available at www.sciencedirect.com
SCIENCE DIRECT
www.elsevier.com/locate/brainres
ELSEVIER
BRAIN RESEARCH

Research Report

The perception of musical phrase structure: A cross-cultural ERP study

Yun Nan[a,b,c], *Thomas R. Knösche*[a,c,*], *Angela D. Friederici*[a]

[a]*Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1A. 04103 Leipzig, Germany*
[b]*National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China*
[c]*Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China*

ARTICLE INFO

Article history:
Accepted 28 March 2006
Available online 18 May 2006

Keywords:
Music perception
Phrase structure
Cultural difference
Event-related potential

ABSTRACT

Electroencephalography (EEG) was used in a cross-cultural music study investigating phrase boundary perception. Chinese and German musicians performed a cultural categorization task under Chinese and Western music listening conditions. Western music was the major subject for both groups of musicians, while Chinese music was familiar to Chinese subjects only. By manipulating the presence of pauses between two phrases in the biphrasal melodies, EEG correlates for the perception of phrase boundaries were found in both groups under both music listening conditions. Between 450 and 600 ms, the music CPS (closure positive shift), which had been found in earlier studies with a false tone detection task, was replicated for the more global categorization task and for all combinations of subject group and musical style. At short latencies (100 and 450 ms post phrase boundary offset), EEG correlates varied as a function of musical styles and subject group. Both bottom–up (style properties of the music) and top–down (acculturation of the subjects) information interacted during this early processing stage.

1. Introduction

Music is a culturally specific phenomenon, which is certainly determined by ethnic background, social environment and traditions. However, the question to what extent the ability to understand and appreciate music is culturally dependent is still controversial. Ethnomusicologists usually assume that extensive experience of music within its cultural context is essential for its basic structural principles to be understood (Blacking, 1973). Lynch and Eilers (1992) explored musical tuning perception in infancy and adulthood and found that adult's perception of musical tuning was culturally specific, and infants between 6 and 12 months of age started to show the same culturally specific performance as adults did. In the same line, Drake and El Heni (2003) proved that acculturation, i.e., the passive exposure to a particular type of music since birth, influenced the perception and cognition of music. On the other hand, some psychological studies have proved the existence of universal, not culturally dependent, aspects of music perception (Castellano et al., 1984; Krumhansl, 1995, 2000). Drake and Bertrand (2001) reviewed a number of rhythm perception and production studies and concluded that some aspects of temporal processing in music might be universal, in the sense that they function in a similar manner irrespective of an individual's cultural exposure and experience. Likewise, in a study of the cross-cultural emotion perception in music,

* *Corresponding author.* Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1A. 04103 Leipzig, Germany. Fax: +49 3425 887511.
E-mail address: knoesche@cbs.mpg.de (T.R. Knösche).

0006-8993/$ – see front matter
doi:10.1016/j.brainres.2006.03.115

Balkwill and Thompson (1999) reported that Western listeners were sensitive to musically expressed emotions in an unfamiliar tonal system (Hindustani ragas). The authors suggested that the perception of emotion in unfamiliar music is facilitated by psychophysical but not cultural cues. Furthermore, there is evidence from developmental studies suggesting that human musicality is present from birth, and that its foundations are not culturally constructed. Smith and Williams (1999) conducted an experiment about children's artistic responses to musical intervals. They asked children from different cultural backgrounds to draw pictures when they were hearing music intervals. The results suggested that the perception of musical meaning is a universal rather than culturally based phenomenon.

As two main representations of auditory information, music and language hold similar acoustic features. The relationship between music and language has recently been investigated by a number of neurophysiological studies at different processing levels. Koelsch et al. (2004) found that music, similar to language, could prime the semantic processing in the human brain. At the syntactic level, incongruities in music elicit brain responses similar to the ones observed in language experiments (Besson, 1998; Maess et al., 2001; Patel et al., 1998). In addition, Maess and co-workers (Maess et al., 2001) used magnetoencephalography (MEG) to demonstrate that Broca's area, held responsible for the processing of syntax during language processing, was also involved in the early processing of music–syntactic incongruities.

Music and language are both composed of sequential events unfolding in time. This property renders segmental information, including phrase structure, very important in the on-line perception in both domains. The phrase perception process is normally induced by certain phrase boundary cues, like the insertion of a pause and the lengthening of the prefinal syllable or musical note. In language, phrasing has been investigated in behavioral (e.g., Warren et al., 1995) and neurophysiological studies (e.g., Steinhauer et al., 1999). Results from an event-related brain potential (ERP) study (Steinhauer et al., 1999) revealed a positive shift (Closure Positive Shift, CPS) in a close relationship with intonational phrase boundaries, which was interpreted to reflect the closing process of the prior intonational phrase. Later studies (Pannekamp et al., 2005; Steinhauer and Friederici, 2001) added more evidence to the notion that indeed prosodic rather than syntactic cues give rise to the CPS. Most interestingly, the investigation of phrase perception in music (Knösche et al., 2005) yielded EEG and MEG correlates for the perception of musical phrase boundaries, which showed a similar timing and topography as the CPS found in language studies. In a follow-up study (Neuhaus et al., 2006), the length of the pause and the length of the phrase-final tone before the pause were identified as the most important markers for the perception of the phrase boundaries. This result is in line with a recent behavioral study by Frankland and Cohen (2004), who found the note-to-note distances to be the most important phrase boundary cues.

Developed through exposure to particular culture-based environments, the processing of music and language is both culturally tuned. From cross-linguistic studies, there is accumulating evidence demonstrating that the same brain regions in the left hemisphere are activated for the processing of semantic and syntactic aspects (across different languages) when subjects process their native language (Gandour et al., 2004; Jacobsen et al., 2004; Klein et al., 2001; Schlosser et al., 1998; for a review, see Friederici and Rueschemeyer, in press). In contrast, cultural differences in the neural basis of music perception, though currently attracting much interest, have yet to be explored more extensively. There are, however, some accounts for cultural specificity of the brain processes associated with music perception (see, e.g., Neuhaus, 2003). As suggested by Arikan et al. (1999) for Turkish subjects, cultural effects could be shown when auditory oddball stimuli were superimposed on white noise, silent background, music played in familiar (ney) or unfamiliar instrument (violoncello). Larger P3 amplitudes were observed in the ney background condition as compared to violoncello one. This ERP result was interpreted as a facilitated attention allocation and memory updating processes when hearing music of a familiar style. In a musicogenic epilepsy case study, Genc et al. (2001) showed that only Turkish music was able to trigger seizures in this 48-year-old Turkish female but not other musical styles, e.g., Western music. Those two studies suggest distinct neural substrates engaged in the processing of culturally familiar music in comparison to unfamiliar music. In contrast, a functional magnetic resonance imaging (fMRI) study (Morrison et al., 2003) failed to prove any activation difference based on cultural familiarity of music either in musical experts or in naive subjects. The authors thus argued that listening to culturally different music might activate quite similar neural regions. Whether this result is due to the low temporal resolution (500 ms to 1000 ms) not allowing a differentiation of early auditory processes as investigated by a standard oddball paradigm and the indirect measures of brain activity by fMRI, or indeed due to a largely common neural substrate for the processing of culturally familiar and unfamiliar music, is not yet clarified.

The aim of the present study is to determine whether musical experts from different culture backgrounds differ from each other in their neural activation associated with cultural familiar/unfamiliar music processes. By using EEG, we aimed at revealing the interactions between cultural variables and music processes by highlighting the time course of processing. In particular, we investigated the perception of segmental information, i.e., phrasing. As mentioned above, the CPS has been proven to be a marker for the phrase perception in both music and language domains. Based on the results from previous comparative studies between music and language (see above and see review: Patel, 2003), it seems reasonable to predict a universal phrase perception mechanism. We used Chinese and Western musical excerpts for the current study. Compared to seven major or minor heptatonic scales for Western music, Chinese music mainly uses five pentatonic basic scales. A simple impression of how these scales sound like could be obtained by composing and playing a piece that can be performed by using only the black keys of the piano. By comparing brain activation differences of German and Chinese musicians in Western and Chinese music listening conditions, we intend to elucidate whether phrase perception is universal in music processing and to what extent it is modified by cultural differences.

