Accepted Manuscript

Title: Non-musicians’ Perception of Phrase Boundaries in Music: A Cross-cultural ERP Study

Authors: Yun Nan, Thomas R. Knösche, Angela D. Friederici

PII: S0301-0511(09)00124-0
Reference: BIOPSY 6101

To appear in:
Received date: 4-1-2008
Revised date: 9-6-2009
Accepted date: 10-6-2009


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Non-musicians’ Perception of Phrase Boundaries in Music: 

A Cross-cultural ERP Study

Yun Nan¹, Thomas R. Knösche², Angela D. Friederici²

¹State Key Laboratory of Cognitive neuroscience and Learning, Beijing Normal University 
(Beijing, China)

²MPI for Human Cognitive and Brain Sciences (Leipzig, Germany)

Corresponding author:

Thomas R. Knösche

Max Planck Institute for Human Cognitive and Brain Sciences

P.O. box 500355; 04303 Leipzig, Germany

Phone: +49 (0) 341-35521735

Fax: +49 (0) 341-35521740

Email: knoesche@cbs.mpg.de
Abstract

The present study investigates neural responses to musical phrase boundaries in subjects without formal musical training, with special emphasis on the issue of cultural familiarity (i.e., the relation between the enculturation of the subjects and the cultural style of the presented music). German and Chinese non-musicians listened to Western and Chinese melodies which contained manipulated phrase boundaries while event-related potentials (ERP) were measured. The behavioral data clearly showed that melodic phrase boundary perception is influenced by cultural familiarity. The ERP revealed a series of positive and negative peaks with latencies between 40 ms and 550 ms relative to the phrase boundary offset, all of which were influenced by the phrase melodic structure type. In contrast, cultural familiarity only influenced phrase boundary processing at longer latencies, reflected by a P3-like component peaking at 280 ms.

At about 450-600 ms post phrase boundary offset, we observed a slightly right-lateralized music closure positive shift (CPS), which has been reported as a marker for phrase boundary processing in musicians in earlier studies. This study demonstrates for the first time that the music CPS can be elicited in non-musicians, suggesting that the underlying phrase boundary processing does not require formal musical training.

Keywords: music perception; phrase structure; cultural differences; event-related potentials; non-musicians
1. Introduction

Over recent decades, the question of when and to what extent particular experiences affect the on-line processing of speech has been quite extensively discussed. Effects of different language experiences have been reported to be present quite early during development (Friederici et al., 2007; Näätänen et al., 1997) and to persist throughout life (Werker et al., 1992). In the domain of music, the influence of a specific music experience provided by different cultures has so far rarely been investigated (but see Lynch et al., 1990; Nan et al., 2006) although the role of musical training has been the subject of research (e.g. Besson and Faita, 1995; Schön et al., 2004; Trainor et al., 1999). For a recent review of the roles of musical enculturation and formal musical training during ontogeny, see Hannon and Trainor (2007).

Musical phrasing

In order to successfully parse music and speech, phrases which structure the incoming information are important (Chiappe and Schmuckler, 1997; Dowling, 1973; Stoffer, 1985; Tan et al., 1981; Wilson et al., 1999) not only for adults but also for young infants (Hirsh-pasek et al., 1987; Jusczyk and Krumhansl, 1993; Krumhansl and Jusczyk, 1990). The fast and accurate recognition of phrase boundaries is crucial for the reliable processing of the input underlying phrase structure. Phrase boundaries are marked by both structural (global) and acoustic (local) cues (Trainor and Adams, 2000). For language, the structural cues include grammatical rules and phonotactic constraints, while for music, they consist of functional relations between the different notes of the scale. Acoustic cues to phrase boundaries in both speech (Cooper and Sorensen, 1977; Scott, 1982) and music (see, e.g. Boltz, 1993; Clarke, 1985; Jusczyk and Krumhansl, 1993) include the insertion of a pause, lengthening of the final element(s) of the phrase, decrease in intensity at the end of the phrase, and changes in pitch. A number of studies on music perception indicate that, quite similar to language (Nakatani and Schaffer, 1978), in general, duration-based cues seem to be more important than those based...
on intensity (Drake and Palmer, 1993; Frankland and Cohen, 2004; Trainor and Adams, 2000),

The single most important cue is clearly the pause (Frankland and Cohen, 2004; Neuhaus et
al., 2006; Riemann, 1900), but the relevance of other cues has been demonstrated in
behavioral (Boltz, 1993; Clarke, 1985; Cuddy et al., 1981; Jusczyk and Krumhansl, 1993; Tan
et al., 1981) as well as electrophysiological (Neuhaus et al., 2006) studies.

Recently, electroencephalography (EEG) and magnetoencephalography (MEG) have been
used to investigate the time course of the processing of musical phrase boundaries in
musicians (Knösche et al., 2005; Nan et al., 2006; Neuhaus et al., 2006) and non-musicians
(Neuhaus et al., 2006). In musicians, a centro-parietal positive event-related potential (ERP)
component approximately 500 milliseconds after the onset of the first tone following a phrase
boundary was identified. Because it bears some resemblance in its mode of elicitation and
scalp topography to the Closure Positive Shift (CPS) indexing the processing of intonational
phrase boundaries in speech (Pannekamp et al., 2005; Steinhauer, 2003; Steinhauer et al.,
1999; Steinhauer and Friederici, 2001), it was labeled music CPS (Knösche et al., 2005).

Based on their similarity, it was suggested that language and music CPS might reflect related
cognitive processes (i.e., the closure of a phrase). If so, this would indicate that phrase
structure processing is a cognitive function shared by both music and language, similar to
other aspects of structural processing of language and music for which similar neural
substrates are reported. For example, ERP components reflecting syntactic processing have
been found to be very similar for speech (Friederici et al., 1993) and music (Koelsch et al.,
2000; Patel, 1998). Moreover, MEG source analysis (Maess et al., 2001) and functional
magnetic resonance imaging (fMRI) (Koelsch, 2005; Koelsch et al., 2002a) showed that many
constituents of the cortical ‘language network’, including Broca’s area, are also involved in
the processing of musical syntax.

The music CPS as a late ERP component has been found to be sensitive to a number of phrase
boundary cues, in particular to pause length, length of the last tone preceding the pause, and
harmonic function of this last tone (Neuhaus et al., 2006), and appears to reflect rather late, high-level aspects of phrase boundary processing.

However, effects of music phrase processing on the early auditory components were also observed (Knösche et al., 2005; Neuhaus et al., 2006). Since, in these studies, the critical condition (phrased condition) contained a pause whereas the control condition (unphrased condition) did not, these effects might reflect a confound with habituation and refractoriness of the neuronal populations in the auditory cortices.

