Perception of Phrase Structure in Music

Thomas R. Knösche,[1]***** **Christiane Neuhaus,**[2] **Jens Haueisen,**[2] **Kai Alter,**[1]
Burkhard Maess,[1] **Otto W. Witte,**[2] **and Angela D. Friederici**[1]

[1]*Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[2]*Friedrich-Schiller-University, Jena, Germany*

◆━━━━━━━━━━━━━━━━━━━━━◆

Abstract: Neither music nor spoken language form uniform auditory streams, rather, they are structured into phrases. For the perception of such structures, the detection of phrase boundaries is crucial. We discovered electroencephalography (EEG) and magnetoencephalography (MEG) correlates for the perception of phrase boundaries in music. In EEG, this process was marked by a positive wave approximately between 500 and 600 ms after the offset of a phrase boundary with a centroparietal maximum. In MEG, we found major activity in an even broader time window (400–700 ms). Source localization revealed that likely candidates for the generation of the observed effects are structures in the limbic system, including anterior and posterior cingulate as well as posterior mediotemporal cortex. The timing and topography of the EEG effect bear some resemblance to a positive shift (closure positive shift, CPS) found for prosodic phrase boundaries during speech perception in an earlier study, suggesting that the underlying processes might be related. Because the brain structures, which possibly underlie the observed effects, are known to be involved in memory and attention processes, we suggest that the CPS may not reflect the detection of the phrase boundary as such, but those memory and attention related processes that are necessary to guide the attention focus from one phrase to the next, thereby closing the former and opening up the next phrase. *Hum Brain Mapp 24:259–273, 2005.*   © **2005 Wiley-Liss, Inc.**

Key words: music perception; prosody; magnetoencephalography; event-related potentials

◆━━━━━━━━━━━━━━━━━━━━━◆

INTRODUCTION

Phrasing is an important means of structuring auditory streams and, hence, facilitating their processing by the human brain. It is equally important for the domains of speech and music. Phrasing in language has been investigated to some extend in linguistics [e.g., Selkirk, 2000] and in psy-

cholinguistics using behavioral methods [e.g., Stirling and Wales, 1996; Warren et al., 1995] and more recently using neurophysiological methods [Steinhauer et al., 1999; Steinhauer, 2003]. In language, phrase boundaries often indicate the closure of a syntactic phrase [Selkirk, 2000]. Phrasing in music denotes the division of the melodic line into structural subunits, i.e. it is the segmentation of a musical thought for purposes of musical sense. According to the theoretician and musicologist Riemann [1900], one of the main representatives of phrasing theory at the end of the 19th century, a *phrase* has to be regarded as a separate musical entity within the melodic line. A musical phrase has a longer duration than a melodic *motif*, but a shorter one than a so-called "musical *period*". Stoffer [1985] viewed the musical phrase as one level in a complex hierarchy describing both formal description and internal representation of music.

For the correct processing of phrasing, the recognition of *phrase boundaries* is necessary. These boundaries are indicated by certain structural markers, which were listed by Riemann and other theoreticians of the 18th up to the 20th

century [Keller, 1955; Mattheson, 1737]. For detecting phrase boundaries, Riemann [1900] considered two cues as most relevant: small breaks immediately after a strong beat and the lengthening of notes, which coincide with a strong beat. In addition, Riemann mentioned several other phrase-establishing cues, including pitch fall and pitch rise in vocal passages or the complete reversal of melodic movement. Harmonic progressions (e.g., cadences and semicadences) have also been described as markers for phrase boundaries [Cuddy et al., 1981, Tan et al. 1981].

While the importance of phrasing for the perception of music by humans has been illuminated by several behavioral studies [Chiappe and Schmuckler, 1997; Dowling, 1973; Stoffer, 1985; Tan et al., 1981; Wilson et al., 1999], there is no report of a direct neural correlate. So far, there has only been one attempt to use event-related potentials (ERPs) to study the processing of musical phrase boundaries in a mental segmentation task [Paulus, 1988]. The author reported a slow positive shift related to the phrase boundaries. However, this study has to be considered preliminary, as no explicit statistical tests were given and the number of subjects was limited to four.

In the domain of language, the subdivision of auditory streams into phrases is of great importance, too. Although phrase boundaries in spoken language can in principle be retrieved by pure syntactic rules in most cases (there are ambiguities), phrase boundary markers greatly aid a speedy and accurate processing. In speech, such markers are of prosodic nature. They mainly consist in changes in the temporal structure, especially in the insertion of pauses, but also in pitch and volume changes; therefore, they are very similar to the ones found in music. (For a thorough account on the role of prosody in language comprehension, see Cutler et al. [1997].) In a recent study, direct electrophysiological evidence was presented for the immediate use of prosodic markers during syntactic parsing [Steinhauer et al., 1999]. The authors used ERP and found that each phrase boundary was marked by a positive shift of some hundred milliseconds duration and a centroparietal distribution. This ERP component was called the *closure positive shift* (CPS)—as it was taken to reflect the process of closing the prior phrase. They also demonstrated that this effect is not simply related to the pause at the boundary but to an entire ensemble of prosodic cues, including lengthening of the last syllable before the boundary and changes in the F0 contour. Steinhauer and Friederici [2001] used filtered delexicalized speech to demonstrate that really *prosodic* rather than *syntactic* cues give rise to the CPS. Moreover, Pannekamp et al. [2004] could show that phrase boundaries in hummed sentences elicit a CPS, although with a right lateralized topography, suggesting that the CPS is not dependent on segmental linguistic information. Further studies demonstrate that the CPS does not necessarily follow each prosodic phrase boundaries during speech perception but that its presence is influenced by the information structure built up before the target sentence. If in a dialogue the focus of attention has been directed toward a certain word (e.g., by a preceding

question), the CPS follows this accentuated word instead [Hruska and Alter, 2004; Toepel and Alter, 2003]. Hence, the CPS seems to reflect closure processes that can be the consequence of the perception of phrase boundaries, rather than the perception of these boundaries itself. The close relationship between music and spoken language on a syntactic/prosodic level [Patel, 2003] and the similarity between the phrase boundary markers in both domains lead to the hypothesis of a similar neural correlate in music.

In summary, there is evidence: (1) for the notion that phrasing is a common phenomenon in both music and language, (2) for the importance of phrasing for the perception of music, and (3) for an electrophysiological correlate of the processing of intonational phrase boundaries in language. However, there is no account for a neural correlate for phrase structure processing in music. Hence, the aims of this study are: (1) to identify an electrophysiological marker for the processing of phrase boundaries in music, (2) to establish its topological and morphological properties as well as its underlying generator structure, and (3) to determine its relationship to the CPS found in language studies. For this purpose, we developed a combined electroencephalography (EEG) and magnetoencephalography (MEG) study, where musical excerpts of different phrase structure were presented to trained musicians. MEG has been successfully used for gaining insight into various aspects of the perception of music, e.g., perception of pitch [e.g., Patel and Balaban, 2000, 2001] and timbre [Pantev et al., 2001], processing harmonic violation [Maess et al., 2001], involuntary activation of motor programs in musicians [Haueisen and Knösche, 2001], and multimodal information processing in musicians [Schulz et al., 2003]; therefore, it seems to be suitable for our purpose.

SUBJECTS AND METHODS

Subjects

We recruited 12 trained musicians (6 women; age range, 23–31 years; mean age, 25.8 years). None of them had any known neurological or hearing disorders. They were all right handed (average handedness 86.2, standard deviation 14.8) according to the Edinburgh Handedness Inventory [Oldfield, 1971]. The average starting age for playing an instrument was approximately 9 years. Both their initial and present main instruments were very diverse, including piano (2/3), guitar (2/2), flute (2/1), violin (3/2), and accordion (2/1). All of them were in the process of some formal musical training, 11 had already passed their intermediate or final exams. They reported on average to exercise for approximately 2 hours a day.

Subjects participated in four sessions on different days (two EEG and two MEG). Before the first experimental run, each subject was generally informed about brain activity and the measurement per se but was not given any further details about the matter of investigation. After each EEG and MEG recording session, participants were asked to fill out a questionnaire. (The questionnaire contained questions on

Figure 1.

Example of stimulus material. Each stimulus existed in a two-phrased basis version (referred to as *phrased*, top) and a single phrase variant (referred to as *unphrased*, bottom), where the pause has been filled by notes. Phrases are generally indicated by a slur, which is extended over a sequence of notes and spreading over the next following bar line in most of the cases. Note that slurs can also serve as playing instructions.

the familiarity and perceived quality of the stimuli, on the physical shape of the subject and on the experienced difficulty of the task. Questions on the subject's musical education were asked only once after the first recording session.) All subjects gave their written informed consent to participate in the study. Approval of the ethical committee was present.

