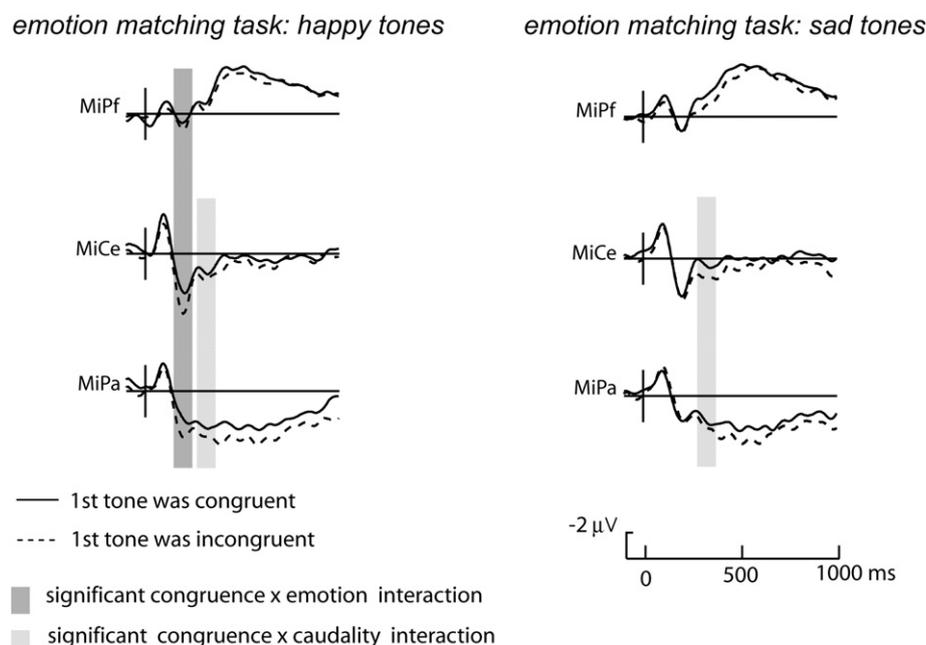


Fig. 1. Grand average ERPs to the 2nd tone of a pair in the emotion matching task at the midline electrodes. Conditions are depicted separately for happy tones (left) and sad tones (right). Time windows with effects reaching significance (< 0.05) are highlighted differently for different effects.

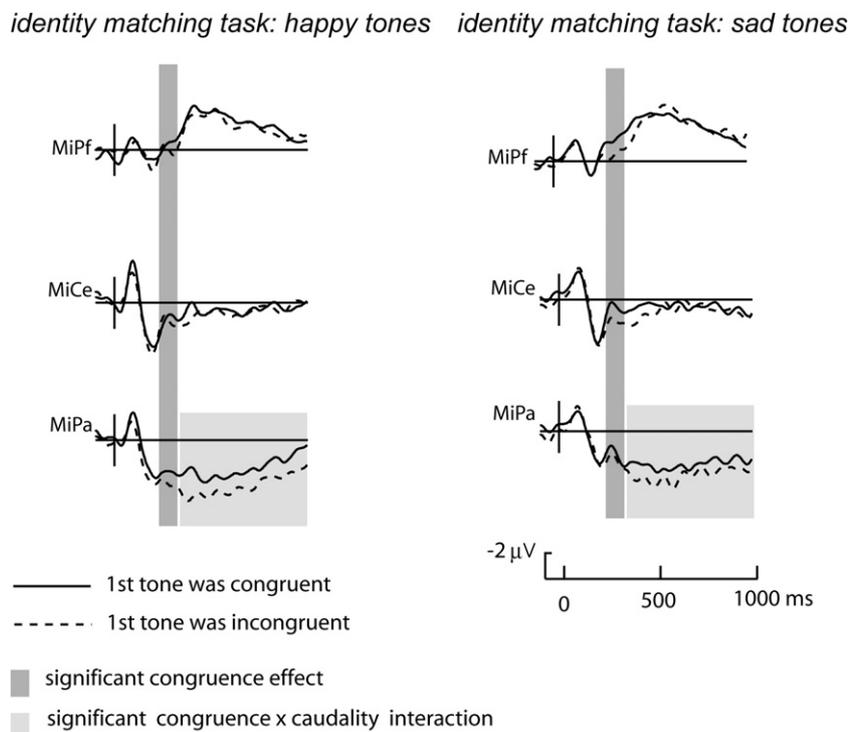


Fig. 2. Grand average ERPs to the 2nd tone of a pair in the identity matching task at the midline electrodes. Conditions are depicted separately for happy tones (left) and sad tones (right). Time windows with effects reaching significance (<0.05) are highlighted differently for different effects.