The influence of formal musical training

Although there is some work indicating that fundamental processes of music cognition may be similar in all normal humans regardless of training (Bigand et al., 1999; Bigand and Pineau, 1997; Koelsch et al., 2000; Regnault et al., 2001), other studies demonstrated that musicians and non-musicians differ in many aspects when processing of music and also other auditory input. These aspects include detection performance for temporal and dynamic perturbations (Repp, 1995), perception of loudness, timbre and pitch (Hoover and Cullari, 1992), as well as brain activity during passive listening to musical stimuli (Morrison et al., 2003; Shahin et al., 2003) and in response to music expectancy violations (Besson et al., 1994; Fujioka et al., 2004; Hantz et al., 1992; Janata, 1995; Jongsma et al., 2005; Koelsch et al., 2005; Regnault et al., 2001; Trainor et al., 1999; van Zuijen et al., 2004). Musical training seems to play a role in music phrase structure perception as well, as indicated by various studies (Chiappe and Schmuckler, 1997; Frankland and Cohen, 2004; Tan et al., 1981). In this line, it has been demonstrated that the occurrence of the music CPS seems to depend on an individual’s formal musical training. In an ERP study, Neuhaus et al. (2006) found that, in contrast to formally trained musicians, non-musicians did not exhibit a CPS in response to phrase boundaries, but an earlier and frontally distributed negativity was found. MEG recordings of non-musicians, on the other hand, did show a CPSm (magnetic equivalent to the CPS) qualitatively similar to the one for musicians, although much lower in amplitude. This apparent difference between
EEG and MEG may be due to the fact that the two methods reflect different aspects of the neural substrate of psychological processes (see discussion in Shahin et al., 2003). MEG, for example, is most sensitive to sulcal activity and almost silent for sources on the crowns of gyri, while EEG yields highest amplitudes for sources which are located near the brain surface and point towards the skull (thus on the gyral crowns). The combined results appear to indicate that phrase boundary processing in musician and non-musician takes place in the time range of the CPS, but that the underlying processes differ. This interpretation is also supported by the finding that the time courses of the music CPS and CPSm in musicians differ considerably (single peak vs. double peak), suggesting that they are not entirely based on the same generators (Neuhaus et al., 2006). These findings suggest that the CPS (at least partially) reflects processes that are modulated by formal musical training and are thus more modifiable than previously thought.

A final conclusion, however, must await further research, as the reported result might have been due to the specific experimental setting. In the study by Neuhaus and colleagues (2006), participants were asked to scan the input for false (off-key) notes, a task that is particularly demanding for non-musicians and requires a rather local, element-directed focus of attention. This might have prevented non-musicians from processing the input in the global, holistic way necessary to appreciate phrase structure and to anticipate phrase boundaries, while for musicians the detection of the off-key notes required little effort and normal holistic processing of the melodies remained unimpaired. Hence, it remains to be investigated whether under different task conditions, non-musicians would exhibit similar electrophysiological correlates for phrase processing as musicians, indicating similar underlying brain processes.

The influence of culture

Music is culturally tuned. Musical styles from different cultures differ in many respects, such as pitch structure, rhythmic and metric rules, and timbre of main instruments, to name but a few. The implicit exposure to music of a particular cultural tint shapes the musical perception
of individuals in that environment. This enculturation starts in infancy, when basic universal features are acquired first, forming the basis for the later acquisition of system specific aspects (Hannon and Trainor, 2007). Behavioral research by Lynch et al. (1990) found that, in contrast to adults, 6-year-old American children were equally sensitive to a mistuned tone in a non-Western scale context as in a Western one. Similarly, Trainor and Trehub (1992) reported that, although adults from a diatonic tonal environment find non-diatonic changes easier to detect than diatonic ones, infants find both types of change equally easy to detect. Jusczyk and Krumhansl (1993) suggested that the basic grouping principles appear to be quite reliable and robust in infancy. These basic grouping principles might facilitate the acquisition of the particular intervalic structures of a musical culture. Cross-cultural behavioral studies (Castellano et al., 1984; Kessler et al., 1984) showed that listeners unfamiliar with a particular musical style tend to use these cues for identifying the structurally important tones.

These results are complemented by a recent cross-cultural neurophysiological study on musical phrase structure perception with musicians (Nan et al., 2006). By manipulating the presence of pauses between two phrases in bi-phrasal melodies from both traditional Chinese and classical Western music, phrase boundary related EEG signals were collected from Chinese and German female musicians, who were asked to perform a categorization task of music style. The processing of musical phrase boundaries was found to be reflected by changes in a number of ERP waves between 100 and 600 ms after phrase boundary offset. The early components (100 - 450 ms) appeared to be influenced by the relation between the cultural affiliations of stimuli and subjects (interaction between musical style and subject group), in that differences between the groups only existed for Chinese music (which was familiar to the Chinese only) and not for Western music (which was familiar to both subject groups). It remains unsolved what the crucial parameter for the differential pattern in the early ERP components is. One possible reason could be that the Chinese had had implicit exposure

---

1 The purpose of the task was to keep the subjects’ focus of attention on the global structure of the musical excerpts.
to both traditional Chinese and Western music throughout their lives, while the experience of the German subjects had been largely limited to Western tonal music. This interpretation would lead to the prediction that similar differential effects between German and Chinese subjects might also be observable in non-musicians. As an alternative explanation, it is possible that the formal training in Western classical music received by the Chinese subjects contributed to the absence of group differences between German and Chinese brain activity for Western music. In this case, the prediction would be that a difference between German and Chinese non-musicians should be observable for Chinese but not for Western music.

The late component, that is, the music CPS (around 550 ms), exhibited an effect of musical style, in that Chinese music generally elicited a larger CPS but no interaction between musical style and subject group. This finding points towards an explanation which assumes that the CPS reflects a culturally universal process that is only influenced by the cue structure of the phrase boundary, which is different between the two different musical styles (e.g. pause lengths, lengths of boundary tones, interval structure, etc., see also Tab. 3 in Nan et al., 2006).

In order to test the above predictions and hypotheses, we applied the paradigm used by Nan et al. (2006) to non-musicians.

