J. Neural Eng. **8** (2011) 036002 (11pp) doi:10.1088/1741-2560/8/3/036002

Sequenced subjective accents for brain–computer interfaces

R J Vlek[1], R S Schaefer[1], C C A M Gielen[2], J D R Farquhar[1] and P Desain[1]

[1] Donders Institute for Brain, Cognition and Behaviour: Centre for Cognition, Radboud University, Montessorilaan 3, 6525 HE, Nijmegen, The Netherlands
[2] Donders Institute for Brain, Cognition and Behaviour: Centre for Neuroscience, Radboud University, Geert Grooteplein-Noord 21, 6525 EZ, Nijmegen, The Netherlands

E-mail: r.vlek@donders.ru.nl

Received 10 August 2010
Accepted for publication 7 January 2011
Published 4 April 2011
Online at stacks.iop.org/JNE/8/036002

Abstract
Subjective accenting is a cognitive process in which identical auditory pulses at an isochronous rate turn into the percept of an accenting pattern. This process can be voluntarily controlled, making it a candidate for communication from human user to machine in a brain–computer interface (BCI) system. In this study we investigated whether subjective accenting is a feasible paradigm for BCI and how its time-structured nature can be exploited for optimal decoding from non-invasive EEG data. Ten subjects perceived and imagined different metric patterns (two-, three- and four-beat) superimposed on a steady metronome. With an offline classification paradigm, we classified imagined accented from non-accented beats on a single trial (0.5 s) level with an average accuracy of 60.4% over all subjects. We show that decoding of imagined accents is also possible with a classifier trained on perception data. Cyclic patterns of accents and non-accents were successfully decoded with a sequence classification algorithm. Classification performances were compared by means of bit rate. Performance in the best scenario translates into an average bit rate of 4.4 bits min^{-1} over subjects, which makes subjective accenting a promising paradigm for an online auditory BCI.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

In general, humans have a good sense for basic metric structures (Michon and Jackson 1985), such as an auditory pattern where every first beat out of two, three or four beats is accented. These metric structures in western music are usually stereotyped as a march (ONE-two), waltz (ONE-two-three) or common rock rhythm (ONE-two-three-four). It has been shown that our sense for metric structures is not only relevant for the perception and production of music (London 2004) but also plays a role in speech (Vatikiotis-Bateson and Kelso 1993) and in motor control tasks (Kelso 1982). The cognitive process responsible for inducing our sense for metric structures occurs in a subconscious and fairly automatic way. This is demonstrated by the so-called clock illusion or 'tick-tock' effect (Brochard *et al* 2003). Here, a binary accenting pattern is automatically induced in the brain when a series of isochronous sound pulses is presented. The name 'clock illusion' refers to one of the best examples of this mechanism where the sound of a clock, which sounds identical for every pulse ('tick-tick-tick-tick...'), is usually perceived with an induced subjective accent ('tick-tock-tick-tock...'). The mechanism inducing these accents is known as subjective accenting or subjective rhythmization (Fraise 1982, London 2004). Brochard *et al* (2003) found that subjects exhibited different neuronal responses to loudness deviations at even and odd positions in a steady pulse train, reflecting binary chunking.

Several studies have explored the perception of metric patterns and stimulus-induced responses in EEG. These studies have shown that both the perception of metric patterns (Snyder and Large 2005) and the expectation of an accent was reflected

in EEG-activity (Zanto *et al* 2006, Jongsma *et al* 2005, Snyder and Large 2005, Desain and Honing 2003, Schaefer *et al* 2010). In a recent study Snyder and Large (2005) reported that (non-phase-locked) gamma-band activity (GBA) in EEG can reflect the metric structure of the stimulus and that at an omission of a stimulus this GBA may persist. This suggests that a form of imaginary rhythm or internal clock is active. Subjective accents can also be added voluntarily, thus making it a deliberate process. Iversen *et al* (2009) investigated this phenomenon and described an effect in the upper beta-band of MEG measurements at subjectively accented versus non-accented tones.

