Available online at www.sciencedirect.com

SCIENCE DIRECT®

Cognitive Brain Research 21 (2004) 351–359

COGNITIVE BRAIN RESEARCH

www.elsevier.com/locate/cogbrainres

Research report

# Changes in emotional tone and instrumental timbre are reflected by the mismatch negativity

Katja N. Goydke[a], Eckart Altenmüller[a], Jürn Möller[b], Thomas F. Münte[b,*]

[a]Institute for Music Physiology and Music Medicine, Hannover School of Music and Drama, Hannover, Germany
[b]Department of Neuropsychology, University of Magdeburg, Universitätsplatz 2, Gebäude 24, 39106 Magdeburg, Germany

Accepted 14 June 2004
Available online 23 July 2004

## Abstract

The present study examined whether or not the brain is capable to preattentively discriminate tones differing in emotional expression or instrumental timbre. In two event-related potential (ERP) experiments single tones (600 ms) were presented which had been rated as happy or sad in a pretest. In experiment 1, 12 non-musicians passively listened to tone series comprising a frequent (standard) single musical tone played by a violin in a certain pitch and with a certain emotional connotation (happy or sad). Among these standard tones deviant tones differing in emotional valence, either in instrumental timbre or in pitch were presented. All deviants generated mismatch negativity (MMN) responses. The MMN scalp topography was similar for all of the three deviants but latency was shorter for pitch deviants than for the other two conditions. The topography of the mismatch responses was indistinguishable. In a second experiment, subjects actively detected the deviant tones by button press. All detected deviants generated P3b waves at parietal leads. These results indicate that the brain is not only able to use simple physical differences such as pitch for rapid preattentive categorization but can also perform similar operations on the basis of more complex differences between tones of the same pitch such as instrumental timbre and the subtle timbral differences associated with different emotional expression. This rapid categorization may serve as a basis for the further fine-grained analysis of musical (and other) sounds with regard to their emotional content.

© 2004 Elsevier B.V. All rights reserved.

*Theme:* Sensory systems
*Topic:* Auditory systems: control physiology

*Keywords:* Music; Emotion; Brain; Tones; Deviant; Mismatch negativity; Pitch; Timbre

## 1. Introduction

In addition to their factual content, language and music often convey emotional information as well. In the speech domain, lesion studies indicate that the comprehension of the semantic content of an utterance and the understanding of affective prosody can be selectively impaired in the sense of a double dissociation [2]. In addition, it has been shown that affective prosody is independently processed from ”syntactic prosody” conveying information about the type of utterance (e.g., question, declarative sentence, or exclamation [14], although the exact neuroanatomical structures supporting the processing of affective and syntactic prosody are far from clear [8]. Animals, too, express emotions via distinct sounds [13,21,30] and the emotional state of a calling animal can be recognized by the specific acoustic structure of certain calls. The same acoustic features are used by different species to communicate emotions [34]. Studies in man aiming to link distinct vocal cues in spoken sentences to perceived emotions have revealed that the rating was mostly influenced by the mean level and the range of the fundamental frequency (F0) [36,41,49]. Low mean F0 was generally related to sadness and high mean F0

---

* Corresponding author. Tel.: +49 391 6718469; fax: +49 391 6711947.
*E-mail address:* thomas.muente@medizin.uni-magdeburg.de (T.F. Münte).

0926-6410/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogbrainres.2004.06.009

level to happiness. Increase of the F0 range was generally associated with high arousal.

In the music domain, a seminal series of experiments by Hevner [15–17] investigated which structural features contribute to the emotional expression conveyed by a piece of music. By systematically manipulating individual factors within the same musical pieces, she concluded that tempo and mode had the largest effects on listeners' judgements, followed by pitch level, harmony and rhythm [17]. In more recent work, Juslin [22] summarized the musical features supporting the impression of sadness (slow mean tempo, legato articulation, small articulation variability, low sound level, dull timbre, large timing variations, soft duration contrasts, slow tone attacks, flat micro-intonation, slow vibrato and final ritardando) and happiness (fast mean tempo, small tempo variability, staccato articulation, large articulation variability, fairly high sound level, little sound level variability, bright timbre, fast tone attacks, small timing variations, sharp duration contrasts and rising micro-intonation).

Many of these features describe changes in the structure of a musical sequence and it has been suggested that the emotional information transported by such suprasegmental features emerges as the result of a lifelong sociocultural conventionalization [43]. Recent studies show that listeners can accurately identify emotions in musical pieces from different cultures [1], however. In contrast, it has been suggested that the appraisal of segmental features [42], i.e., individual sounds or tones, is based on innate symbolic representations which have emerged from evolutionarily mechanisms for the evaluation of vocal expression [22,42]. For opera singers, Rapoport [38], based on spectrogram analyses, has described seven factors that contribute to the emotional expression of single tones:

(1) onset of phonation (voicing);
(2) vibrato;
(3) excitation of higher harmonic partials;
(4) transition—a gradual pitch increase from the onset to the sustained stage;
(5) sforzando—an abrupt pitch increase at the very onset of the tone;
(6) pitch change within the tone; and
(7) unit pulse (a feature produced by the vocal cords).

Many of these features can be mimicked by string and wind instruments, while keyboard instruments are less versatile with respect to the modulation of individual tones.