Stimulus Material and Paradigm

To explore the suitability of stimulus material for the elicitation of phrase boundary related ERP and ERF components, we conducted a pilot experiment. The subjects participating in this study were quite comparable to the ones of the main study in age (22.8 vs. 25.8 years), handedness (83.7 vs. 86.2), musical education, and daily musical practice (2:45 vs. 2:00 h). For stimulation, we used 21 *phrased* melodies (containing two phrase boundaries), which were especially created by two professional composers. For each of the items, an unphrased version was created by filling the second phrase boundary with notes. For an example score, see Figure 1. The task of the subjects was to detect out-of-key notes that occurred at random positions in approximately 10% of the melodies. The melodies were presented in a pseudorandom order, where each of them was occurring eight times. Triggers were set at the offsets of the phrase boundaries. For more details of the experimental paradigm, please refer to the description of the main study.

Apart from clear differences on the N1 and P2 components, no later CPS-like effect could be found, which could be attributed to the presence of one of the phrase boundaries. This finding led us to the conclusion that the stimuli used did not trigger a sufficiently strong impression of phrasing in the subjects for several reasons. First, many of the melodies were, because they were especially composed for the experiment, very complex, and quite unusual in their metric and harmonic structure. In particular, most of the melodies (14 of 21) did not have a symmetric structure (a symmetrically structured melody can be partitioned into two sections of equal length and similar motive and rhythmic structure). Second, the phrasing and the phrase boundaries might not have been accentuated enough in composi-

tion and interpretation. Third, the lengths of the phrases were quite short. Finally, the fact that each melody was presented many times might have covered potential effects. For the main study, we, therefore, decided to use more regularly and symmetrically constructed musical excerpts with more pronounced phrase boundaries. Moreover, each item should be presented only once per subject.

We prepared 101 short melodic fragments, which were clearly divided into two phrases, the boundaries between which were marked by pauses. Some of them (20) were newly composed, whereas most were extracted from preexisting, although not very well-known musical pieces. Most melodies consisted of eight bars, which were equally distributed over the two phrases. For each of the original melodies (labeled *phrased*), a modified counterpart was created, where the pause was filled by one or several notes (labeled *unphrased*). This procedure was done off-line on the MIDI score produced by a synthesizer that was played by a professional pianist. This strategy ensured exact equality of all other aspects of the melody versions apart from the phrase boundary.

An additional set of incorrect melodies was generated by replacing one correct tone at a random position by a dissonant one ("out-of-key note") in each example. These versions served as task items and were excluded from further analysis. Some properties of the stimuli are summarized in Table I, also in comparison to the pilot experiment.

The experimental paradigm is sketched in Figure 2. To prevent any sequence effects, all correct melody versions were presented in pseudorandom order. The resulting 202 trials (each of the *phrased* and *unphrased* examples was presented exactly one time) were mixed up in different manners (balanced for sequence) and split into six blocks of equal length. Approximately 10% of the stimuli were replaced by their incorrect versions (selected randomly). These trials were excluded from any further analysis. Each experimental block was between 20 and 30 min long and contained an average of 17 phrased, 17 unphrased, and 3 out-of-key items. Three stimulus blocks were presented in one session (experimental durations above 90 min are generally experienced as very strenuous.), each subject participated in two experi-

TABLE I. Acoustic and structural parameters of stimuli

	Pilot experiment		Main experiment
	Phrased	Unphrased	phrased/ unphrased
Length of last tones before pause (in ms)			
3rd last tone	239 (45)	226 (51)	251 (35)
2nd last tone	299 (69)	262 (67)	334 (46)
Last tone	416 (86)	344 (76)	555 (61)
Pause length (in ms)	435 (108)		498 (42)
Length of preceding phrase (in ms)	2261 (946)	2519 (913)	4887 (331)
Symmetry of preceding phrase	33%		66%
Harmonic ending	91%		89%
Clear pitch movement of last 3 tones (up or down)	57%		47%

Values in parentheses are 95% confidence intervals. Note that for the main experiment, all these parameters were exactly identical for the phrased and unphrased versions.

mental sessions for EEG and in another two for MEG. The sequence of EEG and MEG recordings was also balanced between subjects, such that approximately half of them underwent MEG first and the other half EEG first.

Subjects were asked to judge, if the musical piece was completely in key or contained any dissonant tone (out-of-key tone). Note that they were not aware that the purpose of the experiment was the investigation of the perception of phrasing.

Recording Procedure

For the EEG recording, small Ag–AgCl electrodes were attached to the scalp, being inserted into a special electrode cap. The subjects were asked to sit down in an electrically shielded room and listen carefully to the acoustic examples. Stimuli were presented by means of loudspeaker, which was placed in a distance of approximately 1.5 m. Before starting the first block, loudness as well as centering of the sound source were balanced individually. For getting accustomed to the stimuli, several examples were presented before the

start of the actual experiment. Subjects were asked to avoid eye blinking and to relax their facial muscles. EEG signals were recorded from 25 electrode positions distributed over the entire scalp. In addition to 19 electrodes from the international 10–20 system (Fp1, Fp2, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1, O2), some electrodes from the 10–10 system [see, e.g., Oostenveld and Praamstra, 2001] were used (FT7, FC3, FC4, FT8, CP5, CP6).

The ground electrode was placed at electrode position C2 in the right hemisphere. The reference electrode was placed at the nose. For registering blink and ocular movement artifacts, vertical and horizontal electrooculograms (EOG) were recorded from above and below the right eye as well as from the outer canthus of both eyes. Impedances of electrodes were kept below 5 kΩ, and the time constant was infinite, using direct current amplifiers. Each measured raw EEG had a bandwidth from 0 Hz to 100 Hz and was digitized on-line with a sampling rate of 250 Hz. In an additional trace, trigger points used as markers for the onset of the first tone of each melody, were recorded and stored to synchronize the stimulus and EEG activity. In an offline procedure, trials with eye artifacts were marked and discarded. A high pass filter of 0.25 Hz was applied to reduce very slow drifts.

A 148-channel whole-head magnetometer system (MAG-NES WHS-2500; 4D Neuroimaging) was used to record the MEG. EOG was measured to identify epochs contaminated by eye artifacts. The head position was measured using five coils attached to the head, which were localized by the MEG system before and after each block. The MEG (sampling rate 508 Hz) data were also passed through a high pass filter of 0.25 Hz. Because it is not possible to restore a certain position of the magnetometer array exactly after the subject has moved, the data recorded from one subject during different sessions and blocks were first averaged within blocks (per condition) and then interpolated to a set of average sensor positions using a method based on linear inverse techniques [Knösche, 2002]. As a result of this procedure, the values from the different blocks represented the magnetic field at the same positions with respect to the head and could be averaged over blocks within each of the two sessions.

Signal Analysis

Triggers were placed at the *offsets* of the phrase boundaries. In the *unphrased* versions (where the phrase boundary
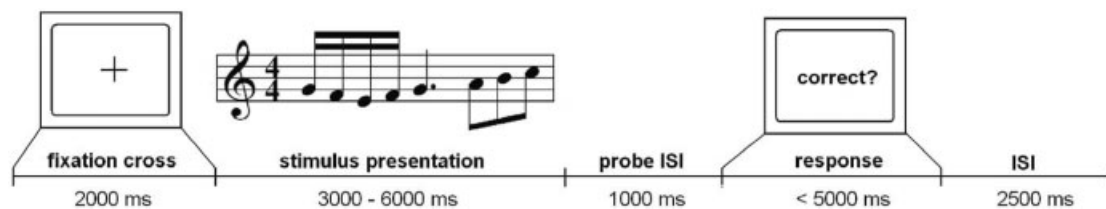
Figure 2.
Timing for the presentation of a single trial. ISI, interstimulus interval.
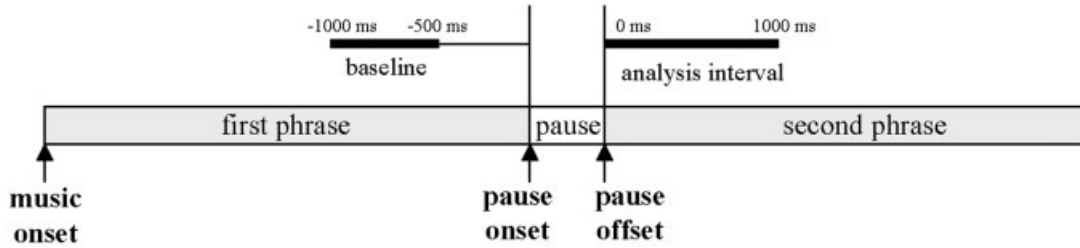
Figure 3.
Definition of the baseline interval. The baseline interval is defined with respect to the pause onset, whereas the analysis interval is defined relative to the pause offset.

had been filled with notes) triggers were set at those time points, which corresponded to the offset of the filled-in phrase boundary. The averaging window was chosen to range from 0 ms to 1,000 ms relative to the trigger points (i.e., to the offsets of the phrase boundaries). As baselines, we chose the time intervals from −1,000 ms to −500 ms relative to the *onsets* of the respective phrase boundaries (see Fig. 3). This step was done to make sure that the baseline could not contain any signal differential between the phrased and unphrased conditions. An obvious choice would have been to use an interval before the melody onsets, but it turned out that this time range was particularly contaminated by ocular artifacts in many data sets.