The focus of this study lies with subjective rhythmization as a mental task for driving a brain–computer interface (BCI) (Dornhege *et al* 2007, Gerven *et al* 2009). A BCI system allows a user to control an output device (for instance a speller, a cursor or an automatic wheelchair) with brain activity. By voluntarily performing a specific mental task a user can encode intentions into his or her brain activity. This activity is measured and in real time (or as close as possible) decoded by a computer and turned into the control signal for an output device. An intuitive mental task, such as subjective rhythmization, could be a useful addition to the existing variety of BCI tasks, such as the P300 (Farwell and Donchin 1988), steady state evoked potential (SSEP, Regan 1977, Müller-Putz *et al* 2005) and imagined movement paradigm (Pfurtscheller *et al* 2006, 1997), some of which can be difficult to perform or require much attention and concentration. The introduction of new mental tasks could also be a way to overcome the so-called BCI illiteracy (Dornhege *et al* 2007) of subjects for tasks commonly used. Recent developments in the domain of auditory BCIs predominantly yield systems that provide feedback in the auditory domain, but are based on (to BCI) rather conservative mental tasks, such as modulation of sensorimotor rhythms (SMR, Nijboer *et al* 2008), slow cortical potentials (SCP, Pham *et al* 2005) or P300 responses to auditory events. Furdea *et al* (2009) attached acoustically presented numbers to a five-by-five classic P300 spelling matrix (Farwell and Donchin 1988), while in a similar way Klobassa *et al* (2009) attached environmental sounds to a six-by-six matrix. Schreuder *et al* (2010) reported using spatial hearing as an informative cue for evoking ERP responses (predominantly P300). Auditory stimuli were presented through five speakers sequentially and in addition to the spatial information thus provided, auditory stimuli were also acoustically different per speaker. Alternatively, systems have been reported related to the concept of auditory stream segregation (Hill *et al* 2004, Kanoh *et al* 2010). In these systems ERP responses to deviants in two streams of auditory stimuli elicited detectable differences in EEG, depending on the subject's attention to one of the streams. With novel paradigms in the auditory domain, accessibility of BCIs may increase for specific groups of users, for instance to users with a visual impairment who are not capable of using a visual P300 speller.

In order to assess the feasibility of subjective rhythmization as a task for BCI, we investigate whether subjective accents can be decoded from EEG on a single-trial basis and, more specifically, compare various approaches to decoding of subjective accents. Data segments are broken down into single accented and non-accented beats and classified. This approach is extended by a sequence classification algorithm. Aiming at an easy-to-use BCI, we also investigate the possibility of training classifiers on perception instead of imagery data. These approaches are compared to a more conservative approach where longer segments of data are classified at once.

2. Materials and methods

2.1. Experimental design and data acquisition

Ten subjects, five females and five males, aged between 22 and 34 years (mean age 27), participated in this study. One subject (S9) had professional musical training, and six participants (S1, S3, S5, S7, S9, S10) actively play a musical instrument. When asked, none of the subjects reported to be diagnosed with any neurological disorder or hearing deficiency. The experiment was undertaken with the understanding and written consent of each subject, approved by the ethical committee of the faculty of social sciences at the Radboud University Nijmegen, and in compliance with national legislation and the *code of ethical principles for medical research involving human subjects* of the World Medical Association (Declaration of Helsinki).

Subjects were seated in a comfortable chair in an electrically and acoustically shielded room at a distance of approximately 0.5 m from a 17″ TFT computer monitor. Two speakers (Monacor, type MKS-28/WS), placed on each side of the monitor, were used to present auditory stimuli to the subjects (stimuli can be found online at http://www.nici.ru.nl/mmm/). A Biosemi active-electrode set (Ag-AgCl) with 64 electrodes was used in combination with an ActiveTwo AD-box to measure EEG at a sampling frequency of 2048 Hz. No further filtering or processing was done at the stage of recording. Simultaneously with the EEG, an electro-oculogram (EOG) was made in order to be able to exclude eye motions as a possible source of information during EEG classification. Two pairs of auxiliary electrodes were placed. One pair was positioned above and below the left eye to measure eye movements in the vertical direction. The other pair was positioned on the temples to measure horizontal eye movements.

The stimulus sequences consisted of three phases, a perception phase, a fade and an imagery phase. A metronome was playing throughout the whole sequence (see figure 1). In the perception phase of the sequence, an accent was superimposed on the metronome every two, three or four beats, thus creating binary, ternary and quaternary patterns, respectively. Throughout this paper we will refer to the span of such a pattern with the term 'cycle', which could be considered equivalent to the musical term 'measure'. The metronome played at 120 BPM (beats per minute), resulting in inter-onset intervals of 0.5 s between successive ticks. The rate of 120 BPM is chosen to avoid overlap of the expected perceptual EEG responses, such as the auditory evoked potential (AEP), which can have components as late as 400 ms (Burkard *et al*