The variations induced in single tones of the same pitch fall within the realm of timbre. Timbre refers to the different quality of sounds in the absence of differences in pitch, loudness and duration. The classical view of timbre, dating back to von Helmholtz [48], holds that different timbres result from different distributions of amplitudes of the harmonic components of a complex tone in a steady state. More recent studies show that timbre also involves more dynamic features of the sound [9,12], particularly with regard to onset characteristics. Timbre has been mostly studied with regard to the recognition of different musical instruments [9–12,27] and multidimensional scaling techniques have revealed that timbre is determined by variations along three dimensions termed attack time, spectral centroid, and spectral flux [27].

Clearly, the timbral variations within a single instrument that are used to transmit emotional expressions are different and are likely smaller than those that are present between instruments. The present study therefore asks whether the brain mechanisms of detecting the timbral variation between notes of different emotional expression played by the same instrument are similar to or different from the variations between instruments playing the same note with the same emotional expression.

Given the importance of emotions for survival, we assumed that the brain may accomplish a fast and probably automatic check [40] on every incoming stimulus with regard to the properties correlated with emotional expression. In the present investigation, we used musical stimuli as a tool to demonstrate the existence of such a fast and automatic checking procedure by employing a mismatch negativity paradigm.

### 1.1. The brain's machinery for auditory change detection

In order to address the early, automatic stages of sound evaluation, the mismatch negativity (MMN) is an ideal tool [32,33,35]. The MMN is a component of the auditory event-related potential (ERP) which is elicited during passive listening by an infrequent change in a repetitive series of sounds. It occurs in response to any stimulus which is physically deviant (in frequency, duration or intensity) to the standard tone. It has also been demonstrated that the MMN is sensitive to changes in the spectral component of tonal timbre [44]. Toiviainen et al. [46] have shown that the amplitude of the MMN obtained for different timbre deviants corresponded to the distance metric obtained in an artificial neural network trained with a large set of instrumental sounds.

The onset latency of the MMN varies according to the nature of the stimulus deviance but for simple, physically deviant stimuli lies at approximately 150 ms. Previous studies have led to the assumption that the MMN reflects the mismatch resulting from a comparison between the physical features of the deviant and the standard stimulus [32]. This implies the existence of a neural sensory–memory trace representing the physical structure of the standard stimulus against which incoming auditory information can be compared. More recent studies (see Refs. [33,35] for a review) have shown, however, that the MMN can also be obtained to deviations within complex series of sounds, suggesting that the memory trace is not only dependent on the physical characteristics of the stimuli but can also contain more abstract properties such as the order of stimuli.

The sensory analysis of the incoming stimulus as well as its encoding appears to take place automatically because the

MMN typically occurs when the subjects do not attend to the eliciting stimuli and are involved in a different task like reading a book [32] or when they are sleeping [26].

The P300 is also evoked by infrequent deviant stimuli, but in contrast to the MMN, it is triggered most effectively when the deviant events are attended and task-relevant [6,31,47]. It is assumed that the P300 is not a unitary component but can be broken down to several subcomponents, one of which is termed P3b. The P3b occurs in response to task-relevant deviant stimuli within a stream of standard stimuli, a sequence known as oddball paradigm. The P3b displays a parietal distribution, the onset latency varies between 300 and 600 ms. Latency and amplitude of the P3b depend on the difficulty of the categorisation task as well as on the task-relevance of the stimulus [20,24]. Thus, the P3b appears to reflect stimulus evaluation and stimulus categorisation processes. It has further been suggested that the underlying processes serve the updating of working memory [7] although not everyone agrees on this interpretation [47].

### *1.2. The current study*

In the current study, two experiments were conducted to assess whether the emotional expression of a single tone allows for attentive as well as preattentive categorization. For that purpose, a standard violin tone of a certain emotional valence (e.g., happy) was presented repeatedly, infrequently interspersed with a tone that deviated from the standard according to its emotional expression (e.g., sad). In addition to this emotional deviant, a tone which differed from the standard tone in pitch level (pitch deviant) and a tone which was played by a flute instead of a violin and therefore differed from the standard stimulus according to instrumental timbre (instr. deviant) were introduced as control stimuli. In experiment 1 (Exp. 1), subjects watched a video and were asked to ignore the sounds (passive condition). In experiment 2 (Exp. 2), a modified oddball paradigm was conducted with subjects required to react to any of the three deviant stimulus types by pressing a button (active condition).

## 2. Methods

### *2.1. Subjects*

Twelve non-musicians participated in the experiment (11 women, 20 to 36 years of age, mean=26). All participants were right-handed, neurologically healthy and had normal hearing.

### *2.2. Stimuli*

Two sets of four different tones were used. Each set consisted of one standard tone and three different deviant tones. All tones were played by a violinist and a flutist, digitally recorded, and edited to equal length (600 ms) and sound level (65 dB) using *cool edit*. These edited tones were rated by 10 naive listeners using a 7-point scale ($-3$=very sad, 0=neutral, +3=very happy). Tones used for the experiment had a mean score of $>1.7$ for the happy and smaller than $-1.7$ for the sad conditions.

In set 1, the standard tone consisted in a violin /c/ played in a happy way. This frequent “happy standard” was combined with a rare violin /c/ played in a sad way (“sad deviant”), a rare flute /c/ played in a happy way (“instr. deviant”) and a happy violin /a/ (“pitch deviant”).

For set 2, the sad violin /c/ was used as a standard (“sad standard”) and combined with the following deviants: happy violin /c/ (“happy deviant”), sad flute /c/ (“instr. deviant”) and sad violin /a/ (“pitch deviant”). A spectrogram of the stimuli is shown in Fig. 1.

In the passive condition, two video films (“Les vacances de monsieur Hulot” and “Playtime”, both by Jacques Tati) were presented to the participants with the sound turned off. In order to minimize eye movements, a small video screen (18″) at a viewing distance of 130 cm was used.