In the questionnaires, several subjects reported to be tired or depressed during the experimental sessions. Such subjects also tended to exhibit an unusual amount of ocular artifacts as well as alpha activity. Hence, only those data sets for which the subjects stated that their state of mind was "happy", "calm", or "alert" but not "tired" or "depressed" were taken to further analysis. (Adjectives to describe the state of mind were predefined and given to the subjects to choose from. They followed the examples set by standard personality tests, e.g., von Zerssen [1976] and Janke and Debus [1978].) Moreover, an artifact rejection procedure was applied to identify artifact-contaminated trials, which were excluded from further analysis. Sessions with more than 30% contaminated trials were excluded altogether. This process resulted in 13 EEG and 15 MEG sessions remaining for further analysis.

Averages of both EEG and MEG recordings were computed for every subject, session, and condition. Trials with out-of-key notes were excluded from the averages. Finally, the averaging over subjects, blocks, and sessions was performed. Visual inspection of the grand averages together with 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.

A selection of the electrodes were grouped into four regions of interest (ROI): left anterior (F7, F3, FT7, FC3), right anterior (F4, F8, FC4, FT8), left posterior (CP5, P3, P7, O1), and right posterior (CP6, P4, P8, O2). For each of the identified time ranges and ROIs, potential values were averaged over time steps and electrodes. For each time window, a three-way analysis of variance (ANOVA) was computed for

the resulting average values with the within-subject factors "melodic condition" (COND, 2 levels: *phrased* and *unphrased*), "anterior–posterior topography" (POST-ANT, 2 levels), and "right–left topography" (RIGHT–LEFT, 2 levels). Additionally, two-way ANOVAs were computed with the factors COND (2 levels) and CHANNEL (25 levels), to take the entire data into account. In all analyses, degrees of freedom were corrected with Huynh–Feldt epsilon.

The statistical analysis of the MEG amplitudes was set up in a very similar way. Four different ROIs (left anterior, right anterior, left posterior, right posterior) were formed from 138 channels (10 midline channels were excluded), each composed of 34 or 35 channels. The analogous three-factor ANOVA was computed. Again, another two-way ANOVA was computed with the factors COND (2 levels) and CHANNEL (148 levels).

Scalp potential and magnetic field maps were generated to shed additional light onto the spatial distribution of the discovered effects. Eventually, source analysis was used to reveal information on the generator configuration underlying the respective components.

Source Analysis

To identify the cerebral network underlying the processing of phrase boundaries, source localization methods were used. To eliminate the influence of any activity that is not related to the processing of phrase boundaries, the source analysis was performed on the differential activity between original and modified stimulus conditions. To obtain statistically meaningful results, this analysis was done for each session separately, rather than for the grand average over all subjects.

Because we expected a complex and distributed activation pattern, we did not attempt any simultaneous estimation of the whole network of active brain areas. Instead we computed an index for each point in the brain, reflecting whether or not this area could possibly have contributed to the measured data. Such a measure has been provided by Mosher et al. [1992] by virtue of the multiple signal classification (MUSIC) method. In this method, the measured data within the analysis time interval are first approximated by a linear combination of spatiotemporal components spanning the so-called *signal subspace*. Then, at each investigated po-
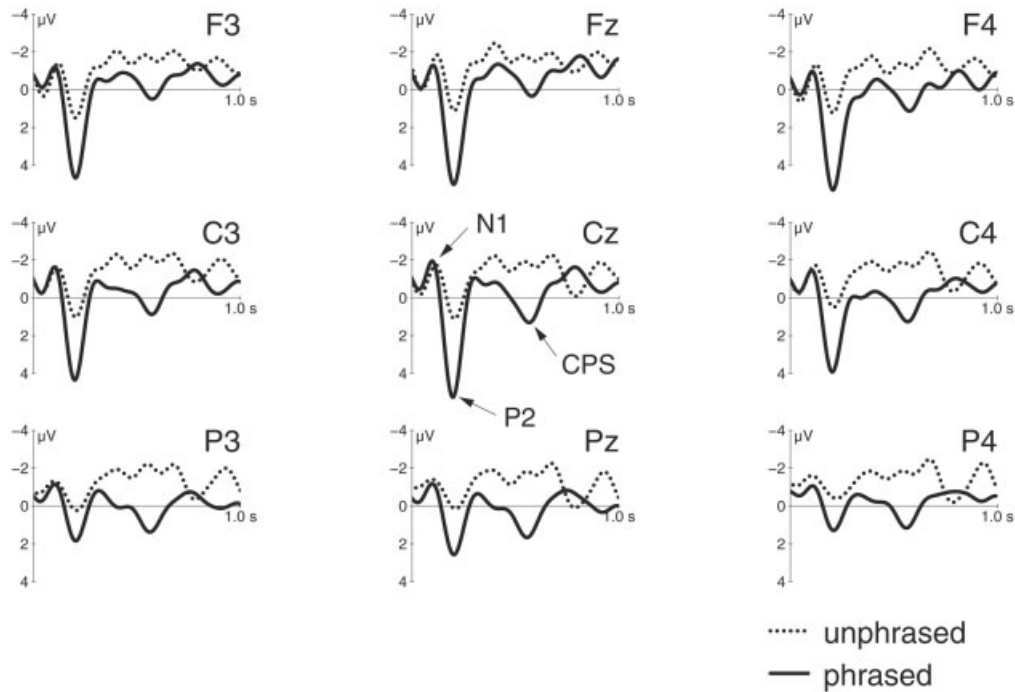
Figure 4.
Traces of selected electroencephalography channels. Solid lines denote the original condition with unmodified phrase boundaries (*phrased*). Dotted lines denote the modified condition, where the phrase boundary was filled with notes (*unphrased*). Units are microvolts. A low-pass filter of 8 Hz was applied for display purposes only.

sition within the brain, three orthogonal unity dipoles are placed and the corresponding EEG or MEG topographies are computed. These three topographies span another subspace, called the *source subspace*. The angle between both subspaces is then used as a criterion for the plausibility of a contribution of the respective brain region to the measured data. This way, we can identify brain regions that could have contributed to what has been observed, and those that could not have.

For MEG, the head geometry was accounted for by a boundary element model consisting of one triangulated surface describing the inside of the skull only [Hämäläinen and Sarvas; 1987]. A model with three surfaces representing the inside and outside of the skull as well as the outer head surface was used for the EEG based source reconstruction. Such models can be generated from magnetic resonance images (MRI). We used a standardized head model based on the Talairach scaled MRIs of 50 individuals. This standard model was then individually scaled to fit the shape of the subject's head as closely as possible in a least squares sense. Five independent scaling factors ensured a close match of the resulting head shape. The head shape information was recorded before the measurements using a *Polhemus FastTrak* system.

The MUSIC solution was computed on a grid of 8-mm width throughout the cerebral areas of the standard brain. This grid was scaled along with the head model to fit the

individual head shape. The analysis time window was chosen according to the intervals, where the experimental conditions differed in their EEG or MEG magnitude (500–600 ms for EEG and 400–700 ms for MEG). For each voxel, the solution value was averaged over all data sets. Areas with an average projection of the source subspace onto the signal subspace of larger than 0.875 (angle between subspaces less than 7 degrees) were accepted as potential candidates for contribution to the measured data.

RESULTS

EEG traces are shown in Figure 4. The results of the statistical analysis using ROIs of selected channels, which are classified according to two topographic factors, are presented in Table II. Statistics for all channels are given in Table III for comparison. The results are compatible. We observe an N1/P2 complex, where the N1 does not exhibit any significant difference between conditions, whereas the P2 is significantly reduced in amplitude for the *unphrased* condition ($P < 0.01$). More interestingly, a broadly distributed centroparietal deflection peaking at approximately 550 ms after pause offset was observed in the *phrased* version compared to the *unphrased* one ($F_{1,12} = 9.5$; $P < 0.01$). Although the topography of this component (see Fig. 5) seems to suggest a slight left lateralization, this finding did not prove statistically significant ($F < 1$).

TABLE II. Results of statistical analysis for EEG and MEG with *regions of interest*

Time window (ms)	EEG or MEG	main effect "condition" COND	Interaction (first order) COND × POST–ANT	Interaction (first order) COND × RIGHT–LEFT	Interaction (second order) COND × POST–ANT × RIGHT–LEFT
400–500	EEG	$F_{1,12} = 2.87$	$F_{1,12} = 0.11$	$F_{1,12} = 0.37$	$F_{1,12} = 0.22$
	MEG	$F_{1,14} = 0.22$	$F_{1,14} = 0.30$	$F_{1,14} = 6.43$ $P = 0.024$	$F_{1,14} = 13.50$ $P = 0.0025$
500–600	EEG	$F_{1,12} = 9.49$ $P = 0.01$	$F_{1,12} = 0.38$	$F_{1,12} = 0.80$	$F_{1,12} = 0.03$
	MEG	$F_{1,14} = 0.62$	$F_{1,14} = 2.62$	$F_{1,14} = 4.72$ $P = 0.048$	$F_{1,14} = 2.50$
600–700	EEG	$F_{1,12} = 4.88$ $P = 0.047$	$F_{1,12} = 0.11$	$F_{1,12} = 1.40$	$F_{1,12} = 0.03$
	MEG	$F_{1,14} = 0.06$	$F_{1,14} = 0.41$	$F_{1,14} = 5.79$ $P = 0.031$	$F_{1,14} = 18.69$ $P = 0.0007$

Only significant P values below 0.05 are shown.
EEG, electroencephalography; MEG, magnetoencephalography; COND, melodic condition; POST–ANT, posterior–anterior topography; RIGHT–LEFT, right–left topography.