2. Results

2.1. *Behavioral results*

The behavioral results are depicted in Fig. 1. The statistics are listed in Table 1. Clearly, both groups perform better with their own native style of music (interaction GROUP × STYLE; significant pair-wise differences between styles in each group; see Table 1b and Fig. 1A). Moreover, the Chinese show a higher performance than the Germans for Chinese traditional music, while there was no difference between the groups for Western music (main effect GROUP, significant pair-wise difference between groups for Chinese music, but not for Western music).

Additionally, it seems that for listening to culturally unfamiliar music (Chinese musicians listening to Western music and German musicians listening to Chinese music), both groups performed better for unphrased as compared to phrased versions, as hinted by a significant interaction COND × GROUP × STYLE. However, the differences were small and the respective pair-wise comparisons only marginally significant (see Table 1c and Fig. 1B).

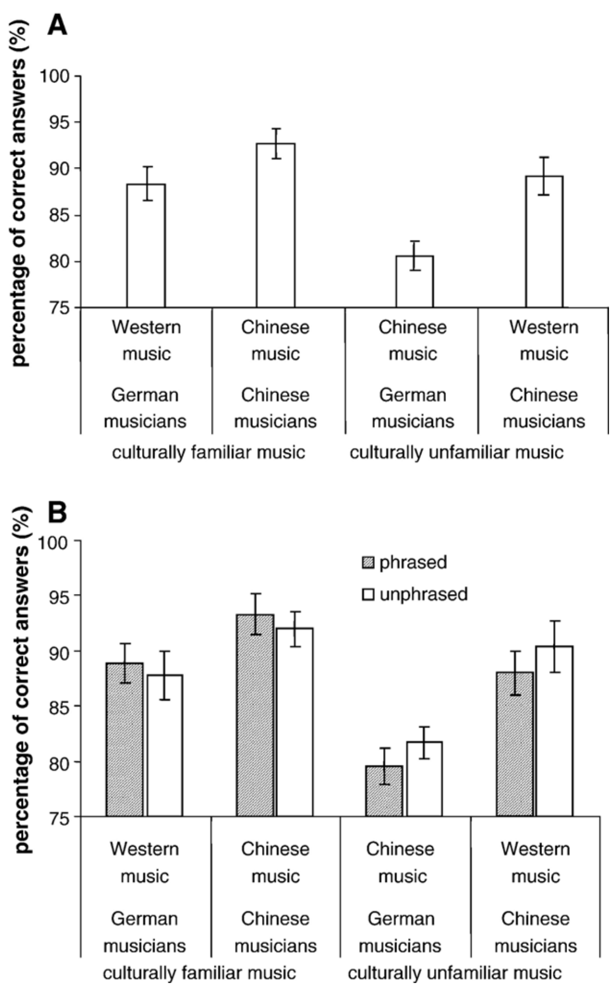

Fig. 1 – Behavioral results: percentage of correct answers. For different music styles over two groups of musicians, the diagrams illustrate (A) the interaction between STYLE (Chinese vs. Western music)*GROUP (Chinese vs. German subjects) and (B) the interaction between COND (phrased vs. unphrased melodies)*STYLE*GROUP. Error bars indicate standard deviations.

Table 1 – Behavioral results: statistics for the percentages of correct answers

*(a) Outcome of the three-way ANOVA: COND*STYLE*GROUP.* The degrees of freedom for all F values are (1, 20). COND stands for experimental condition (*phrased, unphrased*); STYLE indicates style of music (Chinese, Western); GROUP refers to subject group (Chinese, German). Only P values below 0.1 are shown

	Correct answers percent
COND	No significance
STYLE*GROUP	$F = 12.0$, $P < 0.01$
COND*GROUP	No significance
COND*STYLE	No significance
COND*STYLE*GROUP	$F = 7.8$, $P < 0.05$
GROUP	$F = 11.6$, $P < 0.01$

*(b) Pair-wise comparison for STYLE*GROUP interaction* (standard deviation in parenthesis)

	P value	Difference (1st item minus 2nd one)
German musicians: Chinese music vs. Western music	$P < 0.05$	–4.625 (1.62)
Chinese musicians: Chinese music vs. Western music	$P < 0.05$	2.10 (0.843)
Chinese music: Chinese musicians vs. German musicians	$P < 0.001$	7.225 (1.316)
Western music: Chinese musicians vs. German musicians	No significance	0.5 (1.649)

*(c) Pair-wise comparison between unphrased and phrased conditions for STYLE*GROUP*COND interaction.* The table refers to the difference between phrased and unphrased items as dependent variable

	P value	Difference (1st item minus 2nd one)
German musicians: Chinese music vs. Western music	$P < 0.1$	1.250 (0.653)
Chinese musicians: Chinese music vs. Western music	No significance	−0.667 (0.899)
Chinese music: Chinese musicians vs. German musicians	No significance	−0.8 (0.467)
Western music: Chinese musicians vs. German musicians	$P < 0.1$	1.40 (0.703)

2.2. *Electrophysiological results*

2.2.1. *P3 experiment*

The ERP traces for standard and deviant tones in different musician groups are shown in Fig. 2A. For both subject