2. Methods

Subjects

Two groups (20 German and 20 Chinese) of healthy, female subjects with normal hearing were recruited for the current study. They were paid for their participation and gave informed consent prior to the investigation. The study was conducted at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany. It was approved by the local ethics committee and was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

Four German and two Chinese subjects were excluded from final analysis due to excessive
EEG artifacts. Consequently, all data (including subject information, behavioral data and EEG
data) mentioned hereafter refer to the remaining 16 German and 18 Chinese participants. All
subjects were right-handed (average handedness: Chinese, 90.95+/-3.6; Germans, 98.9+/-0.8)
according to the Edinburgh Handedness Inventory (Oldfield, 1971). The mean age of both
groups of subjects was around 25 years (Germans, 25.0+/-0.3 years; Chinese, 25.3+/-0.6 years,
respectively; age range 22–31, no significant difference between groups). On average,
Chinese subjects had lived in Germany for 19 months (range: 1 month to 4 years). All of them
resided in China before coming to Germany. They reported equal familiarity with both
Chinese and Western music. On the other hand, no German non-musician reported any
familiarity with Chinese music. No subject had had any formal musical training lasting longer
than 3 months. None of them played a musical instrument or sang in a choir at the time of
investigation.

**Stimulus Material and Paradigm**

We selected 120 short excerpts (60 Western music, 60 Chinese music) of preexisting music
(i.e., not especially composed for the experiment), which were not generally known. All
excerpts were clearly structured into two phrases (each with 4 bars), with pauses marking the
phrase boundaries (labeled *phrased* throughout this paper). The lengths of the pieces ranged
from 3000 to 17000 ms. In order to investigate the effect of the phrase boundaries, another set
of melodies was created by a professional musician by filling the pauses with musically
plausible note(s) (labeled *unphrased* throughout this paper). This technique, already used in
earlier studies (Knösche et al., 2005; Nan et al., 2006; Neuhaus et al., 2006), ensures that the
meter remains intact. Clearly, the manipulated pieces cannot be considered strictly
“unphrased”, since phrase boundary markers other than the pause (e.g., prefinal lengthening,
harmonic closure, melodic contour, etc.) are still in place. Therefore, the condition is actually
rather “less phrased” than “unphrased”. However, for the sake of brevity and consistency
with earlier publications, we will continue to use the latter term. See also Neuhaus et al. (2006) for behavioral data demonstrating that the impression of “phrasedness” is considerably reduced by the filling of the pause by notes.

An additional set of melodies was created by combining two fragments (one head and one tail fragment, with the cutting point chosen randomly) from different examples belonging to the same cultural style (labeled combined throughout the paper). These versions served as task items and were not used in the final analysis. All melodies were synthesized with piano-like timbre. A brief summary of the acoustic features of both groups of stimuli is listed in Table 1. Examples of stimuli are shown in score format in Figure 1a.

The experimental paradigm was adapted from the previous work of Knösche et al. (2005) and is outlined in Figure 1b. All melodies were presented in pseudo-random order in order to avoid sequence effects. There were 6 blocks of equal length. Each experimental block was between 13 and 15 minutes’ duration and contained an average of 20 Western and 20 Chinese as well as 20 phrased and 20 unphrased stimuli, together with 3 combined task items.

The task was to categorize the presented melodies into three classes, namely combined melodies, Chinese melodies, and Western melodies. Before the EEG measurement, a detailed instruction and a training session were held to ensure that the subjects understood the task well. The subjects’ responses were collected by delayed key-press (a visual prompt sentence was given after the presentation of each melody, and feedback regarding the correctness of the answer was given immediately). Subjects were not informed that the purpose of the experiment was to investigate the phrasing perception in a cross-cultural context.

**Recording Procedure and Signal Analysis**

The experiment was conducted in an acoustically and electrically shielded room. The subjects were seated comfortably in an armchair. EEG was continuously recorded from 64 Ag-AgCl electrodes. The reference electrodes were placed at the mastoid (left and right). Stimuli were
presented via two loudspeakers, which were placed in front of the subject at a distance of 1.5 m. Before starting the first block, loudness was individually balanced. Subjects were asked to avoid blinking, stay still, and relax their facial muscles. They were required to look at a fixation point presented on the screen for the duration of each melody. The EEG signals were recorded with a bandpass filter of 0-100 Hz and digitized on-line with a sampling rate of 250 Hz. The electro-oculogram (EOG) was 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 artifacts were marked and discarded. Additionally, a high pass filter of 0.25 Hz was applied to reduce very slow drifts.

Triggers were placed at the offsets of the phrase boundaries (i.e., at the onset of the second phrase within each melody). As baseline, we used the last 200 ms prior to the phrase boundary offset. Since this baseline interval might contain different signals for phrased and unphrased items, we performed another analysis with the baseline set to the time interval from –800 ms to –300 ms relative to the onset of the respective phrase boundary (i.e., to the offset of the last note in the first phrase). In this way, we double-checked the robustness of the reported findings. The averaging window was set from -200 ms to 1000 ms relative to the trigger points. Separate averages were computed for every subject and condition. Trials for combined melodies (task items) were excluded from the averages. Sessions with more than 50 % artifact contaminated trials were completely excluded, resulting in the total exclusion of 4 German and 2 Chinese subjects. Finally, grand averages over subjects were computed for every combination of subject group (German and Chinese), musical style (Chinese music vs. Western music) and stimulus condition (phrased vs. 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 separately without any multiple test correction), was used to identify time ranges of interest for the various ERP components. The following time ranges were identified (I): 25-55 ms (around the peak of the P1), (II): 75-105 ms (around the peak of the N1), (III):
135-210 ms (around the peak of the P2), (IV): 260-310 ms (around the additional peak on the falling flank of the P2 – see results), and (V): 450-600ms (including the CPS).

The statistical analysis of the results was done by variance analysis (ANOVA). In order to obtain a behavioral index of the information processing, the percentages of correct answers were analyzed using a three-way ANOVA (2×2×2 design) with the within-subject factors “melodic condition” (COND, 2 levels: *phrased* and *unphrased*) and “musical style” (STYLE, 2 levels: *Western* and *Chinese*), and the between-subject factor “cultural background” (GROUP, 2 levels: *German non-musicians* and *Chinese non-musicians*).

The ERP data were analyzed separately for each of the defined time windows. First, the potentials were averaged over the respective time window. Then, a four-way ANOVA (2×2×2×4 design) was computed for the resulting average values, with the factors COND, STYLE, GROUP, and AREA. Note that GROUP is a between-subject factor, while the other three factors are within-subject factors. The factor AREA was created by summarizing the channels in four regions of interest as follows: left-anterior (FP1, AF3, AF7, F3, F5, F7, FC3, FC5, FT7); right-anterior (FP2, AF4, AF8, F4, F6, F8, FC4, FC6, FT8); left-posterior (CP3, CP5, TP7, P3, P5, P7, PO3, PO7, O1); right-posterior (CP4, CP6, TP8, P4, P6, P8, PO4, PO8, O2). The *p* values of all main and interaction effects were corrected using the Greenhouse-Geisser method for repeated-measure effects. All interaction effects involving the factor COND were further detailed by separate ANOVA analyses at all levels of one of the other involved factors. This procedure was repeated recursively until only main effects of the factor COND remained.