### *2.3. Design*

Each subject participated in two different experiments. The experiments were conducted on two different days separated by at least 1 week. Each experiment consisted of two consecutive blocks which differed with regard to the stimulus set used. The order of the two stimulus sets was kept stable for each participant between experiment 1 and 2 but was counterbalanced between subjects. In experiment 1 (passive condition), participants watched a video while the stimulus tones were played in the background. No response to the tones was required. In experiment 2 (active condition), participants held a joy stick in one hand and pressed a button with their index finger in response to any deviant tone. The use of the right or the left hand was counterbalanced between all participants. The order of experiment 1 and 2 was also counterbalanced.

### *2.4. Procedure*

Participants were tested individually while seated in a soundproof chamber in front of a computer screen which was replaced by a television set in the passive condition (Exp. 1).

In each condition, 2600 tones were played to the participants via loud speaker. A series of standard tones was presented, interrupted randomly by emotionally deviant, by instr. deviant, or pitch deviant stimuli. The probability of occurrence was 76.9% for the standard tone and 7.7% for each of the deviant tones. The interstimulus interval was randomised between 400 and 900 ms. No test trials were given but the first 20 trials of each block were excluded from the analysis.

Every 10 min, there was a short break and a longer 15-min-break was taken between the two blocks. Each

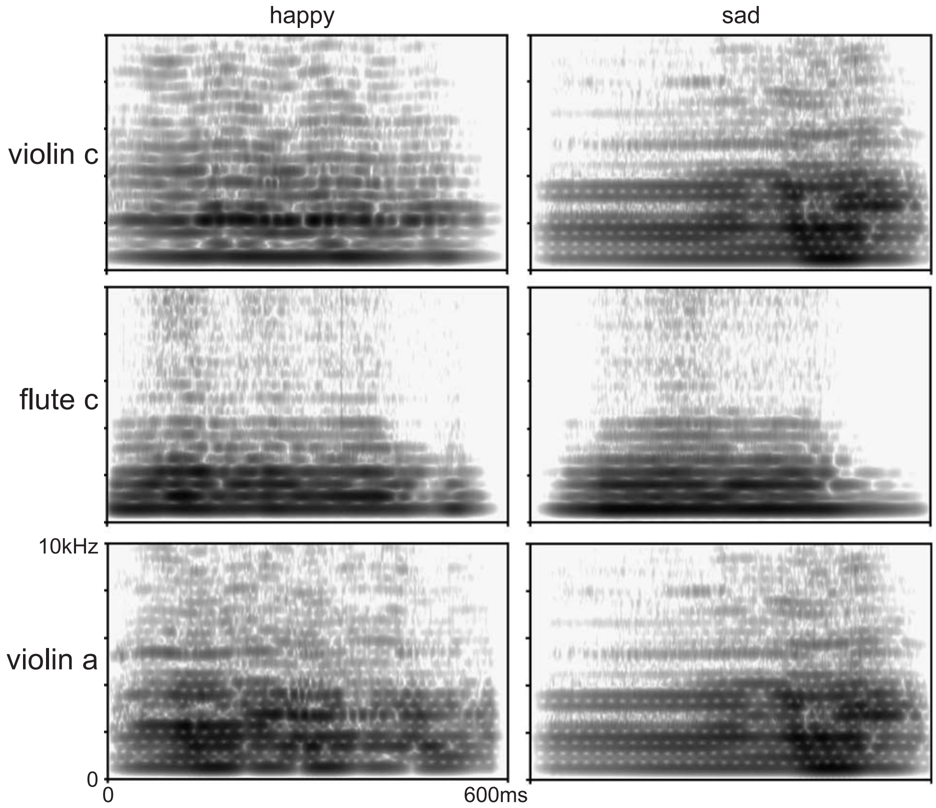

Fig. 1. Spectrograms of stimuli. Note that the legends of $x$- and $y$-axis pertain to all six diagrams.

experimental block lasted about 55 min. One entire experiment lasted about 2.5 h.

In Exp. 1 (passive condition), participants were instructed to watch the video carefully because they would be asked about it later. Following each block, three questions relating to the content of the film were asked by the experimenter that had to be answered by the subject.

In Exp. 2 (active condition), participants were instructed to press a button as fast as possible in response to a deviant tone. During the experiment, the participants looked at a fixation point in the centre of the computer screen.

In both experiments, participants were asked not to speak and to blink or move their eyes as little as possible.

### 2.5. Apparatus and recording

In experiment 2, push-button response latencies were measured from sound onset, with the timeout point (the moment in time after which responses were registered as missing) set at 400 ms poststimulus offset. Timeouts and errors, i.e., wrong responses, were excluded from further analyses. The EEG was recorded from 30 scalp sites using tin electrodes mounted in an electrode cap with reference electrodes placed at the left mastoid and the tip of the nose. Signals were collected using the left mastoid electrode as a reference and were re-referenced offline to the nose electrode. Blinks and vertical eye movements were monitored by a bipolar montage using an electrode placed on the left lower orbital ridge and Fp1. Lateral eye movements were monitored by a bipolar montage using two electrodes placed on the right and left external canthus. The eye movements were recorded in order to allow for later offline rejection. Electrode impedance was kept below 5 kΩ for the EEG and eye movement recording. The EEG was sampled with a Brainlab system (Schwarzer, Munich). Signals were amplified with a notch filter and digitized with 4-ms resolution. Averages were obtained for 1024 ms epochs including a 100-ms prestimulus baseline period. Trials contaminated by eye movements or amplifier blocking within the critical time window were rejected from averaging by a computer program using individualised rejection criteria. On average, 11 % of the trials were excluded from further analysis.