$p < 0.05$). Tones preceded by a tone sung by a different singer are associated with a more positive going ERP than when the preceding tone was sung by the same singer. This congruence effect is also present in the consecutive time window (400–1000 ms), although only at posterior recording sites (congruence \times caudality interaction, $F(2,26) = 4.31$). Moreover, emotion modulates the P2 amplitude in the identity matching task (time window 150 to 250 ms, $F(1,13) = 5.54$, $p < 0.05$) with a larger amplitude for happy tones.

To summarize, there is an earlier effect of “emotional congruence” in the emotion matching than of “identity congruence” in the identity matching task, although only for happy tones. P2-amplitudes appear reduced when the happy voice tone is preceded by a different but similarly happy voice tone. No such early priming effect is present in the identity matching task, although tones in both tasks show a similar, later congruence effect.

One possible explanation for the early P2 congruence effect might be the acoustical similarity of the first and the second tones in the congruent happy condition. On this view, the P2-priming effect would be driven strictly by physical similarity (Wiggs & Martin, 1998). If this were the case, we would expect to see a similar P2 amplitude reduction for the identical stimulus set in the identity matching task. Although an incomplete factorial design was employed, the incongruent tone pairs in the identity matching task were physically identical to the congruent tone pairs in the emotion matching task. In both conditions the prime and target tones were sung by different voices while the emotion expressed was held constant. The emotion-specific acoustical structure of tone 1 and 2 thus were identical in the two tasks. The associated ERPs for these two conditions are superimposed in Fig. 3 and contrasted with the incongruent emotion condition. This comparison clearly shows that the P2 amplitude reduction is specific to the emotion matching task; it is not present for the same physical

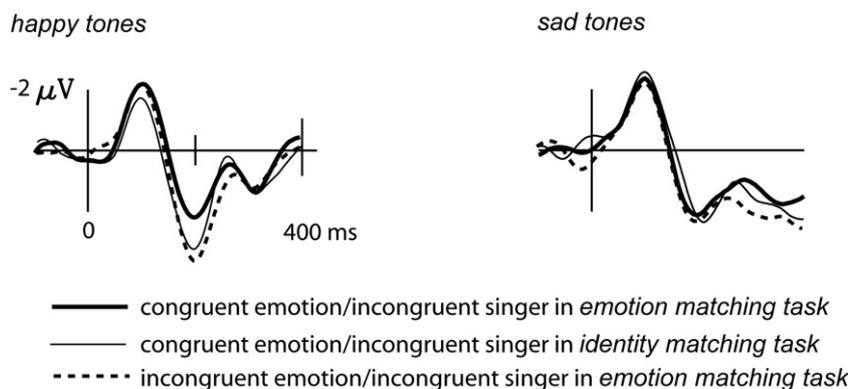


Fig. 3. A P2-priming effect was only seen in the emotion matching task. Though tone pairs were identical in the incongruent identity condition (thin line) and the congruent emotion condition (bold line) priming only occurred in the happy condition of the emotion matching task (left).



Fig. 4. Results of the reaction time experiment: Reactions were faster to happy (bold line) compared to sad voices (dotted line) though only in the emotion matching task (interaction task \times emotion).

stimuli in the identity matching task. The statistical comparison of mean amplitudes to the same (second) tone in both tasks (congruent emotion/incongruent singer in emotion vs. identity matching task) corroborated the difference ($F(1, 13) = 14.49, p = 0.0022$).

4.2. Results of the reaction time experiment

On average, reaction times were statistically indistinguishable for the target (second) tones of the tone pairs in the two matching tasks (emotion 1311 ms vs. identity 1284 ms, $F(1,9) = 0.145, p = 0.712$). However, there was a task by emotion interaction ($F(1,9) = 9.4, p = 0.013$), reflecting faster response times in the emotion than identity matching task when the second tone expressed a happy tone (pairwise comparison of happy in identity task vs. happy in emotion task: $p < 0.01$, see Fig. 4). There was also main effect of emotion (1256 for happy vs. 1339 ms for sad items, $F(1,9) = 7.007, p = 0.027$).

5. Discussion

The aim of the experiment was to delineate the relative time courses of the neural events involved in the recognition of emotion and speaker identity in vocal tone stimuli. Our ERP data suggest that auditory processing is relatively task independent for approximately the first 150 ms after auditory input, although null effects must necessarily be interpreted with caution. In our investigation, auditory N1 amplitudes, known to vary with manipulations of physical features, were not altered by either task or prime-target (in)congruence on either the emotive or the identity dimension. The earliest sign of task-specific processing observed was a reduction in P2 amplitude for congruent relative to incongruent voice pairs, in the emotion matching task, but only for “happy” tones. In line with previous emotional priming studies (Campanella, Quinet, Bruyer, Crommelinck, & Guerit, 2002; Werheid, Alpay, Jentsch, & Sommer, 2005), we view this amplitude attenuation as an index of reduced depth of processing. By contrast, an effect of task-specific congruency emerged almost 150 ms later during performance of the voice identity matching task. Overall, this pattern of results implies that task-relevant auditory feature integration occurs more rapidly for emotion-based than for identity-based cues. It also indicates that the sensory analysis of at least some of the relevant features takes place within the first 150 ms.