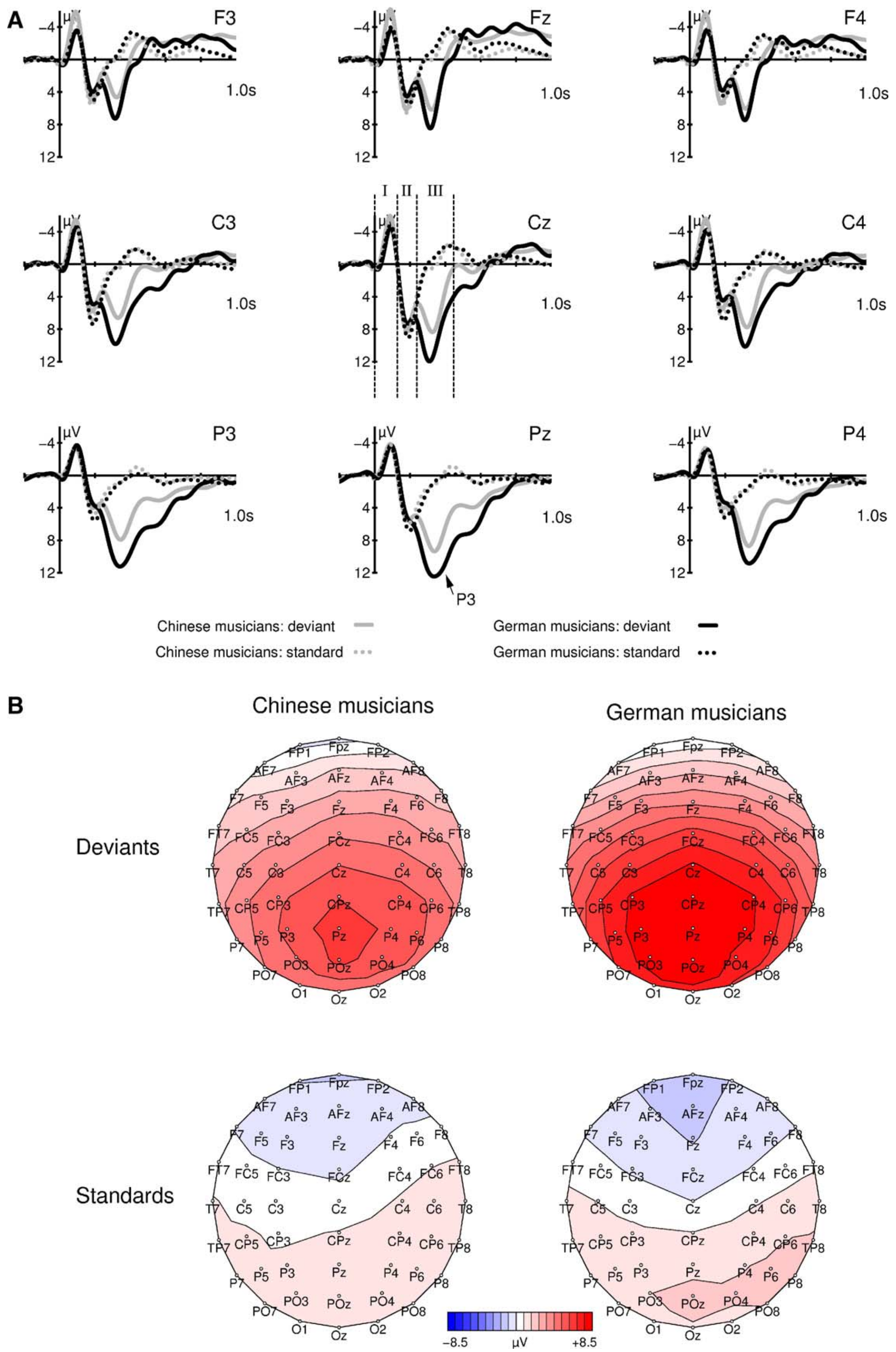

Fig. 2 – Grand average ERPs for the P3 experiment. (A) Time courses: the grey lines refer to Chinese musicians, black lines to German musicians; the solid lines stand for deviant tones, while dotted lines stand for standard tones. The analyzed time windows are labeled as follows: I (0–130 ms), II (130–240 ms), and III (240–450 ms). (B) Topographic maps: the ERPs were integrated over the time window III (300–450 ms). The maps are viewed from above, the nose is pointing upwards.

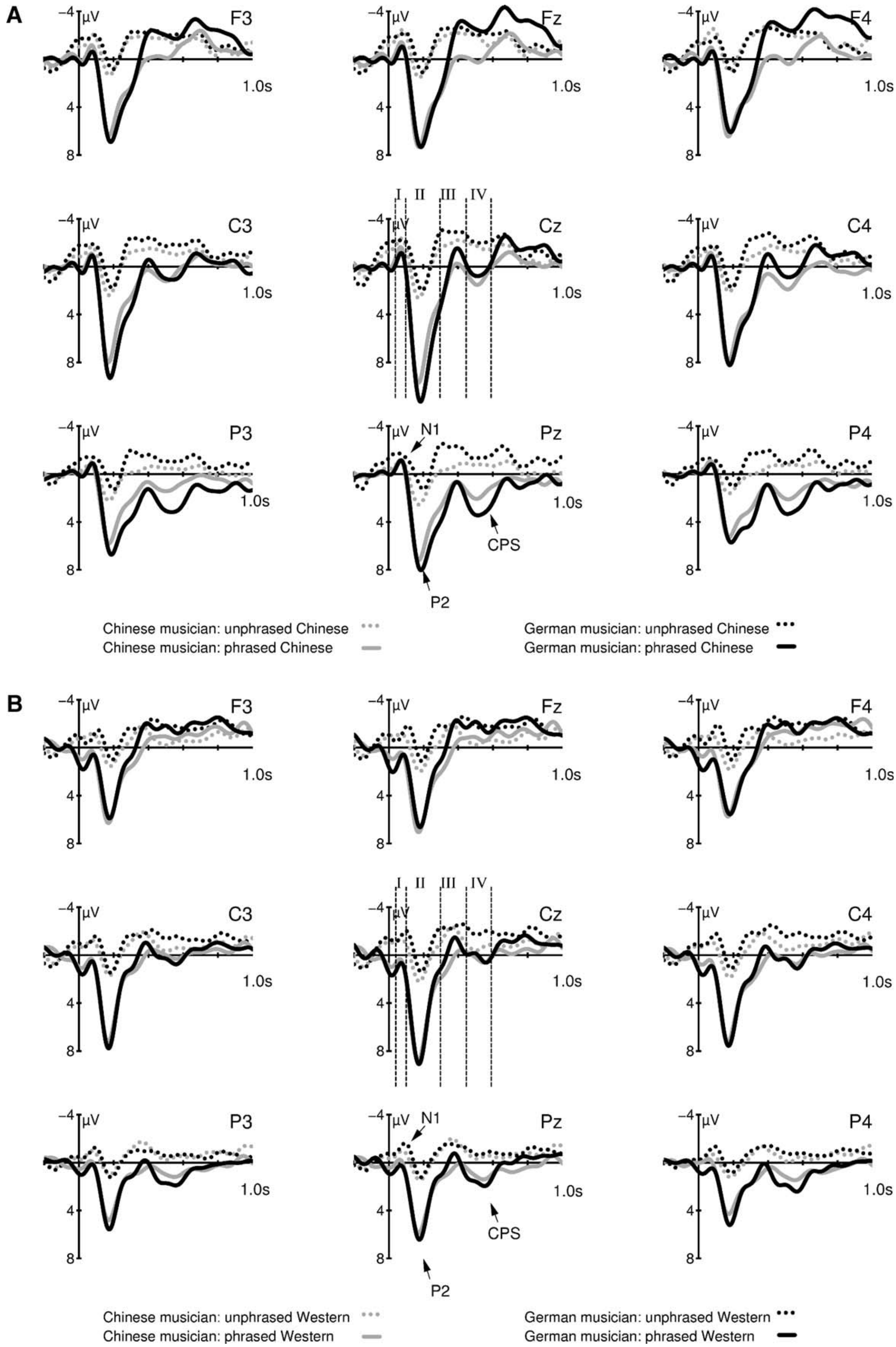

Fig. 3 – Time courses of the grand average ERPs for the main experiment: The grey lines indicate Chinese musicians, black lines for German musicians; the solid lines indicate *phrased* melodies listening, dotted lines for *unphrased* melodies listening. The chosen time windows are labeled as follows: I (40–100 ms), II (100–300 ms), III (300–450 ms), and IV (450–600 ms). Note that the baseline for all conditions was set prior to the onset of the phrase boundary in the *phrased* condition and its respective point in the *unphrased* condition (not the pretrigger interval shown in the diagrams). (A) Chinese music. (B) Western music.

Table 2 – Results of statistical analysis of the main experiment over all the subjects (mean amplitude in time windows)

Analysis window	I: 40–100 ms (N1)	II: 100–300 ms (P2)	III: 300–450 ms	IV: 450–600 ms (CPS)
COND	$F = 11.8$, $P < 0.003$	$F = 52.7$, $P < 0.001$	$F = 25.7$, $P < 0.001$	$F = 7.8$, $P < 0.05$
COND*STYLE	No significance	$F = 6.5$, $P < 0.05$	$F = 5.4$, $P < 0.05$	$F = 6.1$, $P < 0.05$
COND*GROUP	No significance	$F = 4.7$, $P < 0.05$	No significance	No significance
COND*STYLE*GROUP	No significance	$F = 4.2$, $p = 0.055$	$F^* = 4.9$, $P < 0.01$	No significance

COND stands for experimental condition (*phrased, unphrased*); STYLE refers to style of music (Chinese, Western); GROUP indicates subject group (Chinese, German). Only P values below 0.1 are shown. Only effects involving the factor COND are considered. The degrees of freedom for all F values are (1, 20). F^* indicates the significance involving the interactions with recording channel (51 levels), thus the respective degrees of freedom are (50, 1000).

groups and stimulus conditions, we observe a biphasic N1-P2 pattern, peaking at about 100 and 200 ms after stimulus onset, respectively. While the N1 shows no differences between subject groups or between stimulus conditions, the amplitude of the P2 seems to be slightly reduced for the deviant stimuli at some electrodes, in particular for Western subjects (main effect COND in time window II between 130 and 240 ms: $F_{1,20} = 4.26$, $P < 0.05$; interaction COND × CHAN: $F_{1,50} = 2.68$, $P < 0.05$). Moreover, for the deviant condition, the ERP pattern is completed by a large P3 wave (main effect COND in time window III between 240 and 450 ms: $F_{1,20} = 70.66$, $P < 0.001$; interaction COND × CHAN: $F_{1,50} = 27.73$, $P < 0.001$), peaking around 350 ms and showing maximum amplitude at midline central and parietal sites.