3. Results

**Behavioral Results**

The percentages of correct answers for both groups and musical styles are given in Figure 2a. Western music was equally correctly categorized by both groups of subjects, while a large
difference between the groups was revealed when classifying Chinese music: Chinese
subjects’ performance was superior in the classification of Chinese music as compared to
Western music, while German subjects’ performance was considerably worse than for Western
music. These results were further confirmed by the significant main group effect, F(1, 32) =
21.24, p<0.001, and the significant interaction of music style and subject group (STYLE ×
GROUP), F(1, 32) = 55.6, p<0.001. In order to illustrate the STYLE × GROUP interaction, the
mean values for each level of GROUP and STYLE are depicted in Figure 2a.

As demonstrated in Figure 2b, it was only in the native music listening conditions (i.e.,
Germans listening to Western music, Chinese listening to Chinese music) that both groups of
subjects showed better performance for the original music (phrased melodies) than for the
modified unphrased ones. For the non-native conditions (i.e., Germans listening to Chinese
music, Chinese listening to Western music), the opposite was observed (i.e., the performance
was better for the unphrased melodies). This observation is confirmed by a significant three-
way interaction of COND × STYLE × GROUP, F(1, 32) = 10.00, p<0.05. It appears that the
performance difference between phrased and unphrased melodies mainly occurred with
Chinese music for both groups of subjects. The results relating to Western music revealed a
similar tendency but were not significant. These results closely resemble the findings for
musicians from the study by Nan and colleagues (2006).

**Electrophysiological Results**

Figure 3 illustrates the grand average waveforms. Topographic maps are shown in Figure 4.

For all subject groups and conditions, the waveforms show a triphasic pattern of waves, which
we term P1 (peaking at about 40 ms), N1 (peaking at about 90 ms), and P2 (peaking at about
180 ms), followed by another positive peak on the falling flank of the P2 (at about 290 ms) for
the phrased conditions. Subsequently, a negativity at about 400 ms becomes evident for all
conditions, which, for the phrased items, is followed by the centro-parietal CPS peaking at
about 550 ms. These observations, as well as the corresponding statistical assessments, led to the definition of the analysis time windows (see methods section). Table 2 lists the ANOVA results for those time windows.

The reported ANOVA results are based on data, which were baseline-corrected by subtracting the mean ERP of the last 200 ms just before the pause offset. However, we also computed an ANOVA with data that were corrected by subtracting the average of the time window -800 to -300 ms prior to the onset of the pause (see methods section). All reported findings and conclusions are supported by both types of analysis. In the following, we describe the results for the particular analysis time windows.

**Time window I (25-55 ms; around the peak of the P1)**

The P1, peaking at approximately 40 ms post phrase boundary offset, was more positive for the phrased items (Figure 4; Table 2). Post-hoc analysis of the COND × STYLE and COND × AREA effects revealed that this enlarged P1 was mainly restricted to the Western musical style and visible in anterior electrodes only (see also Figures 3 & 4). There was no influence of the group factor for this component. Closer inspection of Figure 3 suggests that it is possible that the effects observed in this early time window originate (completely or in part) from differences in the baseline time window. Hence, the observed interactions with the experimental variables might be attributable to some positive components (i.e., P1 and P2) in response to the pause-filling notes in the unphrased condition.

**Time window II (75-105 ms; around the peak of the N1)**

Immediately following the P1, the negative N1 component was observed, peaking at around 90 ms, with maximal amplitude in the central area (Figures 3 & 4). As compared to the *unphrased* melodies, the *phrased* ones were associated with larger N100 amplitudes, most prominently for the Chinese musical style (see Table 2 and Figures 3 & 4), as confirmed by post-hoc analysis of the COND × STYLE effect: while for Western melodies, the mean
difference between conditions was 0.5 \mu V (marginally significant, p<0.1) and for Chinese melodies it was 1.3 \mu V (significant, p<0.001). Again, there was no detectable influence of the cultural background of the subjects.

**Time window III (135-210 ms; around the peak of the P2)**

The centrally distributed P2 component peaked around 180 ms (Figures 3 & 4). Both groups of subjects exhibited a much larger P2 in the *phrased* as compared to the *unphrased* condition (main effect of COND, Tab. 2). Post-hoc analysis of the COND \times AREA interaction showed that the effect was present in all 4 channel groups, but stronger in the anterior ones. Notwithstanding the apparent increase in the difference between phrased and unphrased items in the culturally unfamiliar condition, that is, Germans listening to Chinese music (see Fig. 3C, 4), the P2 effect of phrasing did not interact statistically significantly with STYLE or GROUP. Close inspection of the data renders it likely that this apparent deviance is a consequence of some overlap between P280 and P2.

**Time window IV (260-310 ms)**

At the falling flank of the P2, the traces belonging to the phrased items exhibit another positive peak at about 280 ms, termed P280 (Figures 3 & 4). In contrast to the peak of the P2, the phrasing effect (factor COND) at the peak of this wave shows strong interactions with each of the factors GROUP, STYLE and AREA (Table 2). The difference between phrased and unphrased melodies was strongest in Germans listening to Chinese melodies (see Fig. 4), as supported by the post-hoc analysis of the COND\timesSTYLE\timesGROUP interaction\(^2\). Close inspection of Figure 3C suggests that a plausible explanation for the (non-significant) familiarity effect on the P2 is the overlap with the P280. This would also explain the longer

\(^2\) First, we computed 2-way ANOVAs for both levels of STYLE. While for Western style, there was only a main effect of COND (F=32.0, p<0.001), mean difference between phrased and unphrased 2.1 \mu V and no interaction COND\timesGROUP, for Chinese music there was a main effect of COND (F=71.8, p<0.001, mean difference 3.3 \mu V) as well as an interaction COND\timesGROUP (F=10.9, p<0.005). Second, we further detailed this interaction by 1-way ANOVAs for each level of GROUP, revealed a significant difference between phrased and unphrased items for both groups, which was much larger for German (F=46.8, p<0.001, mean difference 4.5 \mu V) than for Chinese subjects (F=22.0, p<0.001, mean difference 2.0 \mu V).
latency of the P2 peak for the culturally unfamiliar condition.