The amplitude of the auditory P2 is known to be sensitive to the spectral composition of the eliciting sound waveform (Kuriki, Ohta, & Koyama, 2007). Except for pure sine tones, acoustical stimuli

consist of a number of distinct frequencies. The ‘spectrum’ of frequencies comprising a sound gives the sound its characteristic tonal quality, i.e., its timbre (McAdams, Winsberg, Donnadieu, de Soete, & Krimphoff, 1995). For example, increasing the intensity of higher frequencies relative to the lower frequencies results in an increasing perception of sound ‘brightness’, gradually turning into ‘sharpness’. In studies of timbre processing, P2 amplitudes were found to increase with the number of frequencies present in instrumental tones (Meyer, Baumann, & Jancke, 2006; Shahin, Bosnyak, Trainor, & Roberts, 2003).

In the present experiment the P2 amplitude was larger in response to happy compared to sad tones. An acoustical analysis of the sound stimuli using PRAAT (Boersma & Weenink, 2005) revealed that more high frequencies are present in our happy voice than in our sad voice stimuli in line with the perceived brighter timbre of the happy voices. The analysis also revealed a relatively higher amount of energy within the first third of the tone’s development in happy compared to sad tones which, acoustically, results in the perception of a sharp tone onset (‘attack’). Both, ‘brightness’ and sharp ‘attack’, have previously been identified as characteristic of vocal expressions of happiness (Gobl & Ni Chasaide, 2003; Juslin & Laukka, 2003). The earlier ERP effect for happy voices was paralleled by a shorter mean recognition time for the happy compared to sad tones in the emotion matching task. Shorter reaction times for positive compared to negative valence auditory stimuli, for example violin tones, have been reported by other studies as well (Goydke, Altenmüller, Möller, & Münte, 2004; Schirmer, Zysset, Kotz, & von Cramon, 2004). It thus seems that the acoustic features correlated with a perception of happiness are available early in the acoustic signal and can thus be extracted faster and more easily from the tones than can those correlated with a perception of sadness. These parameters of the acoustic signal thus may serve as early cues for emotional significance and accordingly may facilitate task-specific early sensory processing. The higher spectral complexity of the happy voice stimuli in the present study might thus account for the larger amplitude of P2 compared to the sad voice stimuli in both tasks. It might also serve as explanation for the fact that attenuation of the P2 to happy voices preceded by other happy voices was only present in the emotion matching task but not if the same stimulus pair was presented in the identity matching task. According to hierarchical models of auditory attention (Hansen & Hillyard, 1983) the analysis of simple features of a complex auditory stimulus starts with task-relevant features first. If the task-relevant feature matches one’s expectations and affords a decision, then there is no further processing of (less-relevant) features. The presence of high frequencies early in the timbre spectrum of the happy tones could have served as an early cue that the tone was happy, obviating any further acoustic analysis during the emotion matching task. The reduced P2 amplitude may reflect the discontinued processing of unnecessary acoustic features. In contrast, no such early cue was available for performance of the identity match task.

Our results are consistent with functional imaging data suggesting that emotional significance may alter early sensory processing of vocal input. fMRI data have localized these effects to the temporal lobes (Sander et al., 2005). A linear increase of the hemodynamic response with increasing emotional intensity of happy and angry intonation has been observed in the middle part of the sulcus temporalis superior (mid-STS) (Ethofer, Anders, Wiethoff et al., 2006), an area known as the voice-sensitive area of the human brain (Belin & Zatorre, 2003; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). In a study probing the processing time course of facial expression Werheid et al. (2005) found a priming effect as early as 100 ms post picture onset. The authors suggest that pre-wired configurational representations of prototypical emotional expressions might have facilitated feature analyses. However, to what ex-

tent these early priming effects reflect *emotional* processing remains an open question. In their model of the perception of facial emotions, Eimer and Holmes (2007) suggest that processing steps reflected by early ERP effects to emotional expression are most likely paralleled and preceded by categorization processes of potentially significant content within the limbic parts of the brain, e.g. the amygdala. A recent imaging study reporting amygdala activation in response to positive and negative vocalizations compared to neutral ones (Fecteau, Belin, Joannette, & Armony, 2007) indicates that this may also be true for the auditory domain. However, further research is needed to disentangle the differential roles of the participating structures.