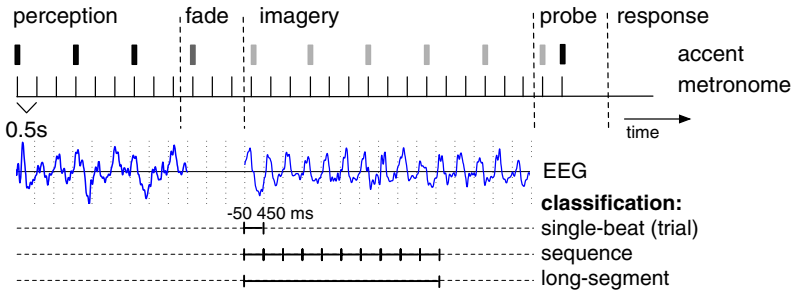

Figure 1. The structure of a single sequence in the experiment is shown, in this case for a three-beat pattern. The sequence starts with three cycles of a ternary metric pattern (perception phase), followed by one cycle (fade phase), where the intensity of the superimposed accent was reduced by 4 dB. Then, the subject had to imagine the accenting pattern for five cycles (imagery phase). At the end, an accented beat was presented to test whether the subject maintained the correct rhythm. The lower part of the figure schematically highlights the different ranges of data taken from the 64-channel EEG for each of the different classification methods described in sections 2.3, 2.4 and 2.5, respectively.

2007), and to stay close to a tempo that is easy to track by human subjects (Fraise 1982). The sound was presented at a peak level of 57 dB(A) for all subjects. In the perception phase, accents were added with the general MIDI sound 'high woodblock'. This accent increased the peak loudness of the stimulus to 65 dB(A). During the fade phase, as a transition from the perception to the imagery phase, the accents were played less loudly, decreasing the peak loudness of the stimulus to 61 dB(A). In the imagery phase the accent was no longer added. A sample sequence is illustrated in figure 1, showing a sequence of a three-beat pattern.

At the start of each sequence, a white fixation cross of 3 cm was shown on the monitor. The appearance of the cross indicated the start of a sequence to the subject and served as a fixation point for the eyes throughout the sequence. After a random delay between 1.0 and 1.8 s after the onset of the fixation cross, the pattern started. The accented pattern was first played for three cycles, which is indicated as the perception phase in figure 1. Subsequently, the pattern was played for one cycle during the fade phase, followed by five cycles containing only the metronome in the imagery phase. In the imagery phase subjects were explicitly instructed to imagine hearing the continuation of the accent pattern, and not to use any other strategies, such as counting, imagining bouncing balls or tapping hands to maintain the rhythm. During the experiment, subjects were visually observed to control for hand, head or other body movements to make sure that no artefacts would influence classification.

To check whether the subjects did not lose track of the accenting pattern, a probe accent was sounded at the end of the sequence and the subjects had to answer the question whether this probe would have coincided with the accent in the pattern, if the accenting sound had not stopped playing. Probe accents were randomly placed on either accented or non-accented positions at the end of the sequence and this information was later used to check the subject's answers. Each next sequence was started with a button press, giving the subject the opportunity to control the interval between sequences, and the opportunity to move freely between sequences. However, during the sequences they were asked to sit still and avoid eye movements or blinks.

A block in the experiment consisted of 12 sequences of two-, three- and four-beat patterns, giving a total of 36 sequences in a block. The order of beat patterns in a block was randomized before the start of the experiment. With four of these blocks per subject we gathered roughly $12 \times 4 \times 5 = 240$ cycles of each imagery pattern and $12 \times 4 \times 3 = 144$ cycles of each perception pattern. Some of the cycles were rejected in further analyses, due to artefacts (see section 2.2).

2.2. *Preprocessing*

Bad channels were identified from the raw EEG signal for each trial with an algorithm sensitive to four properties. Initially, any channel with a dc offset exceeding 30 mV was marked as 'bad', as well as channels with a power exceeding 3500 μV^2 in the 50 Hz band (45–55 Hz) or a maximum derivative larger than 200 μV/sample. Horizontal and vertical EOG channels were band-pass filtered between 0.2 and 15 Hz and decorrelated from the EEG (Schlögl *et al* 2007), thus removing eye drifts or blinks if present. The raw EEG signal, originally sampled at 2048 Hz, was temporally down-sampled to a sampling frequency of 128 Hz. Additionally, as a fourth property for identification of bad channels, within-trial variance was computed and channels exceeding a variance of 2000 μV^2 were marked 'bad'. If—according to the four properties—more than 20% of the channels in a trial were bad, the trial was excluded from further analysis. Trials from a sequence with a wrong answer to the probe accent at the end of the sequence, occurring on average in 10.3% of the sequences, were also excluded. For the remaining trials, bad channels were reconstructed by interpolation from the remaining good channels with a spherical spline interpolation algorithm (Perrin *et al* 1989). The interpolation step assures a stable number of good channels for the classifiers to work with, while avoiding rejection of channels throughout the whole dataset when channels are only occasionally bad. Measures for bad channel identification are based on single trial data only, instead of all available data. This choice was motivated by the intention to use the same preprocessing pipeline in an online BCI system. Data were re-referenced to a common average reference (CAR) and linearly de-trended. The same preprocessing was used for all subsequent analyses.