ERPs were quantified by mean amplitude and peak latency measures using the mean voltage of the 100-ms period preceding the onset of the stimulus as a reference. Time windows and electrode sites are specified at the appropriate places of the result section.

Topographical distributions of the ERP effects were compared by ANOVA designs, with condition (emotion, timbre, pitch) and electrode site (28 levels) as factors. Before computing the statistics, the amplitudes were vector normalised according to the method described by McCarthy and Wood [28].

The Huynh–Feldt epsilon correction [18] was used to correct for violations of the sphericity assumption. Reported

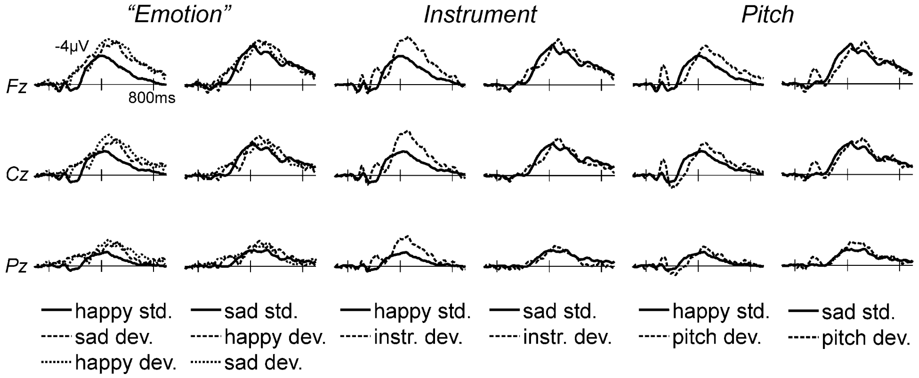

Fig. 2. Grand average ERPs from the passive experiment for three midline electrodes. This experiment was carried out in two versions with either a happy or a sad violin /c/ used as a standard stimulus. Therefore, two columns are presented for each condition (emotion, instrument, pitch) showing the standard and the respective deviant. In the emotion condition, in addition to the deviant differing emotionally from the standard (e.g., rare sad violin /c/ for happy violin /c/ standard), the deviant from the other version (physically identical to the standard stimulus) is presented as well in the same figure. The pitch condition shows a typical phasic MMN with a latency of 140 ms, while the emotion and timbre deviants were associated with a later mismatch response. All three conditions also showed an extended negativity to the deviant stimuli approximately between 400 and 700 ms.

are the original degrees of freedom and the corrected $p$-values.

## 3. Results

### 3.1. Passive condition

Fig. 2, left, shows the grand average waveforms for all three deviant types at three scalp positions (Fz, Cz, Pz). Note that the results from the two blocks, using the happy and the sad violin tone as standard stimuli respectively, are given in separate columns. The waveforms show an initial small negative deflection (N1) at around 100 ms. This is followed by a long-duration negative component with a frontal maximum and a peak around 400 to 500 ms (Fig. 3).[1] The current design allows two different ways to compare emotional deviants. Firstly, deviants and standards collected in the same experimental blocks can be compared. These stimulus classes are emotionally as well as physically different. Secondly, deviants and standards can be compared across blocks such that the same physical stimulus serves as standard and deviant. Regardless of the comparison (Fig. 2, columns 1 and 2), emotional deviants elicited a more negative waveform in the 150–300 ms latency range. Thus, the mismatch response cannot be explained by the fact that different tones elicited the different ERP waveforms. The MMN evoked by instrument deviants is shown in Fig. 2, columns 3 and 4. Finally, stimuli deviating in pitch evoked an early MMN which was of similar size and morphology for 'happy' and 'sad' stimuli (Fig. 2, columns 5 and 6). Statistical analyses (Table 1) show significant effects for pitch deviants in the 100–150 ms time window, whereas effects for emotion and instrument appeared only later, regardless of emotionally deviant stimuli, were compared to the physically identical standard stimulus from the other experimental block or to the standard stimulus of the same block.

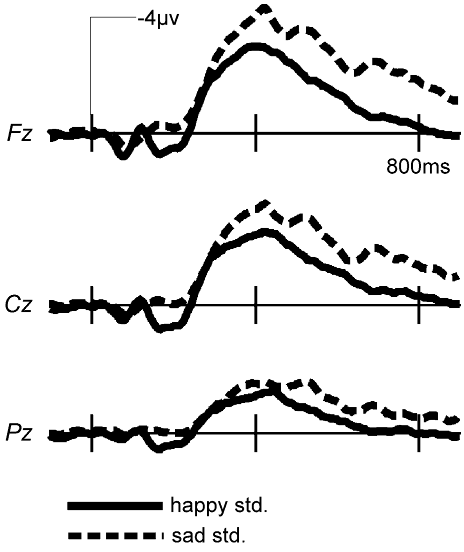

Fig. 3. Comparison of the two types of standard stimuli, violin happy /c/ and violin sad /c/, used in the two blocks of the passive task. The sad stimuli are associated with a higher amplitude tonic negativity (see Footnote 1).