There are obvious differences between the subject groups in that the P3 appears to be considerably reduced for Chinese as compared to Western subjects. In the P3 window (time window III, 240–450 ms), the difference between standards and deviants (i.e., the magnitude of the P3 effect) statistically depended on the subject group (interaction COND × GROUP: $F_{1,20} = 5.33$, $P < 0.03$), confirming the larger P3 effect in Western as compared to Chinese musicians.

As shown by the topographies (see Fig. 2B), the P3 components in both groups of musicians were prominent in parietal areas and without any hemispheric preponderance.

Furthermore, a three-way ANOVA analysis of the factors COND, CHAN, and SESSION (first or second exposure to the stimulus material) conducted within each musician group revealed a main "SESSION" effect in German musicians, $F_{1,10} = 5.41$, $P < 0.05$, which shows that German musicians produced larger P3 component in session 1 compared to session 2. On the other hand, in session 1, the three-way ANOVA of the factors COND, CHAN, and GROUP evidenced a marginally significant GROUP effect ($F_{1,9} = 3.54$, $P = 0.093$) and a marginally significant interaction of COND × GROUP ($F_{1,9} = 3.05$, $P = 0.115$). No effects involving the GROUP factor were found in session 2. No SESSION effect was found for Chinese musicians, either.

2.2.2. *Main experiment*

The grand average ERP traces for the main experiment are illustrated in Fig. 3. Table 2 lists the results of the statistical analysis. Fig. 4 depicts the mean differences between phrased and unphrased items in the time windows between 100 and 600 ms.

The traces in Fig. 3 show that for the unphrased condition, there is already a negative deflection starting before the pause offset. This deflection somewhat merges with the N1 (especially for the Chinese music condition). The remaining ERP pattern exhibits a series of negative and

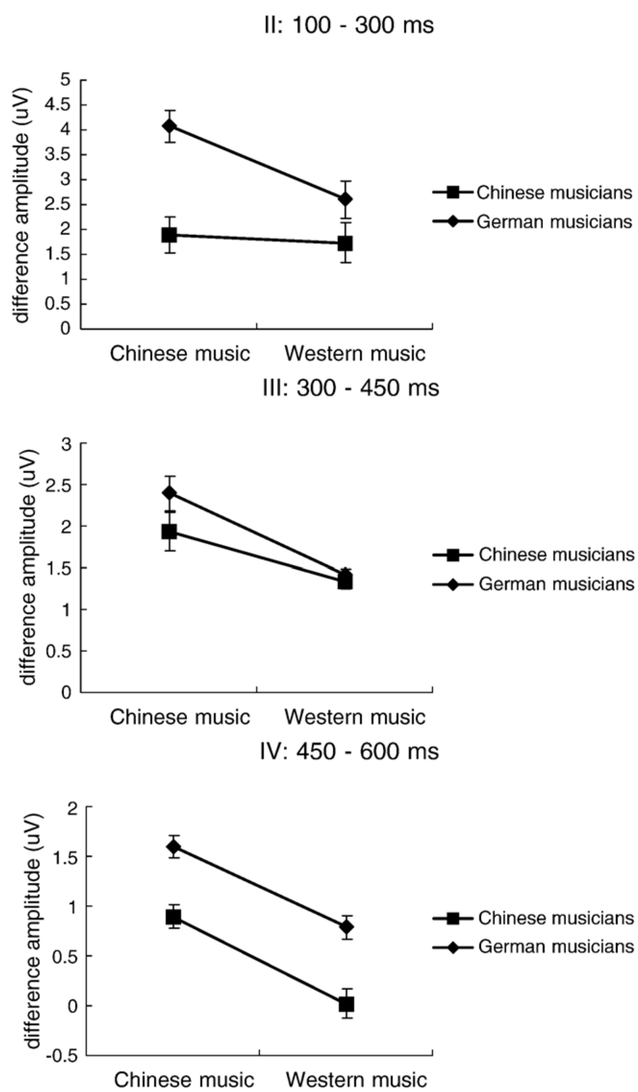

Fig. 4 – Mean amplitudes of the ERP components for the main experiment: graphs show the amplitude differences (averaged over all 51 channels) between *phrased* and *unphrased* conditions (*phrase/ unphrased*) over all the four combinations of music styles and musicians groups, for the time windows II (100–300 ms), III (300–450 ms), and IV (450–600 ms).

positive deflections peaking approximately 80 (negative), 200 (positive), 400 (negative), and 550 ms (positive) after the pause offset. These peaks have different amplitude distributions over electrode sites, phrasing conditions, subject groups, and musical styles. For analysis, time windows have been chosen, each of which encompasses just one component (see Experimental procedures and Fig. 3). Because the present study focuses on the effects of the processing of phrase structure, only effects involving the COND factor (comprising *phrased/unphrased* conditions) are reported. There were significant main effects for the COND factor for all the four time windows, reflecting the fact that over the entire time period of analysis the amplitudes were more positive going for the *phrased* as compared to *unphrased* condition.

The time window I, containing the N1, exhibits a main effect of COND, suggesting that the N1 is more pronounced in the unphrased condition. However, a closer look onto the waveforms (Fig. 3) suggests that this negative shift could stem from another negative wave peaking around the pause offset in the unphrased condition only. This component seems to overlap with the N1.

For both types of music, the P2 is larger for German musicians in the case of phrased melodies and larger for Chinese musicians in the case of the unphrased versions, as confirmed by a significant two-way interaction of COND × GROUP for the average amplitude in time window II. Both effects are considerably stronger for Chinese music (significant two-way interaction COND × STYLE and marginally significant (P = 0.055) three-way interaction of COND × STYLE × GROUP; see also Figs. 3 and 4). No interactions with the SESSION factor were found. Similar observations were made for the time window III (Fig. 4, Table 2).

From Fig. 4, it seems that the CPS is larger for German as compared to Chinese musicians, as well as for Chinese as compared to Western music, without any interaction between the factors. However, while the influence of the musical style could be confirmed by statistics (COND × STYLE interaction), no significant effect could be found for the differences between the groups or sessions.

3. Discussion

Before discussing the electrophysiological effects of phrasing and culture, we first turn to the behavioral findings and the results from our P3 tone perception test, in order to specifically illuminate the differences between the subjects groups.

3.1. *Behavioral data*

In order to examine the influence of enculturation onto the perception of music in musically trained subjects, it would be ideal if each of the chosen subject groups is expert in one and completely ignorant in the other musical culture. Practically, however, the worldwide dominance of classical Western music renders compromises inevitable. As our behavioral data clearly demonstrate, the Chinese subjects recognized classical Western music as reliably as the Germans did. Meanwhile, they are also quite good at recognizing traditional Chinese music (even slightly better than in Western music[1]). German musicians, on the other hand, categorized a much larger portion of the Chinese pieces erroneously as Western. Hence, if it comes to categorizing music, there are no differences between the groups for Western music, while for Chinese music the Chinese subjects have a clear advantage, although they did not receive any formal training in this type of music. This might be a result of the implicit exposure of the Chinese to traditional Chinese (besides classical and popular Western style) music during their life before they left China.