MEG traces for some selected channels are displayed in Figure 6. The statistics for the ROIs is shown in Table II. Table III shows that the use of all 148 channels (including midline channels) produces compatible results. Similarly to the EEG results, the N1m did not differ significantly between conditions, whereas the P2m was considerably reduced in magnitude for the *unphrased* condition ($P < 0.01$). Also for MEG, the traces (Fig. 6) do show additional significant differences between conditions beyond the N1m/P2m complex. These effects are not concentrated in the analysis time window between 500 ms and 600 ms, where a positive deflection was observed in EEG, but extend to a larger time window, i.e., roughly between 400 and 700 ms. Moreover, the largest differences fall into time windows just before (400 ms to 500 ms) and just after (600 ms to 700 ms) the EEG

TABLE III. Results of statistical analysis for EEG and MEG with all channels.

Time window (ms)	EEG or MEG	Main effect "condition" COND	Interaction COND × CHANNEL
400–500	EEG	$F_{1,12} = 2.5$	$F_{24,288} = 0.71$
	MEG	$F_{1,14} = 0.33$	$F_{147,2058} = 4.51$ $P = 0.0001$
500–600	EEG	$F_{1,12} = 8.61$ $P = 0.01$	$F_{24,288} = 0.95$
	MEG	$F_{1,14} = 0.39$	$F_{147,2058} = 3.42$ $P = 0.02$
600–700	EEG	$F_{1,12} = 4.29$	$F_{24,288} = 0.78$
	MEG	$F_{1,14} = 0.13$	$F_{147,2058} = 4.95$ $P = 0.0004$

Only significant P values below 0.05 are shown.
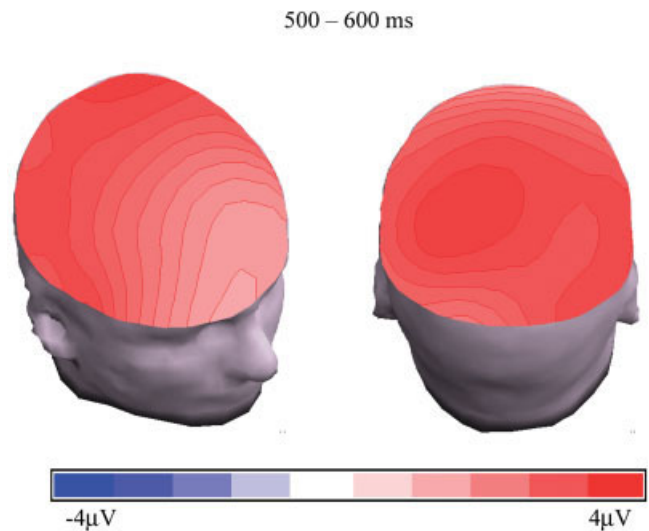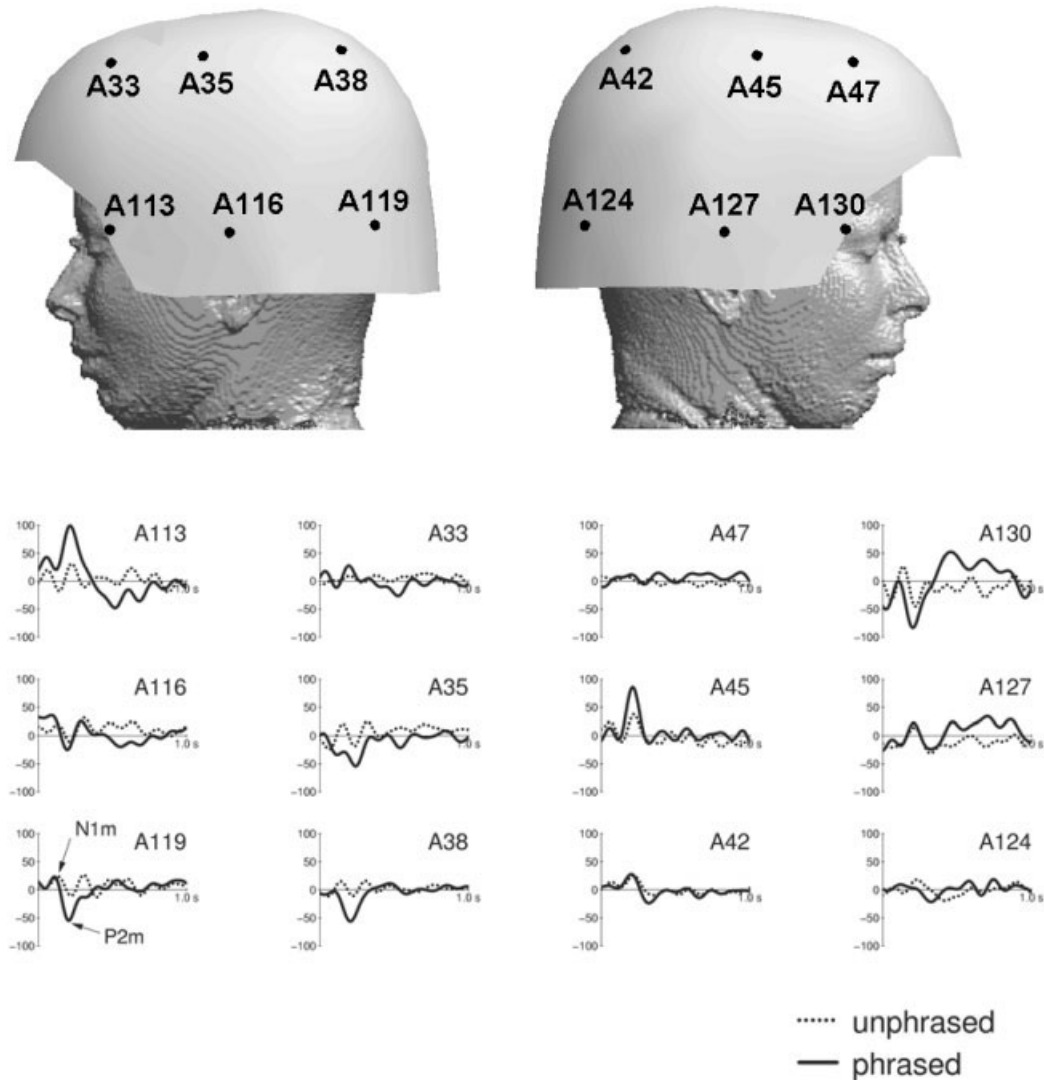EEG, electroencephalography; MEG, magnetoencephalography.

effect. The maps in Figure 7 reveal the topographies of these components. The MEG shows an anterior positivity. (Note that, for MEG, we define positivity as an outward directed field vector.) over the right and a similar negativity over the left hemisphere. Note that bilateral MEG patterns typically show polarity reversals from left to right. In the ANOVA analysis, this leads to interactions between COND and RIGHT–LEFT, rather then to main effects (see Table II). On the other hand, true hemispheric effects would be reflected

500 – 600 ms

−4μV   4μV

Figure 5.
Spline interpolated maps of the relevant electroencephalography components. The maps represent difference activity between the phrased and unphrased conditions, integrated over time windows. They are displayed in three dimensions and viewed from different perspectives.

Figure 6.
Traces of selected magnetoencephalography channels. Solid lines denote the original condition with unmodified phrase boundaries (*phrased*). Dotted lines denote the modified condition, where the phrase boundary was filled with notes (*unphrased*). Units are femtoteslas. A low-pass filter of 8 Hz was applied for display purposes only.

by main effects and interactions without the factor RIGHT–LEFT. Table II demonstrates that there are only MEG effects that do include the RIGHT–LEFT factor; hence, no hemispheric asymmetry can be concluded.

The results of the source localization are summarized in Figures 8 and 9. While looking at these figures, one should bear in mind that the marked areas do not form a proven network subserving the processing of phrase boundaries. Instead, they mark candidate brain regions that could have contributed to the generation of the observed EEG and MEG components. For the EEG analysis, these candidate regions include the posterior cingulate cortex, the retrosplenial cortex, and the posterior part of the left hippocampus. For MEG, it is also likely that the right

posterior hippocampus, the anterior cingulate, and the subcallosal region are involved in the generation of the signal. The sites of subcortical structures like basal ganglia or the diencephalon cannot be excluded from a mathematical point of view, but due to their small size and deep location, any significant contribution to extracranial observations is assumed to be very unlikely.

