ORIGINAL ARTICLE

Decomposing rhythm processing: electroencephalography of perceived and self-imposed rhythmic patterns

Rebecca S. Schaefer · Rutger J. Vlek · Peter Desain

Abstract Perceiving musical rhythms can be considered a process of attentional chunking over time, driven by accent patterns. A rhythmic structure can also be generated internally, by placing a subjective accent pattern on an isochronous stimulus train. Here, we investigate the event-related potential (ERP) signature of actual and subjective accents, thus disentangling low-level perceptual processes from the cognitive aspects of rhythm processing. The results show differences between accented and unaccented events, but also show that different types of unaccented events can be distinguished, revealing additional structure within the rhythmic pattern. This structure is further investigated by decomposing the ERP into subcomponents, using principal component analysis. In this way, the processes that are common for perceiving a pattern and self-generating it are isolated, and can be visualized for the tasks separately. The results suggest that top-down processes have a substantial role in the cerebral mechanisms of rhythm processing, independent of an externally presented stimulus.

Introduction

Many of the auditory patterns we perceive around us, such as speech and music, require the structuring of information over time for efficient perception. Perceiving regularities is

R. S. Schaefer (✉) · R. J. Vlek · P. Desain
Donders Institute for Brain, Cognition and Behaviour,
Centre for Cognition, Radboud University,
Montessorilaan 3, 6525 HE Nijmegen, The Netherlands
e-mail: r.schaefer@donders.ru.nl

essential for interpretation of this information, and leads to predictive processing, which, in turn, is needed for goal-directed behavior (for a recent overview, see Winkler, Denham, & Nelken, 2009). In music listening, it is widely believed that the confirmation and violation of expectations are crucial for a musical piece's own characteristics and what makes it specific and enjoyable (Huron, 2006).

Events in auditory patterns such as musical rhythms are believed to be processed more efficiently when their position in time can be predicted, described in dynamic attending theory (Drake, Jones, & Baruch, 2000; Jones & Boltz, 1989) as well as other theories of fluctuating expectation levels (Desain, 1992). The main premise is that expectancy levels can be manipulated by presenting temporal patterns varying in regularity, thus making impending events more or less predictable. This is also reflected in computational models of rhythm processing, such as the coupled oscillator model presented by Large and colleagues (Large & Jones, 1999; Large & Kolen, 1995), in which the percept of a rhythm is built up out of multiple oscillators with different period lengths, related to different hierarchical levels of the rhythm. These models include a feedback loop in which highly expected events raise the 'confidence' of the oscillator, contributing more to subsequent expectancy. Consequently, these models can predict how in a very simple train of stimuli (e.g. an isochronous rhythm) chunking of a number of events may occur, so that specific future events incur a higher expectancy. This continues even if the actual accent is no longer present in the stimulus, making the percept of a pulse in the event train quite robust (see Fig. 1 for a graphical representation of this process). In the concept of dynamic attending as proposed by Jones and Boltz (1989) and Drake et al. (2000), listeners willfully give more weight to the oscillator we choose, thus attending to different hierarchical levels of the rhythmic structure (i.e. the beat, bar or
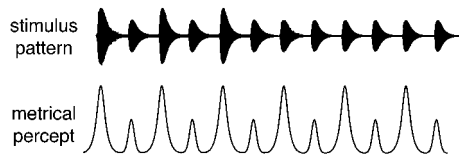
Fig. 1 A schematic representation of a metrical percept is shown to continue its pattern on an isochronous stimulus sequence by first introducing the structure. The perceived pulse will persist, resulting in a purely subjective structure. Based on Snyder and Large (2005)

even phrase level). This adds an internally driven factor to the mechanism, which leaves each individual event with its own unique combination of attention levels for each phase of the coupled oscillators. This concept has also been described in terms of music theoretical considerations (London, 2004), referring to hierarchical levels of metric patterns as cycles. Here, we report the event-related potential (ERP) response to these different events within rhythmic patterns, assuming that varying levels of attention and expectancy will be visible in the ERP of the electro-encephalogram (EEG, for early work demonstrating the effect of attention on the ERP, see Näätänen, 1975; Hillyard, Hink, Schwent, & Picton, 1973). In order to distinguish between the perceptual responses and cognitive mechanisms that are independent of external stimulation, we do this for both externally presented and internally generated patterns.

We commonly think of rhythmic structure as hierarchical (see, for instance, Lerdahl & Jackendoff, 1983; Longuet-Higgins & Lee, 1984), an assumption that has also been supported by showing that brain responses to deviants in different metrical positions resulted in different ERP signatures. This has been shown for the P300 oddball response using intensity decrements on different positions in an isochronous stimulus pattern (Brochard, Abecasis, Potter, Ragot, & Drake, 2003), and for the mismatch negativity (MMN) response to syncopations in different positions (Ladinig, Honing, Háden, & Winkler, 2009). In these studies, it was shown that deviants in strong metric positions result in larger P300 and MMN components, respectively, suggesting enhanced processing of accented events. However, the question of how this processing hierarchy is built up is not easily answered. Two contrasting hypotheses can be formulated. On the one hand the brain signature for each event in a cycle may be unique, as in a Gestalt, and on the other hand, the response may be predictable and built up of low-level components, for instance, due to its own combination of the positions of multiple coupled oscillators. These hypotheses may be tested by decomposing the response to see if any commonalities are found over the different events. Although there are multiple methods of decomposing EEG data, the method most commonly used for ERPs is principal component analysis (see, for instance,

Dien & Frishkoff, 2005 for an overview). This will yield statistically independent components with weight distributions over the different sensors, that combine to form the full signature, and each explain an amount of the variance in the data. If we assume that the EEG traces of the different subprocesses combine linearly in the total signal, we can compare the decomposed EEG response to our own notion of hierarchical processing in rhythm perception.

In the current study, we use three rhythmic patterns: binary, ternary and quaternary groupings, referred to as 2 beat, 3 beat and 4 beat. These patterns roughly correspond to 2/4, 3/4, and 4/4 meters, and consist of cycles of an accented or louder first event called the downbeat, followed by one, two or three unaccented events that are considered to have a weaker metrical function. These groupings are shown to be easiest to synchronize with in terms of numerosity (Repp, 2007). Within these patterns, we defined types of events and pooled together the responses to compare them. First, we compare all accented events to all unaccented events, to find the effect of the downbeat, or accented event. However, the literature on rhythm processing generally posits a more complex structure with more than two types of events (i.e. accented/unaccented, see for instance, Lerdahl & Jackendoff, 1983). Thus, we look for evidence in the brain activity of the processing of a more intricate structure. We postulate that in the different patterns, certain events have something in common, namely the first unaccented event that follows the downbeat (or accented) event, as well as the last unaccented event, also called the upbeat, leading to, perhaps anticipating, the upcoming downbeat. As we are trying to uncover different processes occurring simultaneously, we try to decompose the EEG data to see if we can find a brain signature that is specific to such subprocesses.