The effect of phrasing onto the categorization performance, on the other hand, is equal for both groups of musicians. For both Chinese and Germans, phrasing seems to facilitate categorization of their own type of music and hamper categorization of the respective culturally unfamiliar music type. A possible explanation of this phenomenon is that phrasing causes some sort of "feeling of familiarity", i.e., phrased pieces are more easily categorized as music of one's own culture.

3.2. *The P3 experiment*

The simple oddball paradigm using piano-like tones without any further musical context yielded a P3 like wave, which was larger for the German musicians. The topography of the component as well as the paradigm of the experiment clearly indicates that it is a target P3 or P3b, reflecting attention allocation processes (Friedman et al., 2001; Levy et al., 2003). It has been shown that the target P3 amplitude could be modulated by the arousal states of subjects due to natural or environmental factors (for a detailed review, see Polich and Kok, 1995). More difficult tasks produce larger P3 amplitudes due to the increased arousal states of the subjects (Sommer et al., 1993) and so does the presentation of motivating instructions (Carrillo-de-la-Pena and Cadaveira, 2000). In the current study, the overall larger P3 component for German musicians in the deviant condition is mainly originating from session 1 (first exposure to stimuli). Given the fact that the P3 experiment was conducted after a training procedure for the main experiment, it is likely that German musicians were relatively highly aroused by the more difficult task (unfamiliar Chinese music needed to be categorized) in the training process. This interpretation is in line with the findings that this effect is more obvious for session 1 when the German subjects encounter the unfamiliar Chinese music for the first time[2]. Results from the main experiment, i.e., the somewhat poorer behavioral performance and the larger ERP activities for German musicians compared to Chinese musicians in the Chinese music condition, support this interpretation.

[1] This in fact means that more Western pieces were erroneously categorized as Chinese, than the other way round, i.e., there is a certain inclination to categorize music as Chinese, if in doubt.

[2] Such a P3 reduction effect as a function of repetition has been evidenced in previous studies (Debener et al., 2002; Ravden and Polich, 1998).

Note that any anatomical explanations for the observed differences (e.g., in skull thickness or head shape) are not plausible because then they should have been equally visible in both sessions and all components, including the N1.

Most importantly, the fact that no SESSION effect could be found in the main experiment indicated that similar group differences based on different arousal states did not play a role there.

3.3. *The universality of the CPS component as a marker of phrase perception*

In the main experiment, we found a significant parietally distributed positive deflection for the *phrased* compared to the *unphrased* listening condition in all four combinations of subject groups and musical styles. This component was observed between 450 ms and 600 ms post pause offset and closely resembles the music CPS found in previous studies (Knösche et al., 2005; Neuhaus et al., 2006).

The CPS has been first found for the processing of intonational phrase boundaries in language studies (Hruska and Alter, 2004; Steinhauer, 2003; Steinhauer and Friederici, 2001; Steinhauer et al., 1999), where it was elicited by prosodic cues and was interpreted to indicate the processing of intonational phrase boundaries. Compared to the CPS found in language studies, the CPS in music (Knösche et al., 2005; Neuhaus et al., 2006) exhibits a different latency but a similar amplitude and topography. The difference in latency for the CPS component between language and music studies was interpreted by the different phrase boundary cues in language compared to music, such as lengthened prefinal syllables and changed F0 contours (Knösche et al., 2005). The similar topography and mode of elicitation make it reasonable to draw the conclusion that both components might reflect related cognitive processes.

The current study not only replicated the CPS finding for Western music (Knösche et al., 2005) but demonstrated a CPS in a cross-cultural setting with a categorization task. In the work of Knösche et al. (2005), subjects were asked to detect rarely occurring wrong notes in the presented musical pieces. This task was rather focusing the attention of the subjects to the local structure of the music, while our present categorization task required a more holistic approach. The elicitation of the CPS was not fundamentally influenced by this change of paradigm (see Fig. 3B as compared to Knösche et al., 2005, please note the different scales). This might be a hint that the music CPS reflects processes that are at least partially concerned with phrase boundaries as structuring elements of the entire musical piece, rather than with the mere detection of the boundary markers, such as pauses (i.e., interruptions in a continuous stream of sound). There are also other reasons to believe that the CPS at least partially reflects higher cognitive processing of phrase structure. First, the language CPS has been shown to be elicited independently of the presence of the pause at the intonational phrase boundary (Steinhauer et al., 1999). Second, for the CPS in music phrasing perception, EEG and MEG based source localization result has shown that likely neural generators involve structures of the limbic system associated with attention and memory processes (Knösche et al., 2005). Third, the music CPS is modulated by a number of factors other than pause, e.g., musical experience and harmonic closure of the previous phrases (Neuhaus et al., 2006).

Furthermore, in our study, the presence of the CPS seems to be unaffected by the relation between the acculturation of the subjects and the cultural style of the music (Fig. 4 and Table 2). Tentative influences of subjects' acculturation alone onto the amplitude of the CPS, as suggested by Fig. 4, could not be confirmed statistically (not even marginally) and must remain speculative, while the influence of the musical style could be proven. Independently of the cultural background of the listeners, Chinese music produced larger CPS amplitudes than Western music. The reason might be found in the properties of the stimulus material. Table 3 documents clear differences between the two types of music: Chinese pieces have longer phrases, longer pauses, as well as longer notes immediately before and after the pause. Therefore, this result is in line with the observation by Neuhaus et al. (2006) that longer pauses as well as longer boundary tones produce larger CPS amplitudes.

An alternative explanation for the observed musical style effect lies in the fact that both groups of musicians underwent the same type of formal musical training based on classical Western music. It could be that the CPS, in contrast to the earlier ERP components, reflects a process predominantly influenced by formal musical training, rather than passive exposure to a musical environment. This would be in line with the finding of Neuhaus et al. (2006) that explicit musical training has a profound effect onto the amplitude of the CPS.

Taken together, these results suggest that the CPS reflects a mechanism, which (1) is related to the processing of phrases and phrase boundaries, independently of the particular task of the subjects, (2) seems to reflect not only the mere detection of phrase boundary markers, but also the role of phrase boundaries as structuring elements of the entire musical piece, (3) is acquired or at least influenced by formal musical training, and (4) is elicited by different music styles irrespective of the relationship between the cultural background of the subjects and the cultural style of the music.

3.4. *Early effects of phrasing*

The analysis of the time window I suggests that there might be a difference in N1 between phrased and unphrased items. Due to interference from ERP activity around the offset of the pause interval in the unphrased condition (possibly caused by some auditory processing related to the pause-filling notes), an exact quantification of this effect is not possible. It might, however, be related to the refractoriness mechanism of the neural system (Budd et al., 1998; Jones et al., 2000).

In the time range between 100 and 450 ms, the ERP did not only react to the phrase boundary, but it was also affected by the different musical styles and the cultural backgrounds of the subjects. Hence, the underlying neural processes are driven by both stimulus features (i.e., bottom–up) and cultural-specific knowledge (i.e., top–down). Both types of information interact during this processing phase, expressed by the statistical interactions between the factors COND (phrased or unphrased), STYLE (Chinese or Western music), and GROUP (Chinese or German musicians). As has been postulated by Narmour (1991), combined top–down and bottom–up systems are very important for predicting the congruity of coming musical events when comparing them with preceding contextual implications. Although both groups of musicians have undergone the same formal training in Western music (where they showed little difference both behaviorally and electrophysiologically), they differ in their implicit exposure to traditional Chinese music (giving the Chinese a clear advantage with such music, reflected in better performances). Such implicit cultural influence on music processing has been demonstrated before in behavioral studies (Drake and El Heni, 2003; Schellenberg and Trehub, 1999).