*Time window V (including the CPS)*

The CPS effect of phrasing (more positive for phrased as compared to unphrased items, see Figures 3 & 4 and Table 2) was present in the parietal, but not in the frontal electrodes for both styles of music and across both groups of subjects. However, the effect seemed to be more prominent for Chinese melodies, irrespective of the subject group. Moreover, for both styles of music, there appears to be a right lateralization.

4. Discussion

The present study addresses the brain mechanisms of musical phrase boundary processing in subjects without formal musical training in a cross-cultural (Chinese vs. Western) setting. We investigated if cultural familiarity influences the processing of musical phrase boundaries. Cultural familiarity was quantified by the agreement between the enculturation of the subjects (i.e., whether they grew up in China or in Germany) and the cultural style of the music presented (i.e., classical Western or traditional Chinese). Moreover, we sought to illuminate the time course of phrase boundary processing and draw conclusions on the specific influences of musical training, musical style, and cultural familiarity on the different stages of processing. One question was, at what processing stage the bottom-up processing of phrase boundaries starts to be affected by top-down influences based on musical training (in comparison to earlier published data on musicians – Neuhaus et al., 2006; Nan et al. 2006) and cultural familiarity. This was expected to cast light on the respective roles of procedural and non-procedural long-term memory in music perception. Another question that was investigated in this work was whether the music CPS, which has been reported as a marker for phrase boundary processing in musicians (Knösche et al. 2005, Nan et al. 2006), but found absent in an experiment with non-musicians (Neuhaus et al. 2006), is in actual fact specific to
musicians, or whether it indeed reflects a universal process. If found in non-musicians, it would be interesting to quantify similarities and differences with the CPS in musicians, in particular whether the STYLE effect reported by Nan et al. (2006) (i.e., the CPS is smaller for Western music) is due to the formal musical training both musician groups had received in Western music, or if it is a consequence of the differences in rhythmic, metric and melodic structure between the two styles of music (see Table 1).

In the following sections, we will discuss our findings in light of these questions. Note that there are significant differences in basic stimulus properties (e.g., pause length and length of prefinal tone, see Table 1) between Chinese and Western melodies. It is of course inevitable that the differences in style are somehow reflected by physical stimulus properties. This means however, that main effects of STYLE and interactions STYLE × AREA can be interpreted as pure bottom-up effects of physical stimulus properties. Therefore, we exclusively discuss effects when they involve the factor CONDITION.

**Behavioral findings**

Behavioral results from the melody categorization task were clearly influenced by cultural familiarity. Both groups categorized the musical excerpts of their native style more accurately than the ones of the other culture. This finding is consistent with previous studies (Arikan et al., 1999; Trainor and Trehub, 1992). As predicted, Chinese non-musicians performed equally well on Western classical music as Germans (which is plausible because long-term exposure to Western music can be assumed for Chinese subjects). For traditional Chinese music, the Chinese subjects were even better, while Germans’ performance dropped considerably (Figure 2A). In general, the cultural familiarity effect suggested by the present behavioral results is consistent with previous research (Drake and El Heni, 2003; Lynch and Eilers, 1992; Trainor and Trehub, 1992). Moreover, it was only in the case of agreement between the cultures of listeners and music (i.e., high cultural familiarity) that phrased musical pieces seemed to be
easier to recognize than *unphrased* ones (where the pause was filled with additional notes), suggesting that enculturation interacts with the recognition of phrase boundary cues (Figure 2B). This finding is in line with previous research on musicians (Nan et al. 2006). It seems possible that pauses and other phrase boundary cues serve not only as cues for phrase boundary processing (Frankland and Cohen, 2004; Knöschke et al., 2005; Nan et al., 2006; Neuhaus et al., 2006) but also represent certain musical features differing between musical styles. To summarize, we state that the behavioral data support a clear effect of cultural familiarity in music perception, which interacts with the phrase structure of the stimuli.

**Early ERP components – the acoustic change complex**

While the most primitive features of auditory input are already processed during the first 20-30 ms after stimulus onset in subcortical structures like the brainstem and thalamus (Davis, 2005; Escabi and Read, 2005; Galazyuk and Feng, 1997; Hall, 2005; Hall and Feng, 1986; Jen and Wu, 2006; Kadner et al., 2006), the more refined structural aspects of the stimuli are processed during approximately the next 250 ms in the primary auditory cortex and adjacent temporal areas, and are reflected by ERP components like P1, N1 and P2. These processes involve not only the discrimination of pitch, timbre, duration and intensity (Escabi and Read, 2005; Formisano et al., 2003; Penhune et al., 1999; Zatorre, 1988), but also a whole range of more complex operations, many of which are concerned with the integration of acoustic information over time. These operations include auditory stream segregation (Sussman et al., 1999), sound object formation (Sussman et al., 1998), auditory anticipation (Tervaniemi et al., 1994), extraction of abstract sound patterns and rules (Paavilainen et al., 1998; Paavilainen et al., 2001; Saarinen et al., 1992), as well as the integration of sound elements on the basis of learned rules in the linguistic (Friederici, 1995; Friederici et al., 1993) and musical domains (Koelsch, 2005; Koelsch et al., 2005; Koelsch et al., 2002b).

It is clear that the above-mentioned processing steps, in particular those concerned with
integration of information over time, are essential for any segmentation process. For example,
the detection of musical phrase boundaries requires the integration of the last few tones of the
previous phrase (in order to detect phrase boundary markers like pre-final lengthening and
harmonic function of the last tone) with the presence of a pause and the first tone of the next
phrase (see e.g., Knösche et al. 2005). Similar requirements apply to the detection of prosodic
phrase boundaries in speech (Pannekamp et al., 2005).

We therefore would expect the early ERP to be reflective of the presence of phrase boundaries.
We found an enhancement of P1, N1 and P2 for phrased, as compared to unphrased, melodies.
This increase was strongest for the P2. According to the statistical analysis, these effects were
influenced by the musical style of the input, but not by the cultural background of the subjects
(see Table 2). These data suggest that the underlying processing steps reflected in the early
ERP represent bottom-up processing driven by the physical differences between the stimuli,
rather than top-down processing (e.g., involving actively learned rules or long-term memory
content acquired by passive enculturation).

When comparing the ERP traces with the results reported by Nan et al. (2006) in musicians
using the same stimulus material, we observe some differences. In particular, the musician
study did not yield any detectable P1 while the N1 also seemed to be very small compared to
the present data. In contrast, the P2 appeared to be slightly larger for the musicians. This
apparent P2 difference between the two studies is in agreement with earlier claims that this
component reflects aspects of auditory processing influenced by musical training (Shahin et
al., 2005) and with findings from an artificial language study by De Diego Balaguer and
colleagues (De Diego Balaguer et al., 2007), who showed that the P2 is modulated as a
function of rule-learning. It should be noted, however, that these are qualitative observations,
which are not based on direct statistical comparisons between the two experiments.