DISCUSSION

The major aim of this study was to find an electrophysiological correlate for the processing of phrase boundaries in music. We wanted to know if such a correlate resembles the
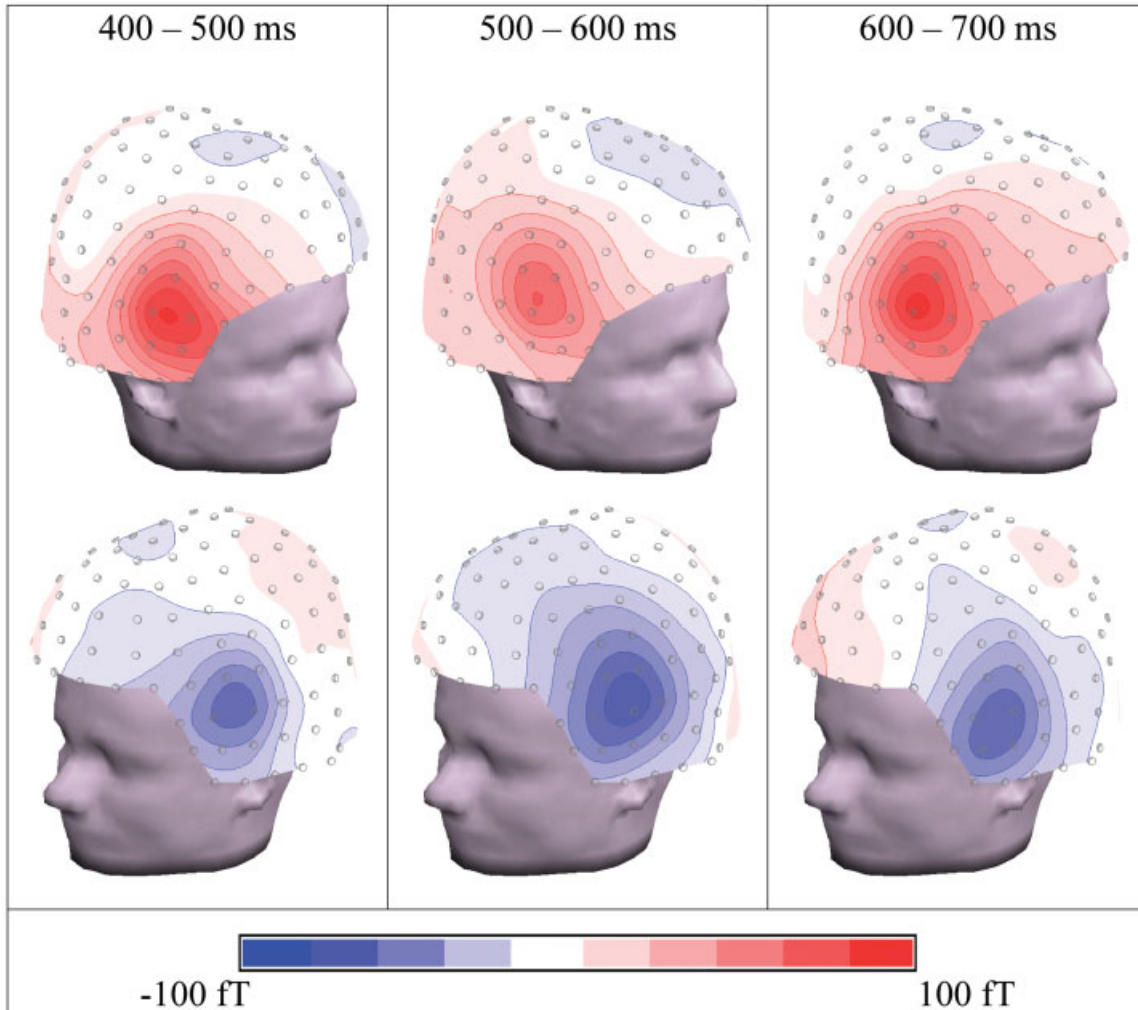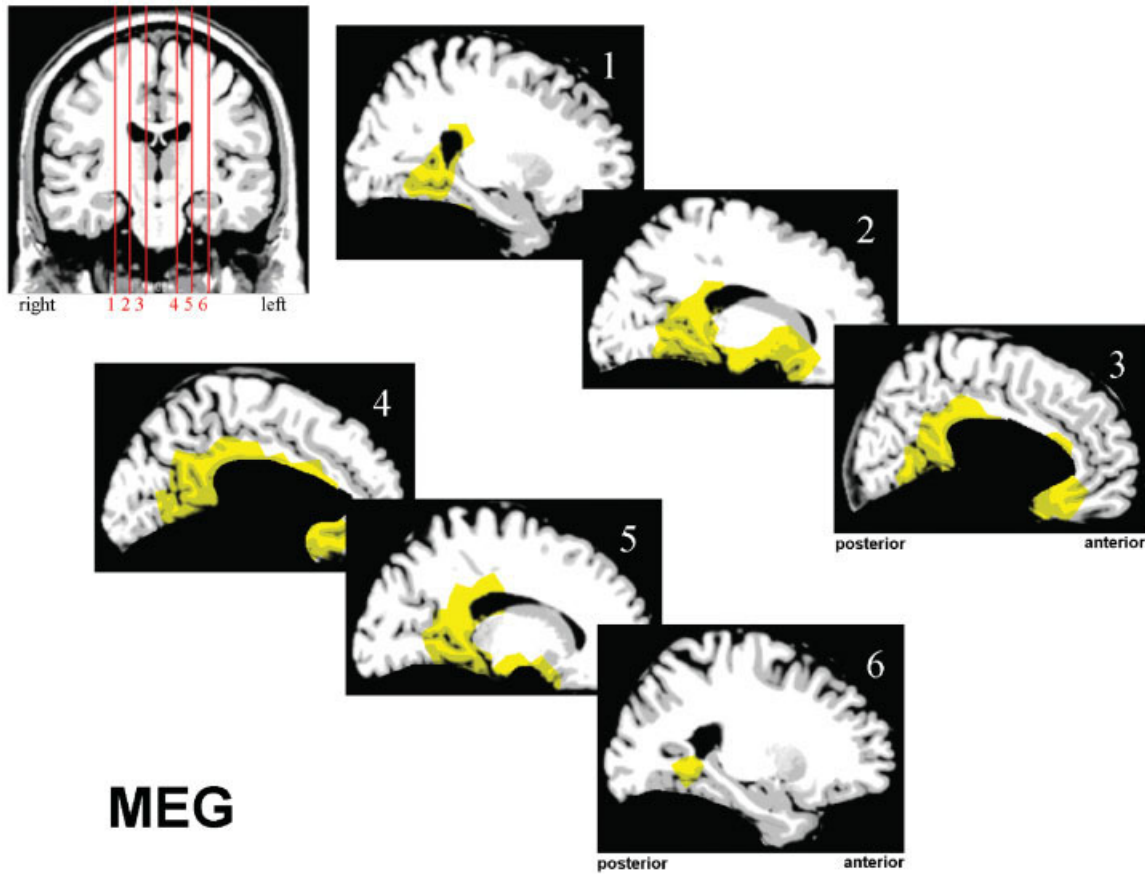
Figure 7.
Spline interpolated maps of the relevant magnetoencephalography (MEG) components. The maps represent difference activity between the phrased and unphrased conditions, integrated over time windows. They are displayed in three dimensions and viewed at from different perspectives.

CPS found by Steinhauer and colleagues [1999] for spoken sentences. Moreover, we aimed at the extraction of information on the network of the brain processes that underlie phrase boundary processing. Indeed, we found statistically significant components correlated to the presence of a musical phrase boundary in both EEG and MEG. In the EEG data, there was a positive variation at central and parietal electrodes peaking at approximately 550 ms after the offset of the phrase boundary. In MEG, we found significant effects mainly in the frontolateral parts of the sensor array, just before (400–500 ms) and just after (600–700 ms) the effect observed in the ERP data. None of the effects was significantly lateralized. Please note that the topography of both the EEG and MEG effects (see Figs. 4 and 6) are clearly distinct from the ones of the respective primary components (P2 and P2m). Hence, they cannot be just the primary response to one of the following notes.

The amplitude and topographical distribution of the found ERP component bear some similarity to the language CPS of Steinhauer and colleagues [1999], who reported an amplitude difference between the conditions of approximately 3 to 4 μV and a nonlateralized distribution with centroparietal maximum. Latency and duration are not so easily matched. The language CPS seems to start directly at the phrase boundary and lasts between approximately 500 and 1,000 ms [Steinhauer et al., 1999; Fig. 2]. Pannekamp and colleagues [2004] reported data from different kinds of speech material, where the CPS became significant between 150 and 700 ms after the onset of the first intonational phrase boundary, exhibiting a duration of approximately 500 ms and an amplitude of several microvolts. The difference in latency between the CPS in language and music may be due to the fact that, in spoken language, the phrase boundary is signaled by parameter changes, which are already present

Figure 8.

Results of the source localization based on magnetoencephalography (MEG) data. The yellow areas show a correlation between the local source space and the signal subspace of less than 7 degrees. Only cerebral areas were considered.

before the onset of the pause. The influence of these early cues might be greater for phrase boundaries in speech compared to music. Such parameter changes are, e.g., lengthened prefinal syllables and changed F0 contours.