4. Conclusion

The current study shows that the music CPS is also present, when subjects are required to perform a rather global categorization task instead of the more local wrong note detection task used in (Knösche et al., 2005; Neuhaus et al., 2006). Moreover, the music CPS was observed in subjects of different cultural background listening both to music of their native and an alien culture. These findings add to the generality of the CPS as a marker for the processing of musical phrasing. The results further suggest that the interaction between the acculturation of the listeners (top–down information) and the cultural style properties of the music (bottom–up information) influences phrase perception in a relatively early time window. In a recent behavioral study, it was evidenced that musical structure organization is facilitated, when listeners perceive music from their own culture (Drake and El Heni, 2003). The current ERP study adds to the description of this acculturation effect on music phrasing by specifying its temporal characteristics.

Taken together, the results support the notion that, although musical phrase perception processes are basically universal (as the CPS effect is present in all conditions), they are influenced by the relationship between musical style and acculturation of the subjects in a relatively early time window.

5. Experimental procedures

5.1. *Subjects*

Twenty-eight healthy trained musicians (14 German musicians, 14 Chinese musicians; all female) were paid for participation. None of the subjects had symptoms of any neurological, psychiatric or internal disease nor had they taken any medication on the day of examination and for at least 3 days before. All subjects had normal hearing abilities. Due to EEG artefacts, two German and three Chinese subjects were excluded from final analysis. Additionally, there was one pifa player among the recruited Chinese musicians, whose data were deleted from the final analysis because this subject was the only one majored in Chinese instrument playing. Thus, all information (including subject information, behavioral data, and EEG data) mentioned hereafter refers to the remaining 12 German and 10 Chinese musicians. All subjects were right-handed (average handedness: Chinese musicians, 91.1 ± 5.97; German musicians, 87.4 ± 3.99) according to the *Edinburgh Handedness Inventory* (Oldfield, 1971). The mean ages of the two group of subjects were both around 24 years (German musicians, 24 ± 1.2 years; Chinese musicians, 24.9 ± 1.4 years, respectively; age range 21–37, no significant difference between two musician groups). The average starting age for instrument playing was about 7 years for German musicians, about 8 years for Chinese musicians. Most of the subjects reported that they were able to play at least two instruments (11 German musicians, 8 Chinese musicians). The main present instrument for German musicians group were very diverse, including flute (5), accordion (1), violin (3), cello (1), piano (2); while for Chinese musicians, the main present instrument mostly focused on piano (7), also some violin (2), and keyboard (1). All subjects were currently still in the process of musical training, eighteen (11 German and 5 Chinese) had already passed their intermediate or final exams. Including the training hours spent on classes, German musicians reported on average to exercise 3.2 h per day, Chinese musicians 3.4 h per day. In addition to the daily exercise, many musicians got chances to perform in public as well (on average, German musicians: 5 times/year, Chinese musicians: 4.2 times/year). Except for one Chinese musician, who trained for Jazz piano, all the other musicians from both groups reported to mainly play music from Baroque, Classical and Romantic eras. This was the case for both training and spare time playing. Two of the Chinese subjects reported that they also preferred to play Chinese music in their spare time. Both groups of musicians reported a prominent Western classical hearing preference (German: 11.6 ± 3.4 h/week, Chinese: 10.9 ± 2.6 h/week), as well as the pop/rock music listening habits (German: 8.1 ± 1.8 h/week, Chinese: 9.5 ± 3.1 h/week). All Chinese musicians had been brought up in China. In order to pursue a higher level of musicality in Western classical music, they moved into the Western musical environment at the age of 18 to 24 years (on average 21.2 ± 0.7 years) from China. Except for two of them, who used to hear Chinese music occasionally in the past, none of German musicians reported experience with Chinese music. Chinese musicians occasionally listened to Chinese music after they left China (on average 1.2 h/week). All subjects gave their informed consent prior to the first investigation.

5.2. *Stimulus material and paradigm*

In this study, synthesized melodies with piano-like timbre were used. In order to explore any preexisting difference in auditory ERP components (e.g., due to anatomical differences) between the two groups of musicians, an oddball experiment was conducted immediately after the short training session for the main experiment. Three pairs of standard/deviant tones were delivered successively within 3 blocks of the experiment (120 standard and 30 deviant tones per block), each block containing one pair of stimuli, which were also

Table 3 – Comparison of phrase boundary features between Chinese and Western melodies

	Chinese melodies	Western melodies
Pause length (ms)	628 (39)	504 (25)
Length of the preceding phrase (ms)	8355 (361)	5290 (209)
Length of the last note before pause (ms)	765 (58)	511 (33)
Length of the first note after pause (ms)	296 (21)	248 (26)

Values are averages, with 95% confidence intervals in parentheses.

synthesized piano-like tones (same timbre as in main experiment). Tones were 500 ms in duration. The inter-stimulus interval was randomized between 400 and 600 ms. The task of the subjects was to silently count the number of deviant tones within each block and to give three final counts in the very end of the oddball experiment.

In the main experiment, we presented short melodic excerpts (120 in total) belonging to Western music (60) or Chinese traditional music (60). They were clearly structured into two phrases (each with 4 bars), the boundaries between which were marked by pauses. In order to avoid excerpts from the musicians' daily practice repertoire, the melodies were extracted from rather unknown musical pieces, with lengths ranging from 3000 to 17,000 ms.

For each of these melodies (labeled *phrased*), a modified counterpart was created (the work was done by a professional musician, who also evaluated all created examples), where the pauses were filled by one or several notes (labeled *unphrased*). An additional set of *combined* melodies was