Hence, it seems that the early ERP components up to about 250 ms post-stimulus might
reflect bottom-up processing of phrase boundaries, which can be modified by intensive training. Top-down influences on the early ERP components based on the cultural background of the subjects are too weak to be detected, and cultural background seems to only fully come into play at later processing stages (see next sections). At least for P1 and N1, familiarity with the presented style of music (i.e., for Germans listening to Western music and Chinese listening to either type of music) or a lack of familiarity (Germans listening to Chinese music) appear to have no influence.

Unfortunately, the specific functional significance of P1, N1 and P2 is at present poorly understood (for a recent review, see Key et al., 2005) and each of them is probably a conglomerate of many overlaying components (for the N1 refer to Näätänen and Picton, 1987). Nonetheless, all three waves have been shown to be sensitive to the physical parameters of the tones and to suppression by short interstimulus intervals (sensory gating, see e.g. Müller et al., 2001). This observation is consistent with the notion of the acoustic change complex (ACC) (Martin and Boothroyd, 1999), meaning an enhancement of the P1-N1-P2 complex with changes in continuous sound input. The ACC is well-documented in studies using auditory stimuli such as clicks (Ponton et al., 2000), and naturally produced speech sounds (Friesen and Tremblay, 2006). Moreover, some recent studies report ACC-like ERP patterns with musical stimuli (Nan et al., 2006; Neuhaus and Knösche, 2006).

Hence, one possible explanation of the observed phrase boundary dependent effects might lie in the simple fact that, for the unphrased condition, the onset of the second phrase was immediately preceded by sound input, while for the phrased condition there was silence. This means that for the phrased condition, there is a greater acoustic change (from silence to tones) than in the unphrased condition (from tones to tones), leading to an enhanced ACC. However, this is most probably not an exhaustive explanation, since it cannot easily account for the

---

1 However, as with many differences observed between musicians and non-musicians, it is not clear whether the observed differences are due to cortical plasticity caused by training, or whether they reflect congenital differences that have eventually led the subjects to become musicians or not.
apparent qualitative differences in the ACC between the musician experiment (Nan et al., 2006) and the present non-musician study. Moreover, there are some recent reports underpinning the relevance of the components of the ACC processing the sequential structure of input. For example, Snyder and colleagues (Snyder et al., 2006) demonstrated the role of the ACC and, in particular, of the P2 for frequency-based auditory stream segregation. Moreover, De Diego Balaguer and colleagues (De Diego Balaguer et al., 2007) showed that the amplitude of the P2 correlates with rule-learning in language and concluded that the changes in the P2 amplitude might reflect reallocation of attention to learning of grouping dependencies. This view is further corroborated by the results of an fMRI experiment on musical phrasing in musicians (Nan et al., 2008), employing a similar paradigm as in the present work. In that study, brain areas belonging to the fronto-parietal attention network were found to be influenced by both the phrase boundary and the musical style. The physiological basis of the ACC components has been investigated by a number of researchers. At least part of N1 is generated in the primary auditory cortex (Ponton et al., 2000) driven by the lemniscal pathway (as suggested by e.g. the tonotopic organization, see Pantev et al., 1988), although intracranial recordings as well as MEG source localization also suggest generators in lateral Heschl’s gyrus and the anterior planum temporale, both of which belong to the secondary auditory cortex (Liegeois-Chauvel et al., 1994; Lütkenhöner and Steinstrater, 1998). It seems that the generators of the N1 are rather complex and not localized to a small area (Liegeois-Chauvel et al., 1994; Näätänen and Picton, 1987).

The generators of the P1 have been localized using intracranial recordings to the lateral portion of Heschl’s gyrus, presumably encompassing parts of both the primary and secondary auditory cortex (Liegeois-Chauvel et al., 1994). Using MEG source analysis, it has been localized to the planum temporale, which incorporates the secondary auditory cortex (Reite et

Note that these fMRI activations are certainly not candidates for the physical sources of the ACC. The reason for this discrepancy may lie in the very different spatio-temporal apertures of the EEG and fMRI methods: EEG has high temporal with poor spatial resolution and fMRI has high spatial with poor temporal resolution.
al., 1988), which is thought to be the principal cortical target site for non-lemniscal thalamic neurons (Weinberger and Diamond, 1987). The non-lemniscal auditory pathway is involved in multisensory integration, learning and discrimination of temporal sound patterns (Kelly, 1973; Layton et al., 1979), and the processing of relevant or emotionally laden stimuli (see Hu, 2003 and the references cited therein). It receives input from the mesencephalic reticular activation system and, apart from the cortex, it projects to limbic structures, like the amygdala (Ponton et al., 2000).

It has been suggested that the generation of the P2 is linked to the mesencephalic reticular activation system (Ponton et al., 2000). In contrast, its immediate physical generators seem to be located in the non-primary auditory cortex (Knösche et al., 2003).

From these considerations, it is evident that P1, N1 and P2 originate from different physical generators and presumably bear differential functional significance, which is in contrast to the ACC notion. This is corroborated by the present study and the results of Nan et al. (2006), suggesting that these early components are influenced differently by the style of the music presented and the musical expertise of the subjects. How the different ERP effects might map on particular sub-processes of phrase boundary processing must remain speculative in the face of the present data and will require future research. The results, however, suggest that all phrase boundary processing stages reflected by the ACC components are subject to influence by musical training, while there is probably no effect of the specific enculturation of the subjects and consequently of cultural familiarity.

**The P280 – a P3?**

Riding on the falling flank of the P2, another positive peak occurs at about 280 ms post-stimulus, which we termed P280. This component is exclusively present for phrased items and represents the first component that, in the present data, shows a significant interaction between GROUP and STYLE in that it is much enlarged when German subjects process the
phrase boundaries in Chinese music, which is unfamiliar to them (note that to Germans the Western music style, and to Chinese both styles, can be considered familiar). It therefore reflects the effect of cultural familiarity on phrase boundary processing. In terms of polarity, topography and latency, this component resembles a novelty P3 (see, e.g. Mecklinger & Ullsperger, 1995). The novelty P3 (often associated, but not strictly identical with the P3a, see Simons et al., 2001) is considered to reflect the evaluation of unexpected events for further cognitive processing or behavioral action. It usually occurs if a deviance or surprise is large enough to capture attention. For a review, see, e.g., Friedman et al. (2001).