Based on the similarities in paradigm, amplitude and topology, the positivity in the EEG associated to the phrase boundaries can be interpreted as the music equivalent of the CPS in speech perception, in the following termed "music CPS".

It should be mentioned that alternative interpretations could be thought of also. For example, the level of expectancy within the analyzed time interval might be greater in the unphrased condition (where we are in the middle of a phrase) than in the phrased one (where a new phrase has just begun). Instead of a positivity for the phrased condition, one could assume a negativity for the unphrased one, reflecting higher expectancy levels. However, such an interpretation would be incompatible with the fact that the difference between the conditions is not present in the entire

analyzed time interval as well as with the MEG observations. However, it is possible that an expectancy-related component contributed to the observed ERP effects.

The question remains, why the music CPS failed to appear in the pilot experiment. We have discussed that the impression of "phrasedness" of the stimuli might have been too low due to several acoustic traits of the melodies (see Subjects and Methods section). Differences between the pilot and main experiments included local acoustic cues in the immediate vicinity of the phrase boundary (e.g., smaller prefinal slowing and shorter pauses in the pilot study, see Table I). These cues should facilitate the detection of the phrase boundary [Riemann, 1900]. Additionally, there were differences concerning the melodic structure as a whole. The stimuli of the pilot experiment were more complicated and less symmetric, and their tonality was often ambiguous. This increased difficulty might have hampered the proper perception of the phrase structure. A post hoc rating test with eight musicians confirmed that the phrase structure of the
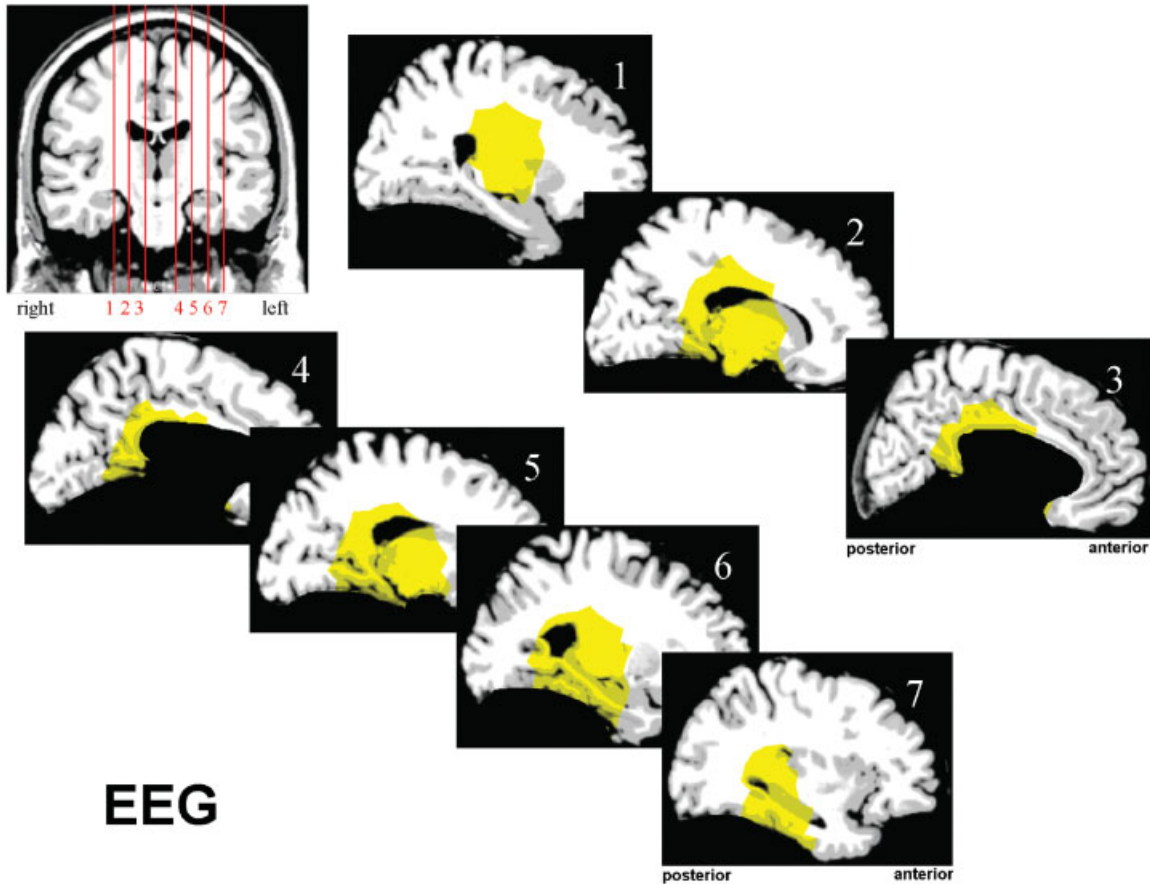
Figure 9.
Results of the source localization based on electroencephalography (EEG) data. The yellow areas show a correlation between the local source space and the signal subspace of less than 7 degrees. Only cerebral areas were considered.

stimuli of the pilot experiment was less clearly perceived. It showed that in 42.8% of the stimuli of the pilot experiment the phrase structure was not recognized correctly, compared to only 21.8% for the main study stimuli. However, it is not clear which of these structural traits of the musical pieces exercise the most decisive influence onto the recognition of phrase boundaries. Further studies are needed to clarify these issues. Another potential problem in the pilot experiment, which has been avoided in the main study, might have been formed by the multiple presentation of the same stimulus. Increasing familiarity with the stimuli might have influenced the normal processing of the music and its phrase structure.

The present study was carried out with trained musicians. This was to make sure that the subjects' degree of implicit musical expertise was comparable to the language expertise of the subjects of Steinhauer and colleagues. On the other hand, musicians usually also have some explicit theoretical understanding on music, whereas comparable knowledge on language was not present in the subjects of the speech studies (who were not linguists after all).

It is known that many aspects of music are processed differently in musicians and nonmusicians. Such differences have been demonstrated for the preattentive processing of auditory stimuli [e.g., Brattico et al., 2001; Koelsch et al., 1999; Pantev et al., 2003; van Zuijen et al., 2004], the processing of temporal patterns [Drake et al., 2000; Jongsma et al., 2004], pitch memory [Gaab and Schlaug, 2003], as well as the processing of harmonic and melodic information [Overman et al., 2003; Schmithorst and Holland, 2003]. On the other hand, music can be enjoyed and understood also by nonmusicians to a considerable degree, as demonstrated by, e.g., Koelsch et al. [2000, 2003]. It is, therefore, an important question, to what extent and under what conditions the music CPS could be observed in a musically untrained group of subjects. This answer would shed light onto the respective roles of inherent capabilities and acquired rules for the perception of musical phrases and should be a subject of future studies.

Equivalence Between Electrophysiological Markers for Phrase Boundaries in Language and Music

The relationship between the two domains and its possible implications on the neural substrate of the observed effects are of special interest here. The speech-like character of music as well as its syntactic and semantic aspects has often been a matter of debate in musicology. Nevertheless, drawing parallels between the two domains should be handled with great care. At first sight, several similarities can be found. Elements of speech and musical sound can be characterized by the same parameters: pitch, duration, loudness, and rhythmic–metrical structure as well as contour, articulation, and timbre. Furthermore, acoustical events in both systems appear as streams of sound, where the principles of grouping and segmentation apply. Speech and musical utterances (in this context, melodies) are based on specific grammatical and phonetic/phonological rules and, therefore, can be described as sequences of structured events along the time axis. Thus, syntactic and prosodic relationships between both domains can be assumed [see for example Riemann's "Funktionstheorie," or Lerdahl and Jackendoff, 1983]. Neurophysiological proof has been provided recently by the result that Broca's area, known for its specialization to syntactic processing in language, was activated by music perception as well [Koelsch et al., 2002; Maess et al., 2001] and by the elicitation of statistically indistinguishable P600 waves caused by structural incongruities in both language and music [Besson and Macar, 1987, Besson et al., 1994; Besson and Schön, 2003; Patel et al., 1998] as well as an early right anterior negativity for structural incongruities in music [Koelsch et al., 2000] similar to the early left anterior negativity for structural incongruities in language [Friederici et al., 1993; Hahne and Friederici, 1999]. In the context of the present study, analogies between the domains of *music* and *language* on a prosodic/syntactic level can be drawn. Since the CPS in language has been shown to rely on purely prosodic information [e.g., Pannekamp et al., 2004], such a level of similarity is in line with the idea of a close relationship between the language and music CPS.

A similar brain response of speech and music boundary processing also sheds new light on general segmentation processes. In speech, a prosodic boundary is often marked by specific shifts in fundamental frequency and durational markers such as prefinal lengthening and pause insertion [Beach, 1991; Selkirk, 1994]. Of interest, similar parameters can be found for musical boundary markers.

However, the majority of psycholinguistic research on the processing of prosodic boundary information in speech examined how prosody helps to resolve syntactic ambiguity resolution [Marslen-Wilson et al., 1992; Warren et al., 1995]. Only a few studies investigated the function of prosodic boundaries during speech segmentation [Cooper and Paccia-Cooper, 1980; Marcus and Hindle, 1990; Streeter, 1978].