To isolate mismatch-related brain activity, deviant minus standard difference waves were computed (Fig. 4). These

---

[1] This negativity is not seen in most MMN studies. One has to bear in mind, however, that in the current experiment, tones with duration of 600 ms were used. Such longer stimuli are known to give rise to a long-standing, tonic negativity [23]. Inspection of the ERPs to the happy and sad standard stimuli suggests that these are different, especially with regard to this long-standing negativity. In Fig. 3, these two ERPs are compared directly. Statistical analyses (successive 100 ms time-windows, Fz/Cz/Pz electrodes) indicated a significant difference between sad and happy tones primarily for the tonic negativity (100–200 ms, $F(1,11)=1.78$, n.s.; 200–300 ms, $F=3.42$, n.s.; 300–400 ms, $F=5.1$, $p<0.05$; 400–500 ms, $F=6.77$, $p=0.024$; 500–600 ms, $F=6.32$, $p=0.029$; 600–700 ms, $F=8.87$, $p=0.013$; 700–800 ms, $F=9.3$, $p=0.011$).

Table 1
Passive experiment; Comparison of standard vs. deviant stimuli; given are the $F$-values ($df$=1,11)

| Comparison | Standard | Deviant | 100–150 ms | 150–200 ms | 200–250 ms | 250–300 ms |
|---|---|---|---|---|---|---|
| Emotion | Happy | Happy | 0.10 | 2.72 | 22.75** | 0.24 |
| Emotion | Happy | Sad | 1.33 | $9.64^{+}$ | $11.28^{+}$ | 3.38 |
| Emotion | Sad | Sad | 1.63 | $6.55^{+}$ | $7.47^{+}$ | 2.72 |
| Emotion | Sad | Happy | 0.19 | 0.06 | 12.02* | 0.24 |
| Instrumental | Happy | Happy | 0.22 | 3.64 | 25.25** | 0.25 |
| Instrumental | Sad | Sad | 0.47 | 0.01 | 3.84 | 0.5 |
| Pitch | Happy | Happy | 10.10* | 2.72 | 22.75** | 17.43** |
| Pitch | Sad | Sad | $4.97^{+}$ | $7.62^{+}$ | 0.13 | 1.1 |

* $p<0.01$.
** $p<0.001$.
$^{+}$ $p<0.015$.

difference waves showed an initial negative peak, identified as the MMN, which was followed by a phasic positivity and finally, the tonic negativity mentioned above. The MMN for the different conditions appeared to differ markedly in latency. This was confirmed statistically by determining the peak latency of the most negative peak in the 100 to 300 time window [Cz site, $F(2,22)=20.3$, $p<0.001$]. Post hoc tests revealed a significant difference between the peak latencies in the pitch and emotion conditions ($p<0.001$) and between pitch and instrument conditions ($p<0.001$). There was no difference between the emotion and instrument conditions, however ($p>0.2$).

While the latency of the negativity was very different for the different classes of deviant stimuli, the distribution of all three effects was virtually identical and typical for the MMN, as illustrated by spline-interpolated isovoltage maps (see Fig. 4, right panel). This was corroborated by an analysis on the vector-normalized [28] mean amplitudes (taken in 40 ms time windows centred upon the peak latency of the negativity in each condition) which revealed no condition by electrode site [$F(27,297)=0.16$, n.s.] interaction.

### 3.2. Active condition

#### 3.2.1. Behavioural results

The level of performance was nearly perfect for all deviant target stimuli (misses<1%) as well as for the standards (false alarms<1%). Differences in mean reaction times (see Table 2) between different types of deviants were only apparent when the standard tone was a happy tone [$F(2,22)=22.45$, $p<0.001$]. Post hoc comparison (Scheffé) revealed that in this condition, the mean reaction to the emotional deviant (sad violin tone) was slower than to the pitch deviant ($p<0.001$) and to the instr. deviant ($p<0.001$).

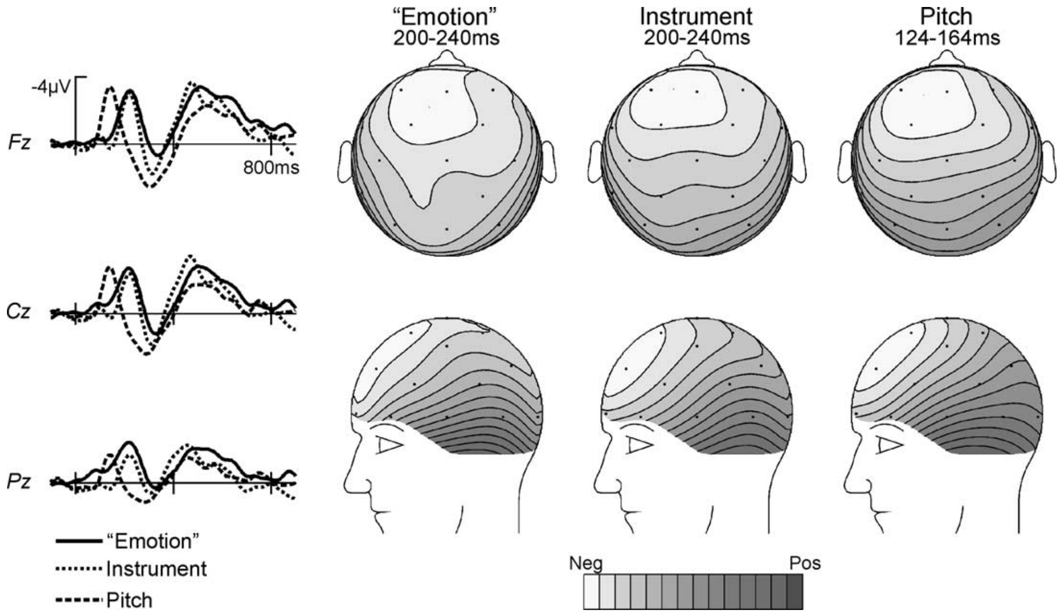

Fig. 4. Deviant minus standard difference waves. For these waveforms, data from both versions of the passive task (violin happy /c/ standard and violin sad /c/ standard) were averaged together. All three conditions show an initial negativity differing in latency. The scalp distribution of this negativity is shown on the right side using spline-interpolated isovoltage maps. These maps are based on the mean voltage in the 40 ms time window centered upon the peak latency of the negativity. The distribution of the negativities from the three conditions is virtually identical.