To investigate rhythm processing independent of perceptual input, we make use of subjective accenting: patterns that are self-imposed on ambiguous, unaccented stimuli. A common manifestation of subjective accenting is the so-called clock illusion, when a regularly sounding 'tick-tick-tick-tick...' may spontaneously induce a 'tick-tock-tick-tock...'-percept in which events are chunked into groups of two. The binary grouping arises spontaneously, and the first beat of every group is perceived as distinctively different from the second. Spontaneous subjective rhythmization has been a topic of study for some time, beginning with the influential work of Bolton (1894). In an early psychology text, it is described as a mechanism inherent to our sense of time, similar to grouping mechanisms inherent to visual perception (Boring, 1942). As opposed to this spontaneous process, we here investigate effortful subjective accenting, by inducing a specific pattern as it is represented in Fig. 1. By investigating rhythm processing based on external input, where the

pattern is present in the stimulus, as well as generation of a rhythmic pattern in the absence of any accent in the stimulus, we can find processing mechanisms that take place independently of physical accenting patterns. Though there will be many shared top-down processes active in both tasks, we refer to the instructional phase as the 'perception'-task, and the latter as the 'imagery'-task. By performing the PCA decomposition over the averaged response to both tasks, brain activity patterns that are common to the two tasks may be isolated and interpreted. In this way, the risk of an effect of the instructional effect adding to the perception task is minimized.

This leads us to a number of hypotheses of rhythm perception, the first and most important one being that the difference between an accented and an unaccented event is detectable in the brain signal. Secondly, we pose that not all unaccented events are equal, more specifically, we hypothesize that events with a similar function in the pattern will show similarities in the EEG response. The unaccented event that follows an accented event (the first unaccented) has a distinctly different function than the upbeat leading up to an accented event (the last unaccented). The former may have some carry-over effect from the downbeat, but is generally considered a weak beat in the pattern, whereas the latter may show some response reflecting the expectation of the downbeat that is approaching. Here, we may see a conflict of rhythmic function, in which the last position in the pattern is never hierarchically important, versus a more cognitive driven view that this event should get most of the anticipatory response leading to the accent. This cognitive view would then result in a variation of the expectation-induced negative deflection in the EEG, the so-called contingent negative variation (CNV, Walter, Cooper, Aldridge, McCallum, & Winter, 1964). However, this is a slow component generally seen to start at up to 1,000 ms before the expected stimulus (see, for instance, Hamano et al., 1997; but also Chen et al., 2010, for an example of an earlier manifestation). Even so, later (250–500 ms) negative responses are often seen in ERPs in musical or rhythmic contexts (Pearce, Herrojo Ruiz, Kapasi, Wiggins, & Bhattacharya, 2010; Jongsma et al., 2005). Another indication of the strength of a metric event may be the presence of a processing negativity (Näätänen, 1982), an early negative response thought to reflect the recruitment of extra attentional resources. Finally, to investigate the role of external input in rhythm processing, we look at both the externally cued ('perceived'), and internally generated (referred to as 'imagined' or subjective) patterns.

Previous work looking into ERP responses to events in a specific metrical context has focused mainly on intensity decrement deviants in different metric positions added to identical or physically accented stimulus trains (Brochard et al., 2003; Abecasis, Brochard, Granot, & Drake, 2005),

resulting in different P300-responses for different metric positions, namely larger P300 amplitudes for accented events in parietal regions. This confirms the spontaneous nature of this process, as no instruction to superimpose a structure was given, and suggests enhanced processing for accented events. This is supported by more recent work from this group, showing that an early, small processing negativity may be seen at the left mastoid channel for accented events in both standard and deviant forms (Potter, Fenwick, Abecasis, & Brochard, 2009). To disentangle the task of deviancy processing from the mechanism of the metric cycle itself, we here look at responses to physically identical sounds (except for the accent in the perception task) in different contexts. As such, no clear predictions can be made for the ERP response to a pattern without deviants. In a recent study, Fujioka, Zendel, and Ross (2010) investigated the brain response to different subjective metrical events as measured with magneto-encephalography (MEG), focusing on accented events (downbeats) and the last unaccented events (termed upbeats). Using 2-beat and 3-beat patterns and spatial-filtering source analysis, they found that responses from hippocampus, basal ganglia, and auditory and association cortices showed a significant contrast between the up- and downbeats of the two patterns while listening to identical click stimuli. However, they did not combine events from different patterns to find any commonalities between them. Another study that also focused specifically on voluntary accenting of ambiguous stimuli, also using MEG, found no difference in the event-related field (ERF, presented as low-frequency content from 1–10 Hz) between subjectively accented and non-accented events (Iversen, Repp, & Patel, 2009).

Based on these studies, we expect the actual accents in the stimulus to result in an increased N1 amplitude (Näätänen & Picton, 1987) due to the intensity differences caused by the accent, not present in the unaccented events. Additionally, the different spectral properties of the accent may enhance the P2 response (Meyer, Baumann, & Jancke, 2006). As for the subjective accents, the literature does not offer a clear-cut prediction. Considering the different types of unaccented event, we expect that if an anticipatory response for the accent is present in the last unaccented events, this will not be present in the other groups of events, thus predicting the first unaccented event not to show either the increased N1/P2 or any sign of anticipation. As no previous work has, to our knowledge, directly compared ERP-responses to different unaccented events in a rhythmic pattern, this part of the work is still exploratory. By comparing events with similar functions we may uncover common processes over different rhythmic patterns, which can be further investigated by decomposing the responses.

Method

Participants

Ten volunteers, recruited at the Radboud University of Nijmegen, participated in the experiment. Each gave their informed consent to participate. All participants were right-handed and had normal or corrected-to-normal vision. None of them had a known history of neurological illness. Musical training was not a criterium for inclusion or exclusion in the study, three of the participants had received formal music training but none were professional musicians. Two datasets were rejected due to a disproportionate number of artifacts (see below for procedure). The reported analyses were carried out for the remaining eight participants (5 male, mean age 38.2, SD 11.6).

Stimuli and equipment

Three stimulus patterns were used: binary, ternary and quaternary rhythms, consisting of 2-, 3-, and 4-beat cycles. As we expected the imagery response to be much smaller than the perception response, twice as much data were collected for this task. The stimulus sequences were constructed to collect a maximal amount of imagery data, and were made up of four parts: a perception part that also functioned as an instruction, a fade into the imagery part, the imagery part itself, and a probe accent as an attention check at the end, explained further in the procedure. A schematic example of one of the sequences is shown in Fig. 2. For every sequence, the metronome tick was played throughout and functioned as the time-lock while keeping the tempo stable. The accents were positioned to establish a pattern, every 2, 3 or 4 metronome beats. After three repeats there was one cycle in which the accent is played softly (fading) and after this no accents are sounded anymore. The subjects were instructed to imagine the accent pattern continuing. At the end of the sequence an extra accent (probe) was played. This probe accent could appear at any point in the pattern, and participants had to indicate whether this probe
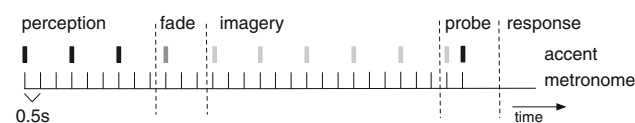

Fig. 2 A schematic overview of a typical stimulus sequence, in this case a ternary beat pattern, with a probe on an unaccented position. With an IOI of 500 ms between events, the sequence consists of 3 perceived patterns or cycles, one transition or fade cycle, and five imagery cycles. As the first cycle in each task is not used, two perception and four imagery cycles per sequence are used for analysis. The sequence was designed to induce an accenting pattern as is represented for a binary pattern in Fig. 1, but applicable to all the patterns used here

coincided with an imagined accent or not. This task was added to control for attention and to check whether the subject was still on track. While the stimulus played, a fixation cross was shown on a screen. All sequences were constructed this way, only differing in the number of events per cycle. The stimuli can be listened to at http://www.nici.ru.nl/mmm.