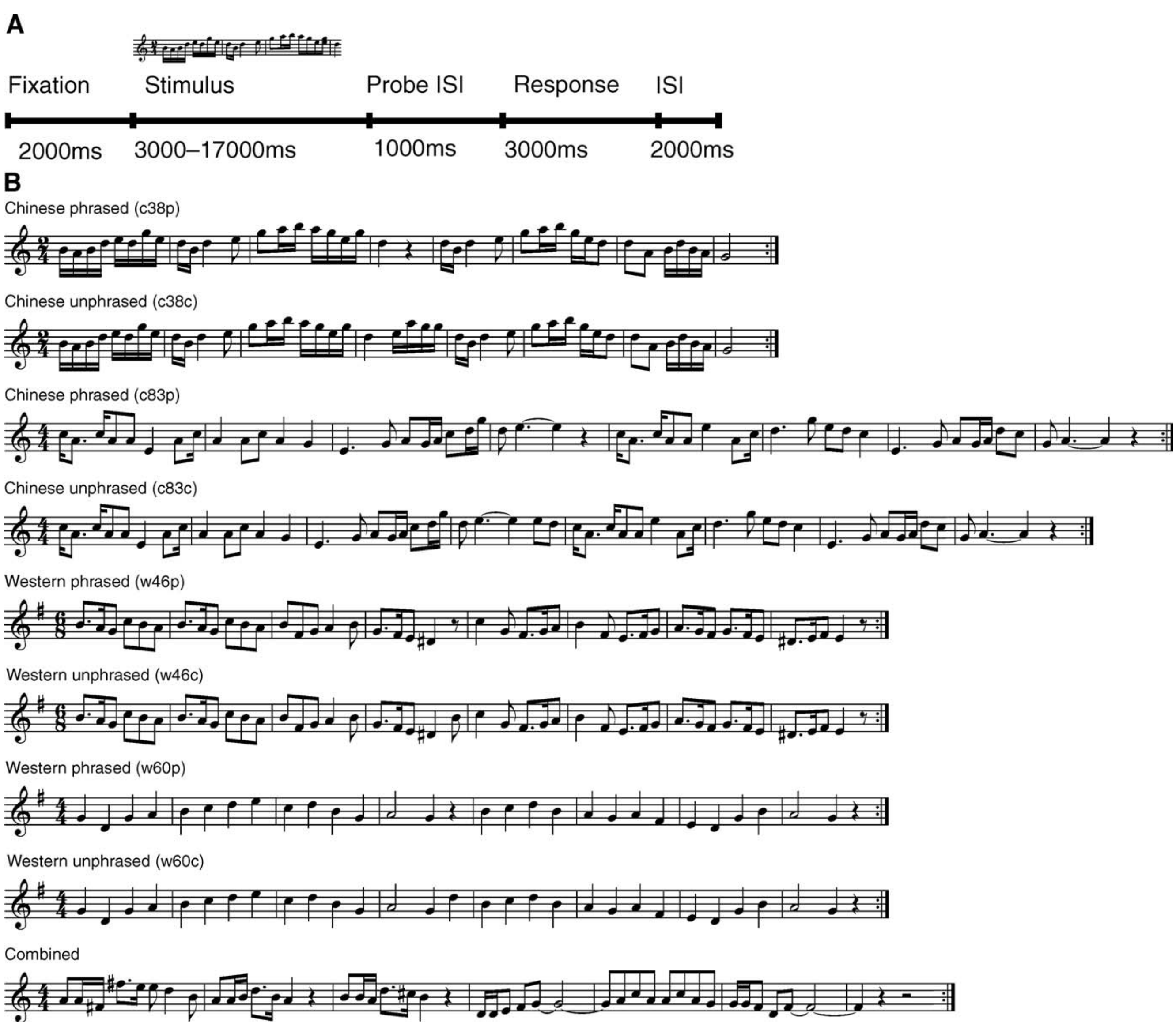

Fig. 5 – (A) The sequence of events within one single trial. ISI indicates the interstimulus interval. The stimulus is one of the biphrasal musical excerpts with different lengths ranging from 3000 to 17000 ms. (B) Typical biphrasal example stimuli. There are 60 stimuli of each of the four stimulus types: *phrased* Chinese, *unphrased* Chinese, *phrased* Western, *unphrased* Western. Compared to the original (*phrased*) versions, the *unphrased* ones were obtained by filling the pauses between two phrases with musically reasonable note(s).

generated by combining two fractions (the cut-off points were set randomly along melodies) from different examples (but belonging to the same style and having the same meter). These versions served as task items and were excluded from final analysis. A brief summary of the acoustic features of both groups of stimuli is listed in Table 3. Examples of stimuli are shown in Fig. 5B.

The experimental paradigm was similar to the previous work of Knösche et al. (2005) and is sketched in Fig. 5A. In order to prevent any sequence effects, all the melodies were presented in pseudo-random order. The resulting trials were mixed up in different manners (balanced for sequence) and split into 6 blocks of equal length. About 10% of the stimuli were task-relevant combined stimuli. Each experimental block was between 13 and 15 min long and contained an average of 20 Western and 20 Chinese as well as 20 phrased and 20 unphrased stimuli, together with 3 task items.

The current ERP experiment was carried out together with a similar MEG experiment using the same subjects and stimulus material (which is not reported due to technical problems). In order to prevent any sequence effects, half of the subjects from each group first participated in the current experiment, the other half in the MEG experiment. There was at least a 10-day interval between two sessions for each subject. The stimuli order was balanced between sessions. For the current experiment, this led to the fact that half of the data sets in each subject group resulted from the first exposure to the stimulus material, the other half from repeated exposure.

The task was to place each of the presented melodies into one of the following 3 categories: "combined", Chinese, or Western. Before the real EEG measurement, a detailed instruction and a training session were given to assure that the subjects understood the task well. The answers were collected by delayed key-press (a visual prompt sentence was given after the presentation of each melody, and also a feedback for the correctness of the answer was given immediately). They were not informed that the purpose of the experiment was the investigation of the perception of phrasing in a cross-cultural context.

5.3. *Recording procedure*

The subjects were seated in the electrically shielded room and fitted with an electrode cap with 51 Ag-AgCl electrodes connected. Additional reference electrodes were placed at the mastoid (left and right). Stimuli were presented via loudspeaker, which was placed in a distance of about 1.5 m. Before starting the first block, loudness as well as centering of the sound source was balanced individually. For getting accustomed to the stimuli, several examples were presented before the start of the actual experiment. The subjects were asked to avoid eye blinking, not to move, and to relax their facial muscles. They were also asked to look at a fixation point presented on the screen for the duration of each melody. Continuous EEG signals were recorded with a bandpass filter of 0 to 100 Hz and digitized on-line with a sampling rate of 250 Hz. Electro-oculograms (EOG) were recorded from electrodes above and below the right eye and at the outer canthi of the eyes. Impedances of electrodes were kept below 5 kΩ. In an offline procedure, trials with eye artefacts were marked and discarded. A high pass filter of 0.25 Hz was applied offline in order to reduce very slow drifts.

5.4. *Signal analysis*

Triggers were placed at the offsets of the phrase boundaries (the onset of the second phrase within each melody). The averaging window was set from −200 ms to 1000 ms relative to the trigger points. In order to make sure that the signal used for baseline correction does not contain any signal difference between the *phrased* and *unphrased* conditions, it was chosen from the time interval −800 ms to −300 ms relative to the *onset* of the respective phrase boundary. That is relative to the offset of the last note in the 1st phrase. Averages of EEG recordings were computed for every subject and condition. Trials with combined melodies (task items) were excluded from the averages. Sessions with more than 50% contaminated trials were excluded altogether, resulting in the total exclusion of 2 German and 3 Chinese subjects. Finally, grand averages over subjects were computed for every subject group (Germans and Chinese) and stimulus condition (Chinese music, Western music, phrased, unphrased). Visual inspection of these grand averages together with the so-called running *t* tests (i.e., *t* tests performed for each time step and channel without any multiple testing correction) was used to identify time ranges of interest for the various ERP components. For the main experiment, the following time ranges were used for analysis: (I): 40–100 ms (including the N1), (II): 100–300 ms (including the P2), (III): 300–450 ms, and (IV): 450–600 ms (including the CPS). For the P3 experiment, the analysis time windows were (I): 0–130 ms (including the N1), (II): 130–240 ms (including the P2), and (III): 240–450 ms (including the P3). For each of the defined analysis time windows, potential values were averaged over time samples and used as dependent variable of a four-way ANOVA with the within-subject factors "melodic condition" (COND, 2 levels: *phrased* and *unphrased*), "channels" (CHAN, 51 levels) and "music style" (STYLE, 2 levels: *Western* and *Chinese*), and the between-subject factor "cultural background" (GROUP, 2 levels: *German musicians* and *Chinese musicians*). Additionally to this four-way ANOVA, a five-way ANOVA, where the SESSION factor has been added, was also tested in order to detect any possible session effect. The *P* values of all effects were corrected using the Greenhouse–Geisser method for repeated-measure effects.

For the P3 experiment, a three-way ANOVA of the factors COND, CHAN, and GROUP was performed for each time window of interest. Moreover, for the time window III (P3 component), a three-way ANOVA of the factors COND, CHAN, and SESSION was conducted within each musician group; and a three-way ANOVA of the factors COND, CHAN, and GROUP was carried out within each session (session 1: first session was EEG measurement; session 2: second session was EEG measurement).

Finally, in order to obtain a behavioral index of the information processing, the percentages of correct answers were analyzed using a three-way ANOVA with the factors COND, GROUP, and STYLE.

Acknowledgments

The authors whish to thank Christiane Neuhaus for her valuable help with the selection and preparation of the stimulation material as well as Cornelia Schmidt and Heike Böthel for carrying out the measurements. The work described in this paper was partially supported by a grant of the *Deutsche Forschungsgemeinschaft* awarded to T.R.K. (KN 588/1-1).