Underlying Neural Network

Source localization based on a signal subspace projection method revealed several limbic structures that are normally involved in memory and attention. It should be born in mind that the applied method cannot uniquely identify a network of active brain areas that underlie the observed electric potentials and magnetic fields. It is indeed very doubtful if any method can reliably do this for the present data. However, what we can do is to decide for every region of the brain separately, whether it could have contributed to the observed data. This way, we create a representation of a class of possible solutions. Such a measure is provided by the signal subspace projection incorporated in the MUSIC method [Mosher et al., 1992]. What is the expected neural response associated to the processing of a phrase boundary? No matter whether we are in the music or in the language domain, the phrase boundary has to be detected using local cues (e.g. pauses, boundary tones, final lengthening) as well as global expectations due to, e.g., harmonic and rhythmic structure of the musical piece or the grammar and prosody of the speech, respectively. Additionally, higher level processes have to take place, in particular those involving memory (e.g., the closed phrase has to be stored in memory as a unified entity—chunking) and attention (e.g., the attention focus has to be directed toward the new phrase).

First, the processing of the local cues, possibly mediated by more global expectations, presumably takes place very quickly and might be reflected by the N100/P200 offset/onset reactions at the end of the previous and the beginning of the new phrase. Source localizations have shown that these components probably originate in the anterior planum temporale [Knösche et al., 2003]. The planum temporale (PT) contains the auditory association cortex, which is known to play a role in cognitive processing of speech and nonspeech [Binder et al., 1998; Griffith and Warren, 2002; Jäncke et al., 2002]. Activation of the PT has also been reported in several studies on music perception [Koelsch et al., 2002; Liegois-Chauvel et al., 1998; Menon et al., 2002; Ohnishi et al., 2001]. Hence, the increased P200 after the phrase boundary might not only reflect simple recovery of neuronal populations after the pause but also some higher order feature extraction necessary for the recognition of a boundary between two phrases. However, there is no way to finally prove this from our present data.

Second, the integrative processing of the phrase boundary appears to be reflected by effects observed beyond 400 ms. The significant ERP and MEG deflections between 400 and 700 ms after the onset of the second phrase are compatible with generators in the posterior and anterior cingulate cortex (PCC, ACC), the subcallosal/medial orbitofrontal area (OFC), the retrosplenial cortex (RC), and the posterior parahippocampal areas (PHC). Recent functional MRI (fMRI) studies investigating the neural network supporting prosodic processing indicate an involvement of the posterior and middle portion of the cingulate gyrus [Meyer et al., 2004]. Activation in OFC, RC, and PHC were not observed, although the thalamus appeared to be involved. Although

there are several subcortical areas (including thalamus) that cannot be excluded from a modeling point of view, their significant contribution to the surface measurements seems to be highly unlikely, so that they are not considered here.

The EEG and MEG time courses show at least three different time windows with different extracranially measurable activity with respect to condition, roughly between 400 and 500 ms, between 500 and 600 ms, and between 600 and 700 ms. Of interest, the first and the last time window show a significant effect only in MEG, with very similar topologies in both periods, whereas the middle window between 500 and 600 ms exhibits a significant difference in EEG, but only reduced activity for MEG. (Please note, that within each of these time periods, the EEG and MEG are not completely stationary.) Hence, although all three periods yield roughly the same collection of possible contributing regions, their relative weightings in the actual networks must be different. It is well known that the MEG method is relatively insensitive toward currents, which are oriented radially. (In an ideally spherical head, MEG would actually be blind to such sources. However, in real situations this is not entirely the case.) On the other hand, EEG is less sensitive to tangentially than to radially oriented generators. Hence, different relative activations in the identified candidate regions can account for topology differences. For example, simulations showed that the PCC does account well for the EEG at the peak latency around 550 ms, but produces only very small MEG. Hence, in the middle period (500–600 ms), the PCC is probably the most active region of the brain. In general, it was not possible to disentangle the precise temporal evolvement of the activities in a reliable way.

It is striking that most of the candidate areas belong to the limbic system and are known to be involved into attention and memory processes. This finding deserves a more detailed discussion. The ACC roughly includes Brodmann areas 24 and 32. Activations are usually found in a broad variety of tasks, including Stroop paradigms, working and episodic memory, semantic generation and processing, direction of attention toward noxious stimuli, processing of emotional stimuli [for a review, see Phan et al., 2002], response and perceptual conflict, motor preparation, and self-monitoring. For a review on a large number of positron emission tomography and fMRI studies, see Cabeza and Nyberg [2000]. Many of the observations are consistent with the hypothesis that ACC plays a role in the direction of attention toward a stimulus or an intended action as well as in the suppression of inappropriate responses. In the present context, it seems that ACC activation could reflect the redirection of attention from the previous phrase toward the new one. Activations of the PCC (Brodmann areas 23 and 31) are consistently found during episodic memory retrieval [see review of Cabeza and Nyberg, 2000] as well as during learning and problem solving tasks [Cazalis et al., 2003; Reber et al., 2003; Ruff et al., 2003; Werheid et al., 2003]. Also mediotemporal (parahippocampal) structures are mainly involved in episodic memory encoding and nonverbal episodic memory retrieval [see review of Cabeza and Nyberg,

2000]. Such memory operations are necessary when stored melody patterns are retrieved and when the elapsed phrase is stored as a unified entity (chunk).

Hence, it seems that the CPS, at least in the case of music, does not directly reflect the detection of the phrase boundary as such but rather memory- and attention-related processes, which are necessary for the transition from one phrase to the next. This process is in line with the finding that, in the language domain, the CPS is not influenced by the presence or absence of certain cues, such as a pause at the prosodic boundary [Steinhauer et al., 1999, experiment 3]. Moreover, it is in agreement with several studies demonstrating that the language CPS is influenced by the focus of attention [Hruska and Alter, 2004; Toepel and Alter, 2003] and the syntactic content of the stimulus material [Pannekamp et al., 2004].

Conclusion

We have identified EEG and MEG markers for the processing of phrase boundaries in music. The EEG marker bears some similarity to the CPS found for the processing of prosodic phrase boundaries in language. Moreover, we used source-modeling techniques to identify a collection of possible generators for the observed waveforms. The identified structures are parts of the limbic system that are known to be involved in memory and attention processes. We conclude that (1) the marker component found is to some extent equivalent to the CPS in language and (2) the CPS seems to reflect memory and attention processes necessary for the transition from one phrase to the next, rather than the mere detection of the explicit acoustic boundary. The identification of the possible neural networks for the language CPS would be an important contribution to further testing this hypothesis. The degree to which the observed effect depends on musical aptitude of the subject and on structural traits of the stimulus material remains to be investigated. In particular, the question of musical expertise could be crucial for the understanding of the significance of the music CPS and its relationship to the language CPS. Musicians may be comparable to the subjects of the language studies in their implicit knowledge in the respective domain, but they certainly have a much higher degree of explicit theoretical knowledge. These issues should be tackled in future studies.

REFERENCES

Beach CM (1991): The interpretation of prosodic patterns at points of syntactic structure ambiguity: evidence for cue trading relations. J Memory Lang 30:644–663.

Besson M, Macar F (1987): An event-related potential analysis of incongruity in music and language. Psychophysiology 24:14–25

Besson M, Schön D (2003): Comparison between language and music. In: Peretz I, Zatorre R, editors. Cognitive neuroscience of music. Cambridge: Oxford University Press.

Besson M, Faita F, Requin J (1994): Brain waves with musical incongruity differ for musicians and non-musicians. Neurosci Lett 168:101–105.

Binder JR, Zilbovicius M, Crozier S, Thivard L, Fontaine A, Masure M, Samson Y (1998): Lateralization of speech and auditory temporal processing. J Cogn Neurosci 10:536–540.

Brattico E, Näätänen R, Tervaniemi M (2001): Context effects on pitch perception in musicians and non-musicians: evidence from event-related potential recordings. Music Percept 19:199–222.

Cabeza R, Nyberg L (2000): Imaging cognition II: an empirical review of 275 PET and fMRI studies. J Cogn Neurosci 12:1–4.

Cazalis F, Valabregue R, Pelegrini-Isaac M, Asloun S, Robbins TW, Granon S (2003): Individual differences in prefrontal cortical activation on the Tower of London planning task: implication for effortful processing. Eur J Neurosci 17:2219–2225.

Chiappe P, Schmuckler MA (1997): Phrasing influences the recognition of melodies. Psychon B Rev 4:254–259.

Cooper WE, Paccia-Cooper J (1980): Syntax and speech. Cambridge, MA: Harvard University Press.

Cuddy LL, Cohen AJ, Mewhort DJK (1981): Perception of structure in short melodic sequences. J Exp Psychol Human 7:869–883.

Cutler A, Dahan D, van Donselaar W (1997): Prosody in the comprehension of spoken language: a literature review. Lang Speech 40:141–201.

Dowling WJ (1973): Rhythmic groups and subjective chunks in memory for melodies. Percept Psychophys 14:37–40.

Drake C, Penel A, Bigand E (2000): Tapping in time with mechanically and expressively performed music. Music Percept 18:1–23.