EEG was recorded using a Biosemi Active-Two system with 256 EEG channels mounted into an elastic cap, and six auxiliary channels (double mastoids, horizontal and vertical EOG), and sampled at 512 Hz. The fixation cross and instructions were displayed on a 15″ TFT screen, and stimuli were played through passive speakers (Monacor, type MKS-28/WS) at a comfortable listening level, adjusted to the preference of the participant. The stimuli were programmed in POCO (Desain & Honing, 1992) and the resulting MIDI file was converted to audio by Quicktime Musical Instruments using general MIDI commands for low bongo (key 61), velocity 0.7×127 as the metronome and high wood block (key 76), velocity 0.8×127 as the accents. The sounds were presented with an inter-onset interval (IOI) of 500 ms and a duration of 200 ms. The analyses were performed in MATLAB (Mathworks, Natick, MA, USA), making use of the FieldTrip toolbox for EEG/MEG-analysis (Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands. See http://www.ru.nl/neuroimaging/fieldtrip).

Procedure

Preceding the actual experiment, a practice session was completed, allowing participants to get used to the task and ensure that they understood it. The practice trials were made slightly easier, with a longer perception-phase and longer fading period. To ensure good understanding of the task, the practice procedure was determined as follows: a counter was set, which counted the correct answers to the probe-tone task. Whenever a wrong answer was given, the counter was set back two points, the practice block ended when the counter had a value of five. A fixation cross was presented at a varying interval before the start of every first beat, appearing between 1 and 1.8 s before the sound started, with a jittered duration to prevent the occurrence of temporal expectation. This fixation cross remained on the screen for the entire sequence. Participants were instructed to neither move nor use motor imagery or inner speech, for instance, by counting. Their specific instruction was to imagine the *sound* of the accent continuing after it had faded. The experimental task for the probe tone at the end of the sequence was to match it to the internally generated pattern, and respond 'yes' to a congruent probe and 'no' to an incongruent probe accent through a button press. One block in the experiment consisted of 12 sequences of each

of these 2-, 3-, or 4-beat patterns, resulting in a total of 36 randomized sequences. Four of these blocks were recorded per subject, yielding roughly 200 instances of every event for the imagery task and 100 for the perceptual task, not taking into account any rejection of data due to artifacts.

Analyses

First, some preprocessing steps were taken. The raw EEG signal, which was originally sampled at 512 Hz, was temporally down-sampled to a sampling frequency of 128 Hz. To segment the data, a time window of −50 to +450 ms was chosen around each metronome tick where 0 is the sound onset. These data segments of 500 s will from here on be referred to as trials. Trials from a sequence with a wrong answer to the probe accent task were rejected (on average 4 sequences per participant, amounting to 2.7% of the data). To avoid possible start-up or state-change effects, the first period of the perception or imagery pattern of a sequence was not used for analyses (marked 'start' in Fig. 2). Channels with poor signal quality were rejected based on the DC offset, with a cut-off of 35 mV, and a variance of 500 μV)^2. From the remaining data, the removed channels were reconstructed by spherical spline interpolation (Perrin et al., 1989). After this a common average was subtracted. If more than 25% of channels were rejected, the trial was rejected as a whole. If more than 30% of trials were rejected based on these criteria, the whole data set was not used. This resulted in exclusion of two participants and left an average of 89 trials (SD 10.6) for every unique event in the perception task, and 184 (SD 22.3) for imagery events.

To test our hypotheses, four different comparisons were made between the types of events, shown in Fig. 3. The ERP was calculated for several types of events by grouping them differently, referring to these groups as conditions. For comparison 1, the accented/unaccented contrast, all the accented (the first beat of the 2-, 3- and 4-beat patterns) and all the unaccented (all other) events were grouped. To investigate the response to different types of unaccented

beat, the first unaccented (the second beat of each pattern) and the last unaccented or upbeat (the last beat of each pattern) were grouped and compared to the accented events (comparison 2 and 3). Here, the two-beat pattern was included in the assumption that the second event in the pattern is a combination of both responses. To investigate the actual differences between different unaccented beats, the first and last were compared to each other (comparison 4). This last comparison is only made up out of the 3- and 4-beat patterns (to avoid the overlap of the 2-beat unaccented event). Because of how the trial sequences were constructed, there were about twice as many trials for the 'imagery' task as for the 'perception' task; however, they are not directly compared to each other. Thus, most conditions are built up of three events, yielding an average of 265 (min 228, max 322) trials per condition for perception, and 560 (min 456, max 725) for imagery per participant. Only 'all unaccented' is built up of twice as many events. When directly compared to each other, first and last unaccented are only based on two events. The condition ERPs were compared using a cluster randomization test. This is a non-parametric statistical test, offering a straightforward way to solve the multiple comparison problem present in EEG data by allowing biophysically motivated constraints, namely clustering over channels, increasing the sensitivity of the test (Maris & Oostenveld, 2007; Maris, 2004). The significance level of the temporal clusters as well as the spatial clusters was set at $p < 0.05$.

We then tested the assumption of decomposability of the response by running a PCA on all the ERP data. This yields a data-driven way of validating the comparisons that we made in a hypothesis-driven way by grouping the trials according to event type. Assuming that the ERP signatures of different subprocesses combine in a linear way, comparing the amount of variance that each component explains for the different event types offers an unbiased method of supporting the choices made top-down in the event groupings. The results yield a weight distribution over the scalp and a time course for each component. We first decomposed the two tasks separately (perception/imagery), and
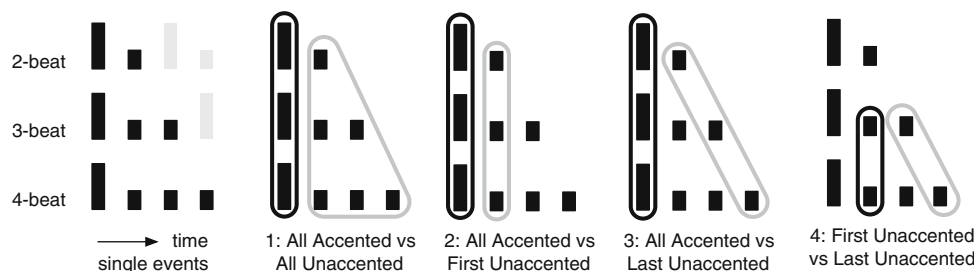

Fig. 3 Condition comparisons made in ERP analyses, reflecting the different hypotheses. Each beat pattern is shown as one accented (bigger) event and a number of unaccented (smaller) events, starting with the single events on the left (with repeating events in *gray*), and the groupings shown for each of the four condition comparisons. The conditions are referred to as AA (all accented), AU (all unaccented), FU (first unaccented) and LU (last unaccented). For the last comparison only the 3- and 4-beat patterns were used to avoid overlap

then also decomposed the whole dataset as one task (rhythm processing). Running the PCA on the average of the perception and imagery data reveals the processes that are common over the two tasks, again with the contribution of each condition to each component to see which component is active when. A cluster randomization test was performed on the contributions of each component to a condition ERP to see if the difference in contribution of a component to a condition was significant.