In the present experiment, for Western classical music neither German nor Chinese subjects should be surprised by the re-onset after the phrase boundary marking pause, since they are familiar with that style of music and can interpret the global structure of the melodies and the specific cues at the end of the first phrase (e.g., harmonic ending with a dominant tone) in such a way that they expect a continuation of the music after the pause. The same applies for the Chinese subjects when listening to traditional Chinese music, which, according to their own judgment and the behavioral results of this study, they also are familiar with. However, when Germans listen to the unfamiliar Chinese melodies, they might be surprised by the re-onset of the melody after the pause and have to revise their model for the structure of the perceived melody. Hence, it seems plausible to interpret the P280 as a novelty P3 component.

The absence of a distinguishable P280 peak in the study by Nan et al. (2006) might mean that it occurs earlier and blends with the P2. This is likely because, in that study, the P2 shows the same cultural familiarity effect on phrase boundary processing that we observed for the P280 in the present experiment. Hence, it might be the case that non-musicians integrate top-down information associated with their own enculturation with bottom-up information related to the musical style somewhat later than musicians do. This could be taken as evidence that musical training facilitates the top-down modulation of bottom-up processing chains. This interpretation is in agreement with recent findings on the influence of musical training on the
ability to detect prosodically incongruent endings in speech (Marques et al., 2007).

The CPS in non-musicians

The present data reveal a music CPS in German and Chinese non-musicians. The presence of the CPS to some extent replicates previous studies about the ERP correlates and of phrase boundary processing in music (Knösche et al., 2005; Nan et al., 2006; Neuhaus et al., 2006), intonational phrase boundary perception in spoken language (Steinhauer et al., 1999; Steinhauer and Friederici, 2001) and in hummed language (Pannekamp et al., 2005). The finding of a CPS in non-musicians in the current study is novel and different from the earlier ERP study by Neuhaus et al. (2006) with non-musicians, where no CPS but an earlier frontal negativity was found. There is, however, an important difference between the present experiment and the study by Neuhaus et al. (2006). The task used by Neuhaus et al. (2006) comprised the detection of off-key notes, which required subjects to focus their attention on the local structures of the music (single notes). In contrast, the current music style distinction task requires relatively holistic processing. Therefore, a possible explanation for the difference between the results of the two studies lies in the fact that, for musicians, the recognition of tonal relationships might be rather automatic and require few resources, while for non-musicians it might require many more resources. Musicians are therefore still able to process the melodies in a natural, holistic way, making use of the phrase structure, while in non-musicians the task used by Neuhaus et al. (2006) interferes with normal processing of the melodies. In contrast, the categorization task used in the present study requires the processing of the melodies as a whole in the first place and therefore does not necessarily compete with the processing of phrase structure. These considerations suggest that the presence of the music CPS reflects a fairly universal process that is not a result of formal musical training. This is in line with previous findings, suggesting that non-musicians might share fundamental processes

Note that a CPSm was found in MEG results by Neuhaus et al. (2006). As discussed in the introduction section, this suggests that EEG and MEG reflect different aspects of the neural activity underlying phrase boundary processing.
with musicians in music cognition (Bigand et al., 1999; Bigand and Pineau, 1997; Koelsch et al., 2000; Regnault et al., 2001), notwithstanding the many differences that have also been documented (Besson and Faita, 1995; Fujioka et al., 2004; Hantz et al., 1992; van Zuijen et al., 2004).

It should be noted that the currently observed CPS in non-musicians is right-lateralized whereas the CPS in musicians showed no such lateralization effect (Nan et al., 2006). With many studies (Besson and Faita, 1995; Fujioka et al., 2004; Hantz et al., 1992; van Zuijen et al., 2004) documenting different neural substrates in music processing between musicians and non-musicians, the topographical difference of the CPS between two groups is not the first to demonstrate a right-lateralized pattern in non-musicians (Shahin et al., 2003). This implies that the presence of the music CPS may not depend on musical expertise, but its topographical distribution seems to rely on music training, suggesting a modification of the underlying neural processes.

Source localization studies using different approaches with different sensitivity profiles suggest that the generators of the CPS might include structures of the limbic system normally involved in cognitive functions like attention and memory (Knösche et al., 2005) as well as cortical sites, e.g., in the planum temporale (Knösche et al., 2006). Also, an fMRI study with similar material (Nan et al., 2008) yielded phrase boundary related activations in brain areas known for their role in working memory (inferior frontal gyrus) and attention (medial frontal gyrus and intraparietal sulcus). Therefore, the component might be associated with fairly universal segmentation processes of musical (or maybe more generally, acoustic) input, as cultural familiarity plays no detectable role. See Knösche et al. (2005) for a more thorough discussion.

Finally, the present results offer some clarification of the effect of STYLE in the CPS time window observed for musicians (Nan et al., 2006) in that the amplitude for Western music was lower for both ethnic groups. Two alternative explanations were offered for this effect:
systematic property differences between the two melody types (i.e., different rhythm, meter, interval structure, etc.) and alternatively the fact that both groups received their formal musical training in Western and not in Chinese music. The finding that this effect was present in the current study with non-musicians as well indicates that indeed the former explanation is more likely.

**Conclusion**

The current study investigates phrase boundary processing in subjects without formal musical training. Special focus was placed on the effect of cultural familiarity (i.e., the relationship between the cultural background of the subjects and the cultural style of the music presented) on the different phases of the process. Moreover, the results were qualitatively compared to previous results with musicians (Nan et al. 2006) in order to gain insight into the role of formal musical training.

The behavioral results show that cultural familiarity indeed affects the cultural categorization of melodies and that this effect interacts with the presence of phrase boundary marking pauses. Therefore, the processing of phrase structure depends on cultural familiarity.

The precise timing of the process was investigated with ERP. The data show that in non-musicians, cultural familiarity affects phrase boundary processing that takes place between approximately 200 and 350 ms post-stimulus. Comparison with previous work (Nan et al., 2006) suggests that, in non-musicians, the integration of top-down information associated with subjects’ enculturation and bottom-up information related to the musical style might occur about 100 ms later than in musicians, indicating that musical training facilitates the top-down modulation of bottom-up processing chains. In contrast to cultural familiarity, musical training appears to modify phrase boundary processing very early, from at least about 20-50 ms after stimulus onset.

Furthermore, we demonstrated that a music CPS can be elicited regardless of musical training
if the task leaves musically untrained subjects sufficient free resources to process the music (and therefore its phrase structure) in a holistic manner. This study thus provides clarification of the result of Neuhaus et al. (2006), who found no music CPS in non-musicians. It suggests that it is a question of capacity, rather than capability as to whether the later phrase boundary processing, as indexed by the music CPS, takes place in a specific experimental situation. These results strengthen the notion that the music CPS reflects a fairly universal component of phrase boundary processing that does not require any formal musical training.