Friederici AD, Pfeifer E, Hahne A (1993): Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. Cogn Brain Res 1:183–192.

Gaab N, Schlaug G (2003): The effect of musicianship on pitch memory in performance matched groups. Neuroreport 14:2291–2295

Griffith TD, Warren JD (2002): The planum temporale as computational hub. Trends Neurosci 25:348–353.

Hahne A, Friederici AD (1999): Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. J Cogn Neurosci 11:194–205.

Hämäläinen MS, Sarvas J (1987): Feasibility of the homogeneous head model in the interpretation of neuromagnetic data. Phys Med Biol 32:91–97.

Haueisen J, Knösche TR (2001): Involuntary motor activity in pianists evoked by music perception. J Cogn Neurosci 13:786–792.

Hruska C, Alter K (2004): How prosody can influence sentence perception. In: Steube A, editor. Information structure: theoretical and empirical evidence. Berlin: Walter de Gruyter.

Jäncke L, Wusenberg T, Scheich H, Heinze H (2002): Phonetic perception and the temporal cortex. Neuroimage 15:733–746.

Janke W, Debus G (1978): Die Eigenschaftswörterliste EWL. Göttingen: Hogrefe.

Jongsma MLA, Desain P, Honing H (2004): Rhythmic context influences the auditory evoked potentials of musicians and nonmusicians. Biol Psychol 66:129–152.

Keller H (1955): Phrasierung und Artikulation: Ein Beitrag zu einer Sprachlehre der Musik. Kassel: Bärenreiter.

Knösche TR (2002): Transformation of whole-head MEG recordings between different sensor positions, Biomed Tech 47:59–62.

Knösche TR, Neuhaus C, Haueisen J, Alter K (2003): The role of the planum temporale in the perception of musical phrases. Proceedings of 4th International Conference on Noninvasive Functional Source Imaging (NFSI), Chieti, Italy, September 2003.

Koelsch S, Schröger E, Tervaniemi M (1999): Superior pre-attentive auditory processing in musicians. Neuroreport 10:1309–1313.

Koelsch S, Gunter TC, Friederici AD, Schröger E (2000): Brain indices of music processing: non-musicians are musical. J Cogn Neurosci 12:520–541.

Koelsch S, Gunter TC, von Cramon DY, Zysset S, Lohmann G, Friederici AD (2002): Bach speaks: a cortical 'language-network' serves the processing of music. Neuroimage 17:956–966.

Koelsch S, Gunter TC, Schröger E, Friederici AD (2003): Processing tonal modulations—an ERP study. J Cogn Neurosci 15:1149–1159.

Lerdahl F, Jackendoff R (1983): A generative theory of tonal music. Cambridge: MIT Press.

Liegeois-Chauvel C, Peretz I, Babai M, Laguitton V, Chauvel P (1998): Contribution of different cortical areas in the temporal lobes to music processing. Brain 121:1853–1867.

Maess B, Koelsch S, Gunter TC, Friederici AD (2001): Musical syntax in processed in Broca's area: and MEG study. Nat Neurosci 4:540–545.

Marcus M, Hindle D (1990): Psycholinguistic and computational perspectives. In: Altmann GTM, editor. Cognitive models of speech processing. Cambridge, MA: MIT Press. p 483–512.

Marslen-Wilson WD, Tyler LK, Warren P, Grenier P, Lee CS (1992): Prosodic effects in minimal attachment. Q J Exp Psychol 45A: 73–87.

Mattheson J (1976): Kern melodischer Wissenschaft [reprint of Hamburg edition of 1737]. Hildesheim: Olms

Menon V, Levitin DJ, Smith BK, Lembke A, Krasnow BD, Glazer D, Glover GH, McAdams S (2002): Neural correlates of timbre change in harmonic sounds. Neuroimage 17:1742–1754.

Meyer M, Steinhauer K, Alter K, Friederici AD, von Cramon DY (2004): Brain activity varies with modulation of dynamic pitch variance in sentence melody. Brain Lang 89:277–289.

Mosher JC, Lewis PS, Leahy RM (1992): Multiple dipole modelling and localization from spatio-temporal MEG data. IEEE Trans Biomed Eng 39:541–557.

Ohnishi T, Matsuda H, Asada T, Aruga M, Hirakata M, Nishikawa M, Katoh A, Imabayashi E (2001): Functional anatomy of musical perception in musicians. Cereb Cortex 11:754–760.

Oldfield RC (1971): The assessment and analysis of handedness. Neuropsychologia 9:97–113.

Oostenveld R, Praamstra P (2001): The five percent electrode system for high-resolution EEG and ERP measurements. Clin Neurophysiol 112:713–719.

Overman AA, Hoge J, Dale JA, Cross JD, Chien A (2003): EEG alpha desynchronization in musicians and nonmusicians in response to changes in melody, tempo, and key in classical music. Percept Motor Skills 97:519–532.

Pannekamp A, Toepel U, Alter K, Hahne A, Friederici AD (2004): Prosody driven sentence processing: an ERP study. J Cogn Neurosci (in press).

Pantev C, Roberts LE, Schulz M, Engelien A, Ross B (2001) : Timbre-specific enhancement of auditory cortical representations in musicians. Neuroreport 12:169–174.

Pantev C, Ross B, Fujioka T, Trainor LJ, Schulte M, Schulz M (2003): Music and learning-induced cortical plasticity. Ann N Y Acad Sci 999:438–450.

Patel AD (2003): Language, music, syntax and the brain. Nat Neurosci 6:674–681.

Patel AD, Balaban E (2000): Temporal patterns of human cortical activity reflect tone sequence structure. Nature 404:80–84.

Patel AD, Balaban E (2001): Human pitch perception is reflected in the timing of stimulus-related cortical activity. Nat Neurosci 4:839–844.

Patel AD, Gibson E, Ratner J, Besson M, Holcomb P (1998): Processing syntactic relations in language and music: an event-related potential study. J Cogn Neurosci 10:717–733.

Paulus W (1988): Effect of musical modelling on late auditory evoked potentials. Eur Arch Psychiatry Neurol Sci 237:307–311.

Phan KL, Wagner T, Taylor SF, Liberzon I (2002): Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. Neuroimage 16:331–348.

Reber OJ, Gitelman DR, Parrish TB, Mesulam MM (2003): Dissociating explicit and implicit category knowledge with fMRI. J Cogn Neurosci 15:574–583.

Riemann H (1900): Vademecum der Phrasierung. Leipzig: Max-Hesse-Verlag.

Ruff CC, Knauff M, Fangmeier T, Spreer J (2003): Reasoning and working memory: common and distinct neuronal processes. Neuropsychologia 41:1241–1252.

Schmithorst VJ, Holland SK (2003): The effect of musical training on musical processing: a functional magnetic resonance imaging study in humans. Neurosci Lett 348:65–68.

Schulz M, Ross B, Pantev C (2003): Evidence for training-induced crossmodal reorganization of cortical functions in trumpet players. Neuroreport 14:157–161.

Selkirk E (1994): Sentence prosody: intonation, stress, and phrasing. In: Goldsmith J, editor. Handbook of phonological theory. New York: Blackwell. p 550–569.

Selkirk E (2000): The interaction of constraints on prosodic phrasing. In: Horne M, editor. Prosody: theory and experiment. Amsterdam: Kluwer.

Steinhauer K (2003): Electrophysiological correlates of prosody and punctuation. Brain Lang 89:277–289.

Steinhauer K, Friederici AD (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 AD (1999): Brain potentials indicate immediate use of prosodic cues in natural speech processing. Nat Neurosci 2:191–196.

Stirling L, Wales R (1996): Does prosody support or direct sentence processing? Lang Cogn Process 11:193–212.

Stoffer TH (1985): Representation of phrase structure in the perception of music. Music Percept 3:191–220.

Streeter LA (1978): Acoustic determinants of phrase boundary perception. J Acoust Soc Am 64:1582–1592.

Tan N, Aiello R, Bever TG (1981): Harmonic structure as a determinant of melodic organization. Mem Cognit 9:533–539.

Toepel U, Alter K (2003): How mis-specified accents can distress our brain. Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain, August 2003, p 619–622.

Warren P, Grabe E, Nolan F (1995): Prosody, phonology, and parsing in closure ambiguities. Lang Cogn Process 10:457–486.

Werheid K, Zysset S, Muller A, Reuter M, von Cramon DY (2003): Rule learning in a serial reaction time task: an fMRI study on patients with early Parkinson's disease. Brain Res Cogn Brain Res 16:273–284.

Wilson SJ, Pressing J, Wales RJ, Pattinson P (1999): Cognitive models of music psychology and the lateralization of musical function within the brain. Aust J Psychol 51:125–139.

van Zuijen TL, Sussman E, Winkler I, Näätänen R, Tervaniemi M (2004): Grouping of sequential sounds—an event-related potential study comparing musicians and non-musicians. J Cogn Neurosci 16:331–338.

von Zerrsen D (1976): Klinische Selbstbeurteilungsskalen. Allgemeiner Teil. Weinheim: Beltz.