Results

ERPs

Significant differences were found in every comparison made, for an overview of the effects on Cz and FPz (chosen for comparability to known 10–20 positions) and FC1 (for the maximal effect), see Fig. 4. Although here, clusters with $p < 0.05$ are shaded, $p < 0.0001$ for the main clusters in each of the comparisons. While keeping in mind that for the perception task, the ERPs are inherently somewhat noisier due to smaller number of trials, we can still see some regularities.

Early effects (100–300 ms)

First of all, the accented events (in comparison 1, 2 and 3) consistently show a larger N1/P2 complex in perception

than any of the unaccented, mainly visible as a larger positive deflection between about 100 and 250 ms. For imagery, where there is no difference in the stimulus, this effect is also significant, albeit smaller. This effect is visible at central locations with the strongest difference for FC1, just left-lateralized from Cz (FC1). Interestingly, the first unaccented events show an early ($\approx$100 ms) positive deflection as well, which distinguishes them from the last unaccented events (comparison 4), and which averages out in the combined condition of all unaccented events. This difference starts earlier in imagery than it does in perception, but is significant in both.

Late effects (300–450 ms)

At higher latencies, an effect at >350 ms with a mainly frontal localization also shows differences between the types of events. The accented and last unaccented events each show a negative deflection, which is not there for the first unaccented. During imagery, this effect is slightly larger for the last unaccented events, differing from the accented at more central electrodes (comparison 3). In perception this is hard to distinguish from the central effect described before, as the P2 increase carries over. The late difference between the first and last unaccented events is consistent for both perceived and imagined patterns, as would be expected considering that in this case there is no difference between the stimuli (all unaccented events).
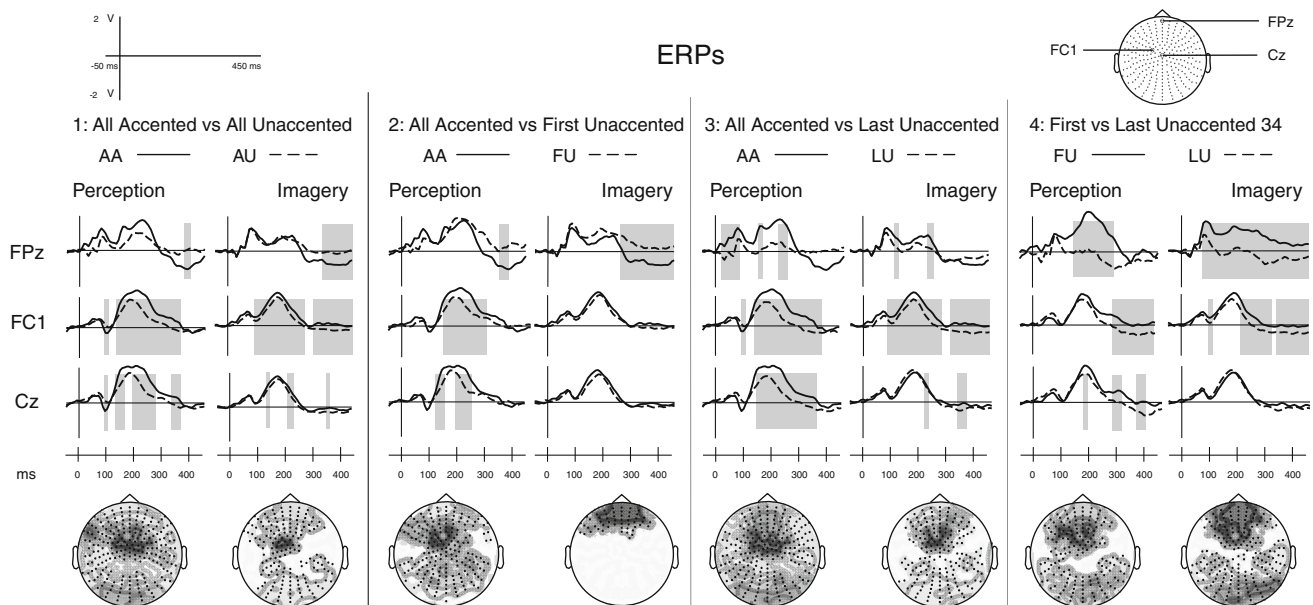

Fig. 4 The ERPs of all the different comparisons, for perceived (*left*) and imagined (*right*) accents, with the *x*-axis running from −50 to 450 ms after the metronome click and the *y*-axis running from −2 to 2 μV. The distribution of channels with significant differences

between these events is plotted below, the channels that show a significant cluster with $p < 0.05$ are highlighted (*bold*) and the *shading* depicts the duration of the significant difference. Below each column of plots, the time-scale is shown in ms

PCA

To test the hypothesis of separate effects with distinct distributions and time courses, a PCA analysis was performed on each task separately. The PCA yields a number of components that each has their own weight distribution over the scalp, an amount of variance of the signal explained by this component and a time course of its activity. The distributions of the components on separate tasks are shown in Fig. 5 (top two rows, P1-2-3 and I1-2-3). The first component explains the most variance by far for both tasks, and their distributions correlate highly ($r \geq 0.99$). The consecutive two components appear similar in terms of distribution and explained variance for the two tasks, but seem swapped in order, with component P2 correlating best with I3 ($r = 0.5$) and P3 correlating best with I2 ($r = 0.7$). The correlations of these distributions suggest that these first three components represent related subprocesses, supporting the next step in which the decomposition was carried out on both tasks together (B1-2-3). Of the resulting components, the weight distribution of B2 correlates highly with both P2 ($r = 0.9$) and I2 ($r = 0.7$) and B3 with P3 ($r = 0.9$) and with I3 ($r = 0.8$). This supports the notion that the processes (or combination thereof) associated with these components are related. Assuming that the three components that explain most of the variance of the data are indeed shared over the two tasks, we only discuss the components identified over both tasks. In this way, we can use the spatial properties of the activity explaining most variance for the mean of the two tasks investigate how active these processes are for the different events. From component 4 on, the

explained variance of individual components is below 5% and will not be discussed further (the scree plot is shown on the right panel of Fig. 5). Looking further into the activation patterns of these components for both tasks separately, Fig. 6 shows the contribution of the three components to each type of event, with time courses shown below each distribution for each task, and significant differences plotted below for the different comparisons. Although the PCA inherently tends to make orthogonal distributions, the subprocesses shown here are also supported by visual inspection of the ERP data when comparing the time courses and locations of significant differences. The three main components are discussed in turn.