REFERENCES

Arikan, M.K., Devrim, M., Oran, O., Inan, S., Elhil, M., Demiralp, T., 1999. Music effects on event-related potentials of humans on the basis of cultural environment. Neurosci. Lett. 268, 21–24.

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

Besson, M., 1998. Meaning, structure, and time in language and music. Cah. Psychol. Cogn. Curr. Psychol. Cogn. 17, 921–950.

Blacking, J., 1973. How Musical is Man? University of Washington Press, Seattle.

Budd, T.W., Barry, R.J., Gordon, E., Rennie, C., Michie, P.T., 1998. Decrement of the N1 auditory event-related potential with stimulus repetition: habituations vs. refractoriness. Int. J. Psychophysiol. 31, 51–68.

Carrillo-de-la-Pena, M.T., Cadaveira, F., 2000. The effect of motivational instructions on P300 amplitude. Neurophysiol. Clin.—Clin. Neurophysiol. 30, 232–239.

Castellano, M.A., Krumhansl, C.L., Bharucha, J.J., 1984. Tonal hierarchies in the music of north India. J. Exp. Psychol. Gen. 113, 394–412.

Debener, S., Kranczioch, C., Herrmann, C.S., Engel, A.K., 2002. Auditory novelty oddball allows reliable distinction of top–down and bottom–up processes of attention. Int. J. Psychophysiol. 46, 77–84.

Drake, C., Bertrand, D., 2001. The quest for universals in temporal processing in music. Ann. N.Y. Acad. Sci. 930, 17–27.

Drake, C., El Heni, J.B., 2003. Synchronizing with music: intercultural differences. Ann. N.Y. Acad. Sci. 999, 429–437.

Frankland, B.W., Cohen, A.J., 2004. Parsing of melody: quantification and testing of the local grouping rules of Lerdahl and Jackendoff's a generative theory of tonal music. Music Percept. 21, 499–543.

Friederici, A.D., Rueschemeyer, S.A., in press. Language acquisition: biological versus cultural implications for brain structure. In: Baltes Reuter-Lorenz, P.B., Rösler (Eds.), Lifespan Development and the Brain: The Perspective of Biocultural Co-Constructivism. Cambridge University Press, New York.

Friedman, D., Cycowicz, Y.M., Gaeta, H., 2001. The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. Neurosci. Biobehav. Rev. 25, 355–373.

Gandour, J., Tong, Y.X., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y.S., Li, X.J., Lowe, M., 2004. Hemispheric roles in the perception of speech prosody. NeuroImage 23, 344–357.

Genc, B.O., Genc, E., Tastekin, G., Ilhan, N., 2001. Musicogenic epilepsy with ictal single photon emission computed tomography (SPECT): could these cases contribute to our knowledge of music processing? Eur. J. Neurol. 8, 191–194.

Hruska, C., Alter, K., 2004. How prosody can influence sentence perception. In: Steube, A. (Ed.), Information Structure: Theoretical and Empirical Aspects, vol. 1. de Gruyter, Berlin, pp. 211–226.

Jacobsen, T., Schroger, E., Alter, K., 2004. Pre-attentive perception of vowel phonemes from variable speech stimuli. Psychophysiology 41, 654–659.

Jones, S.J., Vaz Pato, M., Sprague, L., 2000. Spectro-temporal analysis of complex tones: two cortical processes dependent on the retention of sounds in the long auditory store. Clin. Neurophysiol. 111, 1569–1576.

Klein, D., Zatorre, R.J., Milner, B., Zhao, V., 2001. A cross-linguistic PET study of tone perception in mandarin Chinese and English speakers. NeuroImage 13, 646–653.

Knösche, T.R., Neuhaus, C., Haueisen, J., Alter, K., Maess, B., Witte, O.W., Friederici, A.D., 2005. The perception of phrase structure in music. Hum. Brain Mapp. 24, 259–273.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T.C., Friederici, A.D., 2004. Music, language and meaning: brain signatures of semantic processing. Nat. Neurosci. 7, 302–307.

Krumhansl, C.L., 1995. Music psychology and music theory—Problems and prospects. Music Theory Spectr. 17, 53–80.

Krumhansl, C.L., 2000. Tonality induction: a statistical approach applied cross-culturally. Music Percept. 17, 461–479.

Levy, D.A., Granot, R., Bentin, S., 2003. Neural sensitivity to human voices: ERP evidence of task and attentional influences. Psychophysiology 40, 291–305.

Lynch, M.P., Eilers, R.E., 1992. A study of perceptual development for musical tuning. Percept. Psychophys. 52, 599–608.

Maess, B., Koelsch, S., Gunter, T.C., Friederici, A.D., 2001. Musical syntax is processed in Broca's area: an MEG study. Nat. Neurosci. 4, 540–545.

Morrison, S.J., Demorest, S.M., Aylward, E.H., Cramer, S.C., Maravilla, K.R., 2003. fMRI investigation of cross-cultural music comprehension. NeuroImage 20, 378–384.

Narmour, E., 1991. The top–down and bottom–up systems of musical implication—Building on Meyer theory of emotional syntax. Music Percept. 9, 1–26.

Neuhaus, C., 2003. Perceiving musical scale structures: a cross-cultural event-related brain potentials study. Ann. N.Y. Acad. Sci. 999, 184–188.

Neuhaus, C., Knösche, T.R., Friederici, A.D., 2006. Effects of musical expertise and boundary markers on phrase perception in music. J. Cogn. Neurosci. 18, 1–22.

Oldfield, R.C., 1971. The assessment and analysis of handedness. Neuropsychologia 9, 97–113.

Pannekamp, A., Toepel, U., Alter, K., Hahne, A., Friederici, A.D., 2005. Prosody driven sentence processing: an event-related brain potential study. J. Cogn. Neurosci. 17, 407–421.

Patel, A.D., 2003. Rhythm in language and music—Parallels and differences. Ann. N.Y. Acad. Sci. 999, 140–143.

Patel, A.D., Peretz, I., Tramo, M., Labreque, R., 1998. Processing prosodic and musical patterns: a neuropsychological investigation. Brain Lang. 61, 123–144.

Polich, J., Kok, A., 1995. Cognitive and biological determinants of P300: an integrative review. Biol. Psychol. 41, 103–146.

Ravden, D., Polich, J., 1998. Habituation of P300 from visual stimuli. Int. J. Psychophysiol. 30, 359–365.

Schellenberg, E.G., Trehub, S.E., 1999. Culture-general and culture-specific factors in the discrimination of melodies. J. Exp. Child Psychol. 74, 107–127.

Schlosser, M.J., Aoyagi, N., Fulbright, R.K., Gore, J.C., McCarthy, G., 1998. Functional MRI studies of auditory comprehension. Hum. Brain Mapp. 6, 1–13.

Smith, L.D., Williams, R.N., 1999. Children's artistic responses to musical intervals. Am. J. Psychol. 112, 383–410.

Sommer, W., Leuthold, H., Hermanutz, M., 1993. Covert effects of alcohol revealed by event-related potentials. Percept. Psychophys. 54, 127–135.

Steinhauer, K., 2003. Electrophysiological correlates of prosody and punctuation. Brain Lang. 86, 142–164.
Steinhauer, K., Friederici, A.D., 2001. Prosodic boundaries, comma rules, and brain responses: the closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. J. Psycholinguist. Res. 30, 267–295.
Steinhauer, K., Alter, K., Friederici, A.D., 1999. Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nat. Neurosci. 2, 191–196.
Warren, P., Grabe, E., Nolan, F., 1995. Prosody, phonology and parsing in closure ambiguities. Lang. Cogn. Processes 10, 457–486.