Table 2
Reaction times (ms) to deviant stimuli in the active experiment

| | Block I standard happy | | | Block II standard sad | | |
|---|---|---|---|---|---|---|
| | Emotion | Instrumental | Pitch | Emotion | Instrumental | Pitch |
| Mean ($N$=9) | 527 | 383 | 406 | 449 | 472 | 470 |
| S.D. | 107 | 93 | 115 | 104 | 107 | 118 |

When the standard was a sad tone, no RT differences were found [$F(2,22)=0.341$].

### 3.2.2. ERP data

Fig. 5 shows the ERPs to the target stimuli (Pz electrode site) separately for the happy and the sad version of each deviant. In the emotion condition, the P3b appears to peak much earlier for the happy deviant than for the sad deviant. In the instr. condition, a latency difference in the same direction is suggested upon visual inspection.

The peak latency was quantified in the time window between 300 and 550 ms for the Pz electrode site and subjected to ANOVA with factors condition (emotion vs. instr. vs. pitch) and deviant (sad vs. happy). A main effect of condition was found [$F(2,22)=7.04$, $p<0.005$] reflecting the fact that the P3b was longest in the emotion condition (460 ms, S.D.=85), followed by the instr. (402 ms, S.D.=68) and pitch (383 ms, S.D.=62) conditions. Moreover, a main effect of deviant was also found [$F(1,11)=8.7$, $p<0.015$] reflecting the overall longer latency of sad compared to happy deviants (369 ms, S.D.=81, vs. 441 ms, S.D.=81). The significant condition by deviant interaction [$F(2,22)=8.02$, $p<0.005$] indicated that the latency difference between sad and happy deviants was most pronounced in the emotion condition.

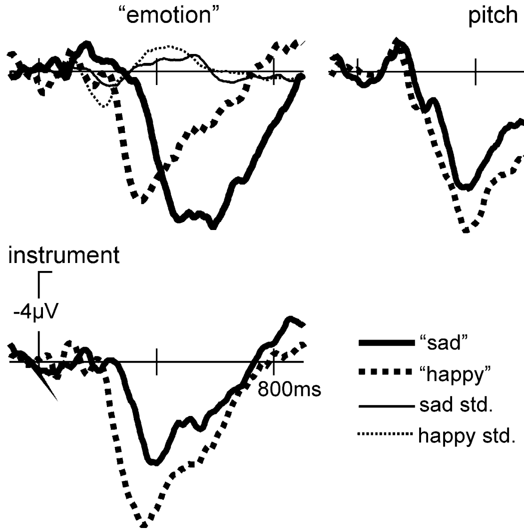

Fig. 5. ERPs from the active experiment for the emotion (top), timbre (middle), and pitch (bottom) conditions (Pz electrode site). In the emotion condition, the latency of the P3 component was dependent on the deviant. A sad violin /c/ target (among violin happy /c/ standards) was associated with a delayed P3 compared to a violin happy /c/ target (among violin sad /c/ standards).

## 4. Discussion

In this study, we used the high temporal resolution of electrophysiological measures to estimate the relative time courses of the brain's response to tones that differed from a standard tone by their emotional expression, by the timbre of the instrument used and by their pitch. The results demonstrate that affective deviants evoke a mismatch response even when subjects do not attend the auditory stimuli akin to the mismatch negativity that was seen for pitch and instrumental deviants. While the peak latency of the mismatch effects to the affective and instrumental deviants was delayed by about 80 ms, the scalp distribution of the three mismatch effects was virtually identical on visual inspection (Fig. 4) and was statistically indistinguishable. In addition, in the active condition, a P3b occurred in response to all three deviant types.

The question arises then, what aspect of the emotionally deviant stimuli triggers the mismatch response in the current study. The finding of a highly similar distribution of all three deviant stimuli suggests that all of these engage the same generators, which are known to reside in the supratemporal plane with additional contribution by frontal cortex [35,39,45]. This further indicates that it is not the emotional quality per se but rather the physical differences between the stimuli of different emotional quality that give rise to the mismatch response. While the finding reveals that tones which differ in physical structure evoke a mismatch negativity is trivial and has been shown repeatedly (see Refs. [32,33,35] for reviews), the current study shows that the subtle physical differences used to convey emotional expression in single musical notes are sufficient to trigger the brain's automatic mismatch response. This automatic detection early in the auditory processing stream at least allows the rapid classification of stimuli according to their emotional quality during further and more detailed auditory analysis that then could be restricted to the emotionally deviant stimulus. The present study does not allow us to determine whether the mismatch detection system indexed by the MMN component to emotional and instrumental deviants would be capable to extract physical invariants from a series of different tone stimuli that are characteristic for particular (standard) emotion. That complex regularities can be extracted from stimulus series has been demonstrated before [33], however. To answer this question, a study using many different happy tones as standards and a set of different sad tones as deviants would be needed.