Component 1

The first component has a central distribution, and shows a positive peak around 100 ms, and then shows a large positivity after about 200 ms. The shape of the time course is similar for the perception and imagery tasks, but the strength is different between the accented and the unaccented events. There is also a significant effect between the first and last unaccented events in a similar pattern for both tasks. For the perception task, the difference between the accented event and the first and last unaccented events appear to separate in time, where the accented events show a late negativity. This is not the case for imagery, where it mainly distinguishes the different unaccented events. This component likely relates to the N1 in perception and the P2 response in both perception and imagery.
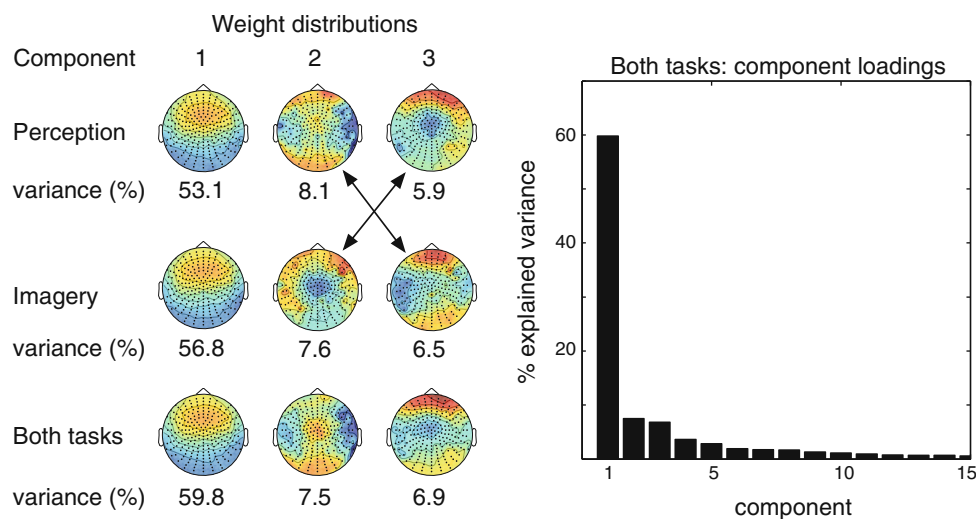

Fig. 5 The *left panel* shows the distributions and explained amount of variance for the PCA performed on separate tasks (components named for the tasks; perception, imagery, and both, numbered in the order of explained variance: P1-2-3, I1-2-3 and B1-2-3). While the separate tasks show weight distributions for the first three components that correlate significantly (shown by the *arrows* between P2-I3 and P3-I2),

the distributions for both tasks together appear to capture the same activity (B2) but isolating additional frontal activity in the third, frontal component (B3). On the *right panel*, the explained variance is shown for the first 15 principal components of the decomposition of both tasks together, showing the first three to be the most important (explaining 74.3% of the total variance)

Fig. 6 The time courses of the contributions of the first three components from the PCA over both tasks are shown in µV, with the weight distributions plotted above. The weights from the combined PCA are used to show the contribution of these components to the tasks separately. For both perceived and imagined patterns, the part of the ERP that is explained by the component is plotted as a time course below for each event type, and significant differences ($p < 0.05$) between these contributions are shown for each comparison (1–4) in the *bars* below the time courses
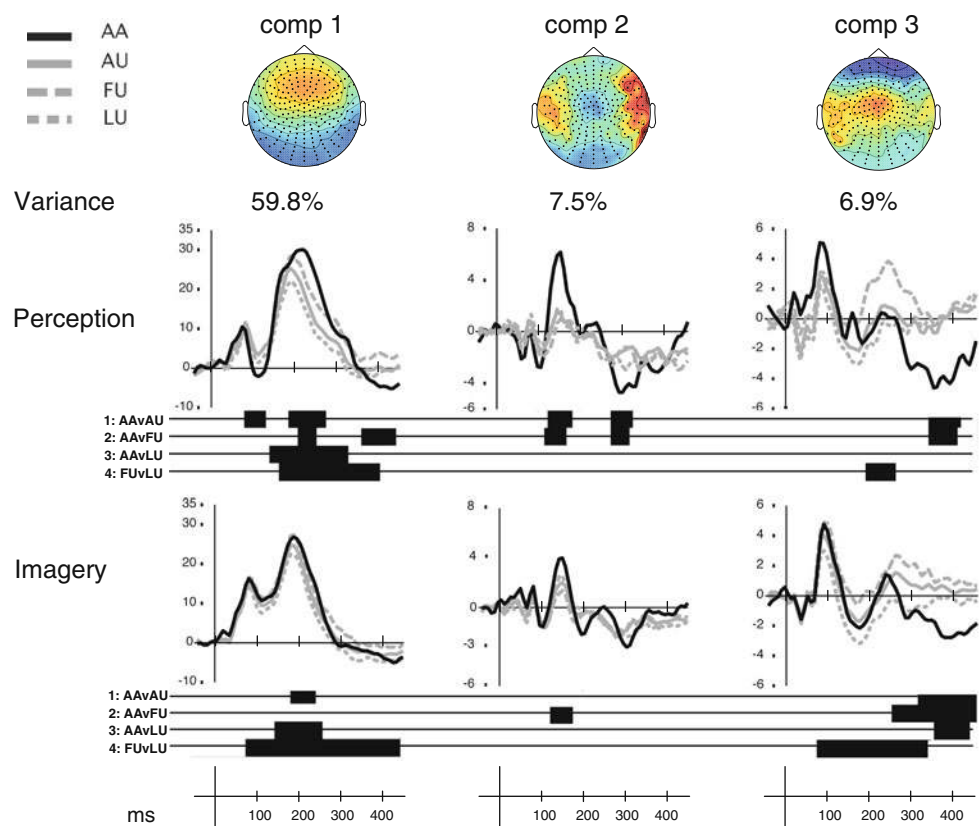
Component 2

The second component is a lateralized activity pattern, explaining 7.5% of the data. It appears to contribute mainly to accented events in the perception task with a strong peak at ≈150 ms and a negativity at ≈300 ms, but does not distinguish between the unaccented events. In the imagery task this response is much smaller, but still significant. Thus, it appears to capture the part of the brain activity associated with the perceptual accent, but, interestingly, the contribution to the ERPs only differs significantly for the second comparison, all accented versus first unaccented, for both tasks. As the latency of the differences between time courses for the first two components appears to coincide with the N1/P2 complex, one interpretation may be that components 1 and 2 represent two subcomponents of this complex.