Acknowledgements

This study was conducted in Leipzig and is part of the first author’s thesis. Part of the work is supported by National Natural Science Foundation of China (30700225).
References


Kadner, A., Kulesza, R.J., Berrebi, A.S., 2006. Neurons in the medial nucleus of the trapezoid body and superior paraolivary nucleus of the rat may play a role in sound duration coding. Journal of Neurophysiology 95, 1499-1508.


Tables

Table 1: Comparison of stimulus properties that might be relevant for phrase boundary processing in Chinese and Western melodies. Values in columns 2 and 3 are averages, with standard errors in parentheses. The last column shows the unpaired two-tailed t-test with adjustment of the equality of variance, computed with SPSS (TM) (118 degrees of freedom).

<table>
<thead>
<tr>
<th></th>
<th>Chinese melodies</th>
<th>Western melodies</th>
<th>unpaired two-tailed t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pause length (ms)</td>
<td>628(39)</td>
<td>504(25)</td>
<td>t=2.676, p=0.009</td>
</tr>
<tr>
<td>Length of preceding phrase (ms)</td>
<td>8535(361)</td>
<td>5290(209)</td>
<td>t=7.766, p&lt;0.001</td>
</tr>
<tr>
<td>Length of the tone before pause (ms)</td>
<td>765(58)</td>
<td>511(33)</td>
<td>t=3.808, p&lt;0.001</td>
</tr>
<tr>
<td>Length of the tone after pause (ms)</td>
<td>296(21)</td>
<td>248(26)</td>
<td>t=1.455, p=0.148</td>
</tr>
</tbody>
</table>
Table 2: Results of 4-way ANOVA COND*STYLE*GROUP*AREA. Only p-values below 0.1 are shown. Marginally significant (0.05<p<0.1) values are shown in gray. The degrees of freedom are given in brackets in the first column. Interaction effects are detailed in post-hoc analysis (see text).

<table>
<thead>
<tr>
<th></th>
<th>I (25-55ms)</th>
<th>II (75-105ms)</th>
<th>III (135-210ms)</th>
<th>IV (260-310ms)</th>
<th>V (450-600ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>COND (1,32)</strong></td>
<td>F=7.05</td>
<td>F=10.1</td>
<td>F=102.9</td>
<td>F=57.7</td>
<td>F=17.9</td>
</tr>
<tr>
<td></td>
<td>p&lt;0.05</td>
<td>p&lt;0.005</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td><strong>COND*STYLE (1,32)</strong></td>
<td>F=4.4</td>
<td>F=6.9</td>
<td>-</td>
<td>F=17.4</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>p&lt;0.05</td>
<td>p&lt;0.05</td>
<td>-</td>
<td>p&lt;0.001</td>
<td>-</td>
</tr>
<tr>
<td><strong>COND*GROUP (1,32)</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>F=6.9</td>
<td>F&lt;0.05</td>
</tr>
<tr>
<td><strong>COND*AREA (3,30)</strong></td>
<td>F=9.4</td>
<td>-</td>
<td>F=9.7</td>
<td>F=10.3</td>
<td>F=40.8</td>
</tr>
<tr>
<td></td>
<td>p&lt;0.001</td>
<td>-</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td><strong>COND<em>STYLE</em>GROUP (1,32)</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>F=6.2</td>
<td>-</td>
</tr>
<tr>
<td><strong>COND<em>STYLE</em>AREA (3,30)</strong></td>
<td>-</td>
<td>F=2.7</td>
<td>-</td>
<td>F=4.0</td>
<td>F=5.0</td>
</tr>
<tr>
<td></td>
<td>p&lt;0.1</td>
<td>-</td>
<td>p&lt;0.05</td>
<td>p&lt;0.01</td>
<td>p&lt;0.01</td>
</tr>
<tr>
<td><strong>COND<em>GROUP</em>AREA (3,30)</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>COND<em>STYLE</em>GROUP*AREA (3,30)</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>STYLE (1,32)</strong></td>
<td>-</td>
<td>F=10.9</td>
<td>F=3.6</td>
<td>F=26.4</td>
<td>F=3.5</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>p&lt;0.005</td>
<td>p&lt;0.1</td>
<td>p&lt;0.01</td>
<td>p&lt;0.1</td>
</tr>
<tr>
<td><strong>STYLE*GROUP (1,32)</strong></td>
<td>-</td>
<td>-</td>
<td>F=5.2</td>
<td>F=3.5</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>p&lt;0.05</td>
<td>p&lt;0.1</td>
<td>-</td>
</tr>
<tr>
<td><strong>STYLE*AREA (3,30)</strong></td>
<td>F=5.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>p&lt;0.01</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>STYLE<em>GROUP</em>AREA (3,30)</strong></td>
<td>F=3.4</td>
<td>F=3.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>p&lt;0.1</td>
<td>p&lt;0.05</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>GROUP (1,32)</strong></td>
<td>-</td>
<td>-</td>
<td>F=6.4</td>
<td>F&lt;0.05</td>
<td>-</td>
</tr>
<tr>
<td><strong>GROUP*AREA (3,30)</strong></td>
<td>-</td>
<td>F=5.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>p&lt;0.05</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Figures

A

Chinese phrased (c38p)

Chinese unphrased (c38c)

Chinese phrased (c38p)

Chinese unphrased (c38c)

Western phrased (w40p)

Western unphrased (w40c)

Western phrased (w40p)

Western unphrased (w40c)

Combined

B

Figure 1: Experimental setup. A: Paradigm of one experimental trial. ISI – inter-stimulus interval. B: Examples scores for stimuli.
Figure 2: Overview of the behavioral data. Error bars indicate standard deviation. A: Percentage of correct answers for all combinations of STYLE and GROUP, collapsed over COND, illustrating the STYLE×GROUP interaction. B: Percentage of correct answers for all combinations of COND, STYLE and GROUP, illustrating the COND×STYLE×GROUP interaction.
Figure 3: Overview of the ERP data, average with respect to phrase boundary offset, in 9 representative channels. The analysis time windows are shown as gray shaded bars. The arrows indicate the peaks of the identified components. A: Western music. B: Chinese music. C: Differences between phrased and unphrased conditions for both styles and groups.
Figure 4: Topographic maps at the peak latencies of the different components for the phrased and unphrased conditions as well as the difference phrased-unphrased for the four music style/subject groups. Note that the color scale is different for each time window. TW I-V represent the different time windows analyzed. P1, N1, P2, P280 and CPS represent the different ERP components discussed in the text. The maps are viewed from above, with the nose pointing upwards (see legend).