Of relevance to this issue, Bostanov and Kotchoubey [4] compared brain responses to short joyful ("Yeeh!", "Heey!", "Wowh!", "Oooh!") exclamations to those to a single woeful ("Oooh!") vocalization, while subjects were required to "listen attentively" without a further task. These authors found a negative component between 200 and 400 ms for the woeful stimulus compared to the joyful stimuli, which was remarkably similar to the ERP effect found for emotional and instrumental deviants in the passive experiment of the current study. In the Bostanov and Kotchoubey [4] study, all five exclamations occurred equally often, however, such that the woeful stimulus could be considered deviant only if the brain had grouped the four joyful exclamations together. This implies that the invariant physical attributes characterizing the majority of the stimuli as joyful in the experiment must have been extracted by the auditory system, thereby allowing the differential processing of the single woeful stimulus.

While we are unaware of any brain imaging study using musical tones of varying emotional quality, a PET study [37] requiring the active discrimination of a subtle timbral aspect of musical stimuli (dull vs. bright oboe) identified the right superior and middle frontal gyrus as candidate regions supporting selective attention to timbre. Timbre-specific activations of temporal brain regions might have been missed in this study, however, because a comparison between selective attention to timbre vs. attention to pitch had been employed. Both of these tasks might have engaged the auditory cortex to a similar extent. Likewise, when attention to a specific target word or attention to a specific emotional tone was compared in a verbal dichotic listening task, no fMRI activation differences were found in the planum temporale and superior temporal sulcus [19].

A more recent fMRI study [29] comparing the brain responses to melodies played with two synthetic instrumental timbres revealed activation differences in the posterior Heschl's gyrus and superior temporal sulcus, i.e., areas that are involved in the initial analysis of incoming sounds. Importantly, in this study, the timbral difference was irrelevant for the task of the subjects, supporting our view that timbral aspects of sounds are processed early and automatic in the auditory system.

Thus, the results of the current study, in conjunction with earlier work, demonstrate that the brain is in possession of a tool for the preattentive analysis of auditory input that allows for a fast and automatic categorization not only according to simple physical characteristics but also according to more complex acoustic features like instrumental timbre and emotional expression. The speed of the detection indicates that the categorization happens automatically. Following Scherer [40], the result of this fast appraisal may serve as a basis for further evaluation processes, for example, the ultimate assignment of the correct emotion by secondary auditory and frontal areas [37] and the triggering of emotional and autonomous responses by limbic structures [3,5,25].

## Acknowledgements

We thank Dana Heinze and Monique Lamers for their help during recording and analysis of the data. Supported by grants of the DFG to EA and TFM.

## References

[1] L. Balkwill, W.F. Thompson, A cross-cultural investigation of the perception of emotion in music: psychophysical and cultural cues, Music Percept. 17 (1999) 43–64.

[2] A. Barrett, G. Crucian, A. Rayner, K. Heilman, Spared comprehension of emotional prosody in a patient with global aphasia, Neuropsychiatry Neuropsychol. Behav. Neurol. 12 (1999) 117–120.

[3] A.J. Blood, R.J. Zatorre, Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 11818–11823.

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

[5] M. Davis, The role of the amygdala in fear and anxiety, Annu. Rev. Neurosci. 15 (1992) 353–375.

[6] E. Donchin, Surprise?...Surprise!, Psychophysiology 18 (1981) 493–513.

[7] E. Donchin, M.G.H. Coles, Is the P300 component a manifestation of context updating? Behav. Brain Sci. 11 (1988) 357–374.

[8] A.D. Friederici, K. Alter, Lateralization of auditory language functions: a dynamic dual pathway model, Brain Lang. 89 (2004) 267–276.

[9] J.M. Grey, Multidimensional perceptual scaling of musical timbres, J. Acoust. Soc. Am. 61 (1977) 1270–1277.

[10] J.M. Grey, Timbre discrimination in musical patterns, J. Acoust. Soc. Am. 64 (1978) 467–472.

[11] J.M. Grey, J.W. Gordon, Perceptual effects of spectral modifications on musical timbres, J. Acoust. Soc. Am. 63 (1978) 1493–1500.

[12] J.M. Grey, J.A. Moorer, Perceptual evaluation of synthetic music instrument tones, J. Acoust. Soc. Am. 62 (1977) 454–462.

[13] M.D. Hauser, The Evolution of Communication, MIT Press, Cambridge, 1997, 776 pp.

[14] K. Heilman, D. Bowers, L. Speedie, H. Coslett, Comprehension of affective and non-affective prosody, Neurology 34 (1984) 917–921.

[15] K. Hevner, The affective character of the major and minor modes in music, Am. J. Psychol. 47 (1935) 103–118.

[16] K. Hevner, Experimental studies of the elements of expression in music, Am. J. Psychol. 48 (1936) 246–268.

[17] K. Hevner, The affective value of pitch and tempo in music, Am. J. Psychol. 49 (1937) 621–630.

[18] H. Huynh, L.A. Feldt, Conditions under which mean square ratios in repeated measure designs have exact $F$-distributions, J. Am. Stat. Assoc. 65 (1980) 1582–1589.

[19] L. Jäncke, T.W. Buchanan, K. Lutz, N.J. Shah, Focused and nonfocused attention in verbal and emotional dichotic listening: an FMRI study, Brain Lang. 78 (2001) 349–363.

[20] R. Johnson, A triarchic model of P300 amplitude, Psychophysiology 23 (1986) 367–384.

[21] U. Jürgens, Vocalization as an emotional indicator. A neuroethological study in the squirrel monkey, Behaviour 69 (1979) 88–117.