Taken together, the present experiment provides some interesting information on the time course of neural events underlying the processing of acoustic cues that are unarguably important for social interaction: recognizing a person from their voice, and making inferences about a speaker's emotional state. However, given the relatively low number of correct trials obtained by our participants a word of caution is in order. Though our data indicate that paralinguistic features encoding different pieces of information about a speaker are processed differently depending on the specific information that the listener is focusing on, more data is needed including different kinds of expressive voice stimuli (words, syllables, tones) to gain a differentiated understanding of paralinguistic feature processing in the brain.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bandc.2008.06.003.

References

- Antinoro, F., Skinner, P. H., & Jones, J. J. (1969). Relation between sound intensity and amplitude of the AER at different stimulus frequencies. *Journal of the Acoustical Society of America*, 46(6), 1433–1436.
- Banase, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3), 614–636.
- Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, 8(3), 129–135.
- Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport*, 14(16), 2105–2109.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403(6767), 309–312.
- Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer (Version 4.3.14).
- Campanella, S., Quinet, P., Bruyer, R., Crommelinck, M., & Guert, J. M. (2002). Categorical perception of happiness and fear facial expressions: An ERP study. *Journal of Cognitive Neuroscience*, 14(2), 210–227.
- Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia*, 45(1), 15–31.
- Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., et al. (2006). Cerebral pathways in processing of affective prosody: A dynamic causal modeling study. *Neuroimage*, 30(2), 580–587.
- Ethofer, T., Anders, S., Wiethoff, S., Erb, M., Herbert, C., Saur, R., et al. (2006). Effects of prosodic emotional intensity on activation of associative auditory cortex. *Neuroreport*, 17(3), 249–253.
- Fecteau, S., Belin, P., Joannette, Y., & Armony, J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage*, 36(2), 480–487.
- Gobl, C., & Ni Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech communication*, 40, 189–212.
- Goydke, K. N., Altenmüller, E., Möller, J., & Münte, T. F. (2004). Changes in emotional tone and instrumental timbre are reflected by the mismatch negativity. *Brain Research Cognitive Brain Research*, 21(3), 351–359.
- Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., et al. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, 8(2), 145–146.
- Hansen, J. C., & Hillyard, S. A. (1983). Selective attention to multidimensional auditory stimuli. *Journal of Experimental Psychology Human Perception and Performance*, 9(1), 1–19.
- Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H., et al. (1997). Vocal identification of speaker and emotion activates different brain regions. *Neuroreport*, 8(12), 2809–2812.
- Johnstone, S. J., Barry, R. J., Anderson, J. W., & Coyle, S. F. (1996). Age-related changes in child and adolescent event-related potential component morphology, amplitude and latency to standard and target stimuli in an auditory oddball task. *International Journal of Psychophysiology*, 24(3), 223–238.
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129(5), 770–814.
- Kuriki, S., Ohta, K., & Koyama, S. (2007). Persistent responsiveness of long-latency auditory cortical activities in response to repeated stimuli of musical timbre and vowel sounds. *Cerebral Cortex*, 17(11), 2725–2732.
- McAdams, S., Winsberg, S., Donnadieu, S., de Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, 58(3), 177–192.
- Meyer, M., Baumann, S., & Jancke, L. (2006). Electrical brain imaging reveals spatio-temporal dynamics of timbre perception in humans. *Neuroimage*, 32(4), 1510–1523.
- Näätänen, R., Sams, M., Alho, K., Paavilainen, P., Reinikainen, K., & Sokolov, E. N. (1988). Frequency and location specificity of the human vertex N1 wave. *Electroencephalography and Clinical Neurophysiology*, 69(6), 523–531.
- Picton, T. W., Goodman, W. S., & Bryce, D. P. (1970). Amplitude of evoked responses to tones of high intensity. *Acta Oto-laryngologica*, 70(2), 77–82.
- Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., et al. (2005). Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody. *Neuroimage*, 28(4), 848–858.
- 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., Zysset, S., Kotz, S. A., & von Cramon, Y. D. (2004). Gender differences in the activation of inferior frontal cortex during emotional speech perception. *Neuroimage*, 21(3), 1114–1123.
- Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *Journal of Neuroscience*, 23(13), 5545–5552.
- Traunmüller, H. (1997). Perception of speaker sex, age, and vocal effort. *Phonum*, 4, 183–186.
- von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Research Cognitive Brain Research*, 17, 48–55.
- Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRI study. *Neuron*, 30(3), 829–841.
- Werheid, K., Alpay, G., Jentzsch, I., & Sommer, W. (2005). Priming emotional facial expressions as evidenced by event-related brain potentials. *International journal of psychophysiology*, 55(2), 209–219.
- Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, 8(2), 227–233.
- Winkler, I., Takegata, R., & Sussman, E. (2005). Event-related brain potentials reveal multiple stages in the perceptual organization of sound. *Brain Research Cognitive Brain Research*, 25(1), 291–299.
- Woldorff, M. G., & Hillyard, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography and Clinical Neurophysiology*, 79(3), 170–191.