Component 3

Component 3, a central/frontal activity pattern that explains 6.9% of the variance, shows an early peak in explained variance for all events in both tasks, but after about 30 ms starts to distinguish different unaccented events during imagery, and later on (at ≈300 ms) starts to contribute to the difference between accented and unaccented events. The difference in this contribution is markedly smaller for the perception task, and only reaches significance in

relatively small time windows (comparison 1 and 2 at ≈400 ms, comparison 4 at ≈250–300 ms). The localization and time course suggest that this is an attention-related process, and may include effects of anticipation.

As an exploratory check on the groupings chosen to form the conditions, the time courses of the components on single events are shown in Fig. 7. The dash pattern represents the grouping made in Fig. 3, so comparable activation patterns for similar dashed lines support our grouping. Looking at these time courses clarifies some of the significance results shown in Fig. 6, namely the absence of significance for Component 3 in perception, the grouping does not appear to reflect structure here. However, for components 1 and 2, and in imagery component 3, the type of event tends to group together, supporting our design. Most obviously, for perception, component 2 indeed isolates the response to all accented events at about 150 ms, and in imagery component 3 isolates the response to first unaccented events (3b2 and 4b2) at about 300 ms. Component 1 (shown at a much larger scale than the other components) reveals the same grouping, supported by the statistical testing of the group means.

Discussion

In the current study, the ERP signatures of rhythmic processing were investigated for both actual and subjectively
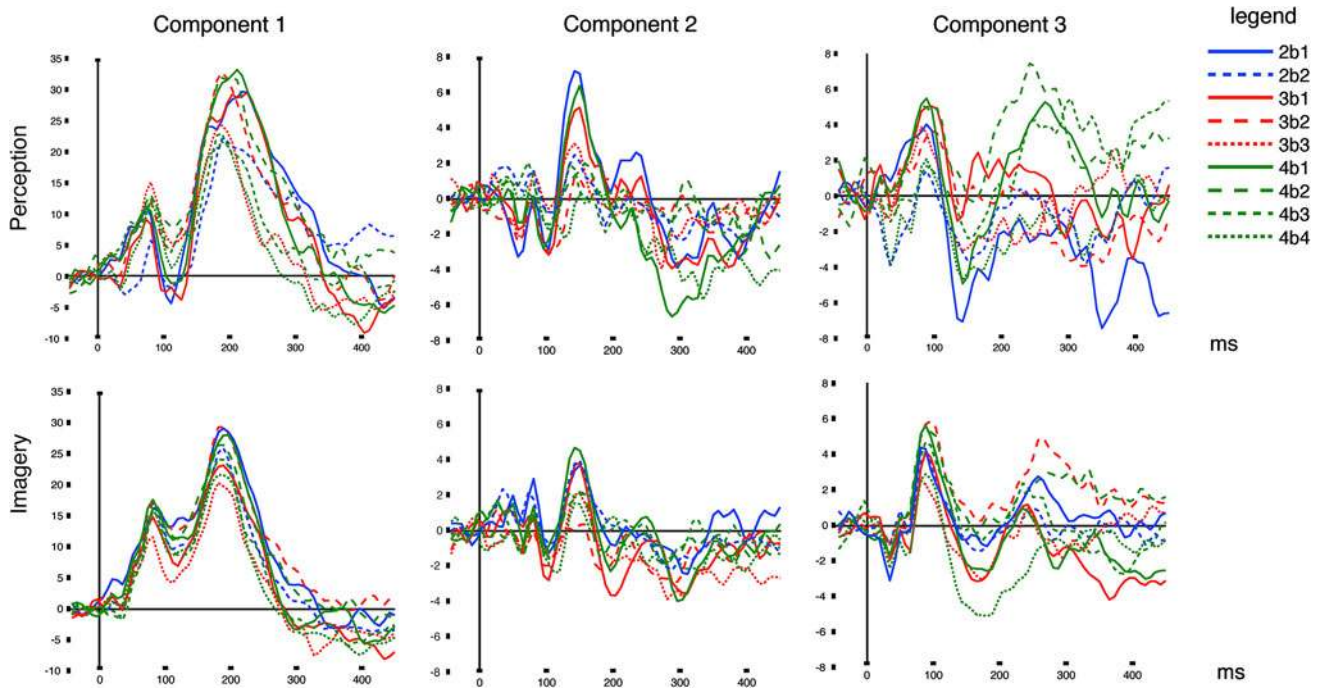
Fig. 7 The time courses of the component contribution of single events are shown in µV with the colors separating the patterns, the *solid lines* representing the accented events and the different *dashed lines* representing different unaccented events as is shown in the legend. Events are denoted by the pattern and the position (i.e. 4b3 is the 3rd beat in a 4-beat pattern). The scales vary to maximally visualize the time course shape; the first component shows a far larger response than the other two

accented rhythmic patterns. Significant differences were shown between responses to metronome ticks on different positions within a rhythmic pattern. Both hypotheses were confirmed; differences were seen between accented and unaccented events in perceived and imagined rhythms, as well as further differentiation of unaccented events. The ERPs showed the predicted increased central N1-P2 response for actual and, to a lesser degree, subjective accents as compared to all unaccented events. This effect is stronger when comparing all accented events to the last unaccented event of a pattern than when compared to the first unaccented. Although this effect is strongest on channel FC1, slightly left-frontal from Cz, we do not interpret this as a lateralized effect, given the distribution of the significant clusters shown in Fig. 4. Additionally, a late, frontal positive response is seen in 'first-unaccented' events but not in 'last-unaccented' events, as compared to the accented events. The final comparison between different unaccented events supports the notion of independence of these two different responses. This implies that rhythm processing entails more than simply serially processing only accented and unaccented events, but that there are different responses to unaccented events with a different context but an identical sound. Although comparing events with a different immediate history poses some problems, here the results appear to be quite straightforward in that the events with an actual accent show a stronger N1/P2 response ending after

≈350 ms, whereas all other events are based on an identical stimulus (the metronome tick). As the time between ticks (500 ms) is long enough not to expect purely perceptual responses to leak into the next event, any difference we see is due to cognitive aspects of rhythm processing. This is true especially in the imagery task, where all differences between events are completely subjective. Thus, the later differences between unaccented beats, here interpreted as purely cognitive instead of perceptual, can only be caused by the rhythmic context.

The most interesting finding here, which has not been shown before, is the difference between different types of unaccented event. Given that, in both perception and imagery tasks, the sound stimuli for the unaccented events are identical, it is not surprising that the significant effects are similar in location and latency. It does, however, imply that the mechanism we see is independent of external input, and thus is active for perceiving and self-generating a rhythmic pattern, and is mediated by metric position. Although the pattern of significant differences is comparable for the two tasks, the largest difference is visible in the comparison between the accented and first unaccented, in which the increased N1/P2 effect we see for perceived patterns is completely absent. It appears that, without the perceptual response to the actual accent, the last unaccented events show a decreased N1/P2 amplitude, and the first unaccented events only show a decreased late frontal negativity when

compared to accented events. Although their interpretation is not straightforward, the impression of two distinct processes is given.