[22] P.N. Juslin, Communicating emotion in music performance: a review and theoretical framework, in: P.N. Juslin, J.A. Sloboda (Eds.), Music and Emotion—Theory and Research, University press, Oxford, 2001, pp. 309–337.

[23] W.D. Keidel, DC-potentials in auditory evoked response in man, Acta Oto-Laryngol. 71 (1971) 242–248.

[24] M. Kutas, G. McCarthy, E. Donchin, Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time, Science 197 (1977) 792–795.

[25] J.E. LeDoux, Emotion circuits in the brain, Annu. Rev. Neurosci. 23 (2000) 155–184.

[26] D.H. Loewy, K.B. Campbell, C. Bastien, The mismatch negativity to frequency deviant stimuli during natural sleep, Electroencephalogr. Clin. Neurophysiol. 98 (1996) 493–501.

[27] S. McAdams, S. Winsberg, S. Donnadieu, G. de Soete, J. Krimphoff, Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes, Psychol. Res. 58 (1995) 177–192.

[28] G. McCarthy, C.C. Wood, Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models, Electroencephalogr. Clin. Neurophysiol. 62 (1985) 203–208.

[29] V. Menon, D.J. Levitin, B.K. Smith, A. Lembke, B.D. Krasnow, D. Glazer, G.H. Glover, S. McAdams, Neural correlates of timbre change in harmonic sounds, NeuroImage 17 (2002) 1742–1754.

[30] E.S. Morton, On the occurrence and significance of motivational–structural rules in some bird and mammal sounds, Am. Nat. 111 (1977) 855–869.

[31] T.F. Münte, T.P. Urbach, E. Duezel, M. Kutas, Event-related brain potentials in the study of human cognition and neuropsychology. In F. Boller, J. Grafman, G. Rizzolatti, Handbook of neuropsychology, 2nd edition, Vol. 1, Elsevier, Amsterdam, pp. 139–235.

[32] R. Näätänen (Ed.), Attention and Brain Function, Erlbaum, Hillsdale, 1992, 494 pp.

[33] R. Näätänen, M. Tervaniemi, E. Sussman, P. Paavilainen, I. Winkler, “Primitive intelligence” in the auditory cortex, Trends Neurosci. 24 (2001) 283–288.

[34] D.H. Owings, E.S. Morton (Eds.), Animal Vocal Communication: A New Approach, Cambridge University Press, Cambridge, 1998, 296 pp.

[35] T.W. Picton, C. Alain, L. Otten, W. Ritter, A. Achim, Mismatch negativity: different water in the same river, Audiol. Neuro-otol. 5 (2000) 111–139.

[36] H. Pihan, E. Altenmüller, I. Hertrich, H. Ackermann, Cortical activation patterns of affective speech processing depend on concurrent demands on the subvocal rehearsal system—A DC-potential study, Brain 123 (2000) 2338–2349.

[37] H. Platel, C. Price, J. Baron, R. Wise, J. Lambert, R.S.J. Frackowiak, B. Lechevalier, F. Eustache, The structural components of music perception. A functional anatomical study, Brain 120 (1997) 229–243.

[38] E. Rapoport, Singing, mind and brain—unit pulse, rhythm, emotion and expression, in: M. Leman (Ed.), Music, Gestalt, and Computing: Studies in Cognitive and Systematic Musicology, Springer, Berlin, 1997, pp. 451–468.

[39] M. Sams, E. Kaukoranta, M. Hamalainen, R. Näätänen, Cortical activity elicited by changes in auditory stimuli: different sources for the magnetic N100m and mismatch responses, Psychophysiology 28 (1991) 21–29.

[40] K.R. Scherer, On the nature and function of emotion: a component process approach, in: K.R. Scherer, P. Ekman (Eds.), Approaches to Emotion, Erlbaum, Hillsdale, 1984, pp. 293–318.

[41] K.R. Scherer, On the symbolic function of vocal affect expression, J. Lang. Soc. Psychol. 7 (1988) 79–100.

[42] K.R. Scherer, Emotional effects of music: production rules, in: P.N. Juslin, J.A. Sloboda (Eds.), Music and Emotion—Theory and Research, Oxford University Press, Oxford, 2001, pp. 361–392.

[43] J.A. Sloboda, Empirical studies of the emotional response to music, in: M.R. Jones, S. Holleran (Eds.), Cognitive Bases of Musical Communication, American Psychological Association, Washington, 1990, pp. 33–46.

[44] M. Tervaniemi, I. Winkler, R. Näätänen, Pre-attentive categorization of sounds by timbre as revealed by event-related potentials, NeuroReport 8 (1997) 2571–2574.

[45] H. Tiitinen, K. Alho, M. Huotilainen, R.J. Ilmoniemi, J. Simola, R. Näätänen, Tonotopic auditory cortex and the magnetoencephalographic (MEG) equivalent of the mismatch negativity, Psychophysiology 30 (1993) 537–540.

[46] P. Toiviainen, M. Tervaniemi, J. Louhivuori, M. Saher, M. Huotilainen, R. Näätänen, Timbre similarity: convergence of neural, behavioral, and computational approaches, Music Percept. 16 (1998) 223–241.

[47] R. Verleger, Event-related potentials and cognition: a critique of the context updating hypothesis and an alternative interpretation of P3, Behav. Brain Sci. 11 (1988) 343–356.

[48] H.L.F. von Helmholtz, On the Sensations of Tone, Dover, New York, 1863/1954, (A.J. Ellis, Trans.).

[49] C.E. Williams, K.N. Stevens, Emotions and speech: some acoustical correlates, J. Acoust. Soc. Am. 52 (1972) 1238–1250.