This hypothesis was tested by decomposing the ERP data with PCA. The distributions of the components of the different tasks separately (perception and imagery) correlate highly, which supports the validity of decomposing the two tasks together, namely rhythm processing, with or without external input. This is likely a composite process, including elements of mnemonic processing, tempo tracking, regularity detection, expectancy generation and others. The first three components explain almost 75% of the total variance. The distributions found for the different subcomponents connect well with the existing literature. Kuck, Grossbach, Bangert, and Altenmüller (2003) found, when researching the distributions of rhythm and meter processing, that there was sustained cortical activation over bilateral frontal and temporal brain regions, that did not differ much for the two tasks. The different subcomponents of accenting were obviously present in their stimuli as well, and perhaps may also be decomposed. In a study concerning speech rhythm, Geiser, Zaehle, Jancke, and Meyer (2008) found that adding an explicit rhythm judgement task, thus directing attention to the rhythmicity of the (spoken) stimulus, increased activity in the supplementary motor areas and the inferior frontal gyrus, both bilaterally. These sources, related to the explicitness (i.e. directed attention) of a rhythmical task, may well be implicated here. However, more work is needed to confirm this. Even so, Geiser, Ziegler, Jancke, and Meyer (2009) separated out meter and rhythm deviants and found the ERP response to rhythmic deviants to be maximal in frontal areas, and dependent on directed attention, while meter changes elicited a response more centrally and laterally distributed.

To asses the activity of the processes explaining most variance over both tasks for each task individually, their distributions were used to visualize the activity for the two tasks separately. The time courses of the components show specific activations for specific aspects of the rhythmic patterns. The first and biggest component contributes to the N1/P2 activity, also showing the first unaccented event to be more like an accented event than the last unaccented event. As this component explains around five times as much variance as the other two, for both tasks, the fact that a difference between the conditions is visible here provides the main support for the grouping into event categories that was decided on. The second component appears to respond mainly to perceived accents, but does not distinguish between the accented events and the last unaccented events, likely also contributing to the increased N1/P2 complex seen in the ERP in perception. Then finally, the third component contributes to the later, more frontal effect, distinguishing well between unaccented events only in the

imagery task at a relatively early latency ($\approx$100 ms) and between the accented and all unaccented events a bit later (at $\approx$350 ms). Inspection of the contributions of single events to the different components supports the groupings of events used here, according to metric context, save for component 3 in perception. This difference may be interpreted as a result of increased focus or effort during the self-generation of the rhythmic pattern in imagery which is not necessarily there during perception. Alternatively, it may be due to the smaller number of trials for the perception task. The finding that decomposing both tasks together as one still yields interpretable results was unexpected, and indicates that the cognitive processing of rhythms, be they externally presented or internally generated, shares a common mechanism. The relevance of the decomposition is obviously dependent on the assumption that the EEG traces of the components combine linearly to form the total signal. Other methods (that make the same assumption) can also be used to decompose EEG data into subprocesses, such as ICA (Makeig, Jung, Bell, Ghahremani, & Sejnowski, 1997) or linear regression (Hauk, Pulvermüller, Ford, Marslen-Wilson, & Davis, 2008; Schaefer, Desain, & Suppes, 2009), and a solid comparison of these different methods may be a subject of future work.

A number of assumptions made in the design may have consequences for the interpretation of the results. First, the assumption that the second event in a 2-beat pattern includes characteristics of both first and last unaccented events in a pattern may have influenced the contrast with the accented events. More detailed analyses are needed to test this, but as this would not exaggerate the difference but instead diminish it, the effect may actually be a bit larger than demonstrated here. The decomposed time courses per event shown in Fig. 7, however, support our assumption. Then, considering that imagery is never completely controlled we must allow for the possibility that participants were in fact using inner speech or imagery after all, in contrary to explicit instructions. Also, as we were interested in collecting a maximal amount of imagery data, the stimuli were constructed to always have perception preceding imagery. Although the first cycle of each sequence was never used and treated as an instruction cycle, there was a fixed order of tasks in the design. However, as it is not possible to 'continue a pattern internally' that is not presented first, this opportunity was used to extend this presentation to a series of perception trials. Finally, there were subtle differences between the tasks that involve more than the task itself. As previously mentioned, the fact that the 'perception' part of the sequence also includes an element of instruction, and preparation for imagery, has to be kept in mind. However, by decomposing the data based on the mean over both tasks the risk of these processes causing the found effects is minimal. Moreover, if these effects would

be present, the increase in variance would again cause an underestimation of the effect instead of an overestimation.

Considering the literature cited earlier, we can say that the N1/P2 effect that was expected for the accented events is actually affecting the unaccented events as well, although the difference is found mostly in the P2-part of the complex, in a similar time window as where Fujioka et al. (2010) found an effect of accenting. Interestingly, this component has been found to be affected by spectral aspects (or timbre) of an auditory stimulus (Shahin, Roberts, Pantev, Trainor, & Ross, 2005; Meyer et al., 2006). Given that the stimuli were identical in the imagery task, in this case the percept is completely self-generated. This is even more interesting in the comparison between first and last unaccented events, where even in the Perception task the stimulus is identical. The later, frontal effect is harder to interpret. The reduced negativity seen in the first unaccented events may support the interpretation of a CNV-like response for the last unaccented events, however the fact that it is also present in the accented events contradicts this. To a certain extent, there is of course anticipation for every event, as the stimulus is intentionally rhythmic. Also, as the first unaccented events all lead to events with different functions the level of anticipation, the levels of anticipation would likely differ. Again, the decomposed time courses per type of event shown in Fig. 7 tend to support the grouping we made here in terms of how components contribute to each event, especially for the primary component explaining most of the variance. If however we interpret the absence of the positivity in the accented events as extra anticipation for the first unaccented beat, this would lend new importance to this event in the cycle, not suggested by either music theory or cognitive theory. On the other hand, if we interpret this as a somewhat late processing negativity present for the accented and the last unaccented, this would fit quite well. Although not explicitly discussed as a component related to rhythmic processing, a frontal component with a similar latency is seen in other studies that involve rhythmic musical stimuli (for instance Pearce et al., 2010; Jongsma et al., 2005), and further work is called for to elucidate this response. If we consider this negativity a default, then its absence for the first unaccented events may be interpreted as reduced processing, which is supported by the lack of information present in the stimulus at this position (i.e. no accent and no anticipation). The early processing negativity found by Potter et al. (2009) was not seen here for the accented events, in either perception or imagery. Looking back at the coupled oscillator models, the decomposition results do not support the view of multiple processes resulting in the responses to the rhythmic events. The difference between the two types of unaccented events is captured mostly in the main PCA component, as is the difference between accented events and the first unaccented

events. Thus, the interplay between varying levels of attention, expectation and processing is not separable by statistical decomposition in our study.

To conclude, the current report shows processing of metronome clicks in a different metric context to result in different ERP responses, and thus to be heavily influenced by attention levels, even without differences in the perceptual input. The decomposition through PCA yields an informative look at the subprocesses involved, offering a decomposition that at least partly relates to the hierarchical levels of rhythm processing. By identifying components that were active over both tasks (perception and imagery), we found support for the notion that similar cerebral sources are active in perceived and self-imposed patterns, although they are clearly not identical. The time courses of these components could be interpreted to separate a more low-level effect on the N1/P2 complex for the perception task, distinguishing accented from unaccented events, from a later, more frontal effect that distinguishes different types of unaccented events in both tasks. As the current data are based only on simple, regular metre, further work is needed to clarify the nature of these responses in the framework of processing models such as coupled oscillators. Also, other IOIs may produce different responses. Even so, a strong case has been made to distinguish between different types of unaccented events within one rhythmic pattern when researching cerebral mechanisms of rhythm processing. Additionally, in the absence of any externally driven process, self-generated or imagined rhythms were shown to be measurable in EEG, differentiating responses based on the rhythmic context.

References

Abecasis, D., Brochard, R., Granot, R., & Drake, C. (2005). Differential brain response to metrical accents in isochronous auditory sequences. *Music Perception, 22*(3), 549–562.

Bolton, T. L. (1894). Rhythm. *American Journal of Psychology, 6*, 145–238.

Boring, E. G. (1942). *Sensation and perception in the history of experimental psychology*. New York: Appleton-Century.

Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The "ticktock" of our internal clock: Direct brain evidence of subjective accents in isochronous sequences. *Psychological Science 14*(4), 362–366.

Chen, Y., Huang, X., Yang, B., Jackson, T., Peng, C., Yuan, H., & Liu, C. (2010). An event-related potential study of temporal information encoding and decision making. *NeuroReport 21*, 152–155.

Desain, P. (1992). A (de)composable theory of rhythm perception. *Music Perception, 9*(4), 439–454.

Desain P., & Honing, H. (1992) *Music, Mind and Machine: Studies in Computer Music, Music Cognition and Artificial Intelligence*. Amsterdam: Thesis Publishers.

Dien J., & Frishkoff, G. A. (2005). Principal components analysis of event-related potential datasets. In T. Handy (Eds.), *Event-related potentials: A methods handbook*. Cambridge. MA: MIT Press.

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition, 77*, 251–288.

Fujioka, T., Zendel, B. R., & Ross, B. (2010). Endogenous neuromagnetic activity for mental hierarchy of timing. *Journal of Neuroscience, 30*(9), 3458–3466.

Geiser, E., Zaehle, T., Jancke, L., & Meyer, M. (2008). The neural correlate of speech rhythm as evidences by metrical speech processing. *Journal of Cognitive Neuroscience, 20*(3), 541–552.

Geiser, E., Ziegler, E., Jancke, L., & Meyer, M. (2009). Early electrophysiological correlates of meter and rhythm processing in music perception. *Cortex, 45*, 93–102.

Hamano, T., Lüders, H. O., Ikeda, A., Collura, T. F., Comair, Y. G., & Shibasaki, H. (1997). The cortical generators of the contingent negative variation in humans: a study with subdural electrodes. *Electroencephalography and Clinical Neurophysiology*, 104:257–268.

Hauk, O., Pulvermüller, F., Ford, M., Marslen-Wilson, W. D., & Davis, M. (2008). Can I have a quick word? early electrophysiological manifestations of psycholinguistic processes revealed by event-related regression analysis of the EEG. *Biological Psychology, 80*(1), 64–74.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science, 182*, 177–180.

Huron, D. (2006). *Sweet anticipation: music and the psychology of expectation*. Cambridge, MA: MIT Press.

Iversen, J. R., Repp, B. H., & Patel, A. D. (2009). Top-down control of rhythm perception modulates early auditory responses. *Annals of the New York Academy of Sciences, 1169*, 58–73.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96*(3), 459–491.

Jongsma, M. L. A., Eichele, T., Quian Quiroga, R., Jenks, K. M., Desain, P., Honing, H., & van Rijn, C. M. (2005). Expectancy effects on omission evoked potentials in musicians and non-musicians. *Psychophysiology, 42*(2), 191–201.

Kuck, H., Grossbach, M., Bangert, M., & Altenmüller, E. (2003). Brain processing of meter and rhythm in music: electrophysiological evidence of a common network. *Annals of the New York Academy of Sciences, 999*, 244–253.

Ladinig, O., Honing, H., Háden, G., & Winkler, I. (2009). Probing attentive and preattentive emergent meter in adult listeners without extensive music training. *Music Perception, 26*(4), 377–386.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*(1), 119–159.

Large, E. W., & Kolen, J. F. (1995). Resonance and the perception of musical meter. *Connection Science, 6*, 177–208.

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

London, J. (2004). *Hearing in time: Psychological aspects of musical meter*. Oxford: Oxford University Press.

Longuet-Higgins, H. C., & Lee, C. S. (1984). The rhythmic interpretation of monophonic music. *Music Perception, 1*, 424–441.

Makeig, S., Jung, A., Bell, D., Ghahremani, A. J., & Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences of the United States of America*, 94:10979–10984.

Maris, E. (2004). Randomization tests for ERP topographies and whole spatiotemporal data matrices. *Psychophysiology, 41*(1), 142–151.

Maris, E., & Oostenveld, R. (2007). Nonparametric testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*, 177–190.

Meyer, M., Baumann, S., & Jancke, L. (2006). Electrical brain imaging reveals spatio-temporal dynamics of timbre perception in humans. *NeuroImage, 32*, 1510–1523.

Näätänen, R. (1975). Selective attention and evoked potentials in humans–a critical review. *Biological Psychology, 2*, 237–307.

Näätänen, R. (1982). Processing negativity: An evoked-potential reflection of selective attention. *Psychological Bulletin, 92*(3), 605–640.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology, 24*(4), 375–425.

Pearce, M. T., Herrojo Ruiz, M., Kapasi, S., Wiggins, G. A., & Bhattacharya, J. (2010). Unsupervised statistical learning underpins computational, behavioral, and neural manifestations of musical expectation. *NeuroImage, 50*, 302–313.

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current mapping. *Electroencephalography and Clinical Neurophysiology, 72*(2), 184–187.

Potter, D., Fenwick, M., Abecasis, D., & Brochard, R. (2009). Perceiving rhythm where none exists: Event-related potential (ERP) correlates of subjective accenting. *Cortex, 45*, 103–109.

Repp, B. H. (2007). Perceiving the numerosity of rapidly occurring auditory events in metrical and nonmetrical contexts. *Perception and Psychophysics, 69*(4), 529–543.

Schaefer, R. S., Desain, P., & Suppes, P. (2009). Structural decomposition of EEG signatures of melodic processing. *Biological Psychology, 82*, 253–259.

Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., & Ross, B. (2005). Modulation of p2 auditory-evoked responses by the spectral complexity of sounds. *NeuroReport, 16*(16), 1781–1785.

Snyder, J. S., & Large, E. W. (2005). Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cognitive Brain Research, 24*, 117–126.

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: an electric sign of sensorimotor association and expectancy in the human brain. *Nature, 203*, 380–384.

Winkler, I., Denham, S. L., & Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences, 13*(12), 532–540.