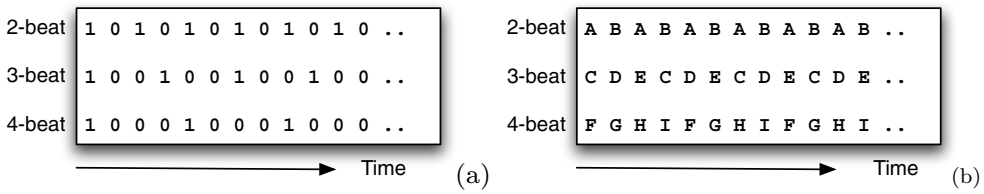

Figure 2. Two different ways of labelling the individual beats in a two-, three- and four-beat sequence are shown, corresponding to two different hypotheses on cognitive processing. Each character in a row represents a single beat within a specific beat pattern. Panel (a) shows the labelling when only a binary distinction between beats is made (labels 1 and 0 for accented and non-accented). This information is also relevant for sequence classification, since it can be interpreted as a 'codebook' that lists the possible 'code words' as rows. Panel (b) shows the labelling when distinguishing nine different classes of beats (labels A to I).

2.3. Single-beat classification

Data acquired during two-, three- and four-beat cycles from the stimulus sequences were split into individual beats, which will be called trials from now on (see figure 1). A time window from −50 ms to 450 ms was chosen around each pulse of the metronome, where time 0 ms corresponds to the time of occurrence of the pulse. Data collected during the first cycle of a beat pattern in the perception or imagery phase were excluded from analysis. After preprocessing, all trials were further processed using Fieldtrip (Oostenveld 2010). A high-pass filter with a 3 dB cutoff at 0.5 Hz and low-pass filter with a 3 dB cutoff at 15 Hz were applied (both filters were of sixth order Butterworth type). A more common low-pass cutoff at 40 Hz was also tried but only seemed to have a negative effect on classification performance, presumably due to additional noise without additional information about the classes. All remaining time points (64) per trial on all available EEG channels (64) were used as features for classification.

Single beats were classified with a regularized (L_2) logistic regression algorithm (Bishop 2006) applied directly to the pre-processed spatiotemporal data matrices. Since the brain's representation of subjective accents is not fully understood, two approaches to classification have been explored. The first approach makes a distinction between accented and non-accented beats only, requiring a binary classifier. The number of trials remaining after preprocessing of perception data was on average 264 (between 226 and 284) per subject for the accented class and 528 (between 452 and 567) for the non-accented class. For the imagery data on average 593 (between 509 and 638) trials were available for the accented and 1186 (between 935 and 1165) for the non-accented class. In the binary classification approach with this experimental design, many more trials are available for the non-accented condition than for the accented condition. In order to avoid a bias towards the non-accented class in the classifier's output, the classifier's loss function was weighted per class to compensate for the unbalanced training data, such that both classes become equally important (Bishop 2006).

The second approach assumes that each beat has a unique representation, depending on the metric pattern as well as the position of the beat within the pattern. Classification under this assumption requires a nine-class classifier. Figure 2(b) shows how the nine classes (labels A to I) are distributed over the two-, three- and four-beat patterns. The multi-class classifier was constructed from 36 binary classifiers, one for each possible pair of classes—also referred to as a sub-problem—on a 1-against-1 basis (Bishop 2006).

Figure 2 illustrates the order and class labels of single beats from the overarching beat patterns, for each of the two classification approaches. Figure 2(a) shows the accented and non-accented beats as '1' and '0', respectively. Figure 2(b) illustrates the nine different beats in the two-, three- and four-beat patterns. The total number of collected beat patterns led to about 88 perception trials (between 70 and 96, depending on the number of trials rejected in the preprocessing) and 187 imagery trials (between 140 and 216) for each of the nine classes for each subject. The number of trials per class is not structurally unbalanced in the nine-class approach.

To measure performance of the classifiers, double-nested cross-validation (Bishop 2006) was performed, using five inner folds for hyperparameter optimization and ten outer folds for the final performance evaluation. As a precaution against overfitting, folds were constructed with the aim of keeping the trials in each fold consecutive in time (also known as in-order cross-validation). The significance of the classifier performance was calculated using a *t*-test. Since test-sets in the binary classification approach can be unbalanced and since we want to enforce equal importance of classes, balanced classification rates (defined as the average of per-class performance) will be reported.

As shown by Vlek *et al* (at press) it is possible to classify imagery subjective rhythmization data with a classifier that was trained on the perception data. This method, which we will refer to as 'cross-condition classification', was also used here as a variation on the binary classification approach. It also distinguishes imagined accents from non-accents in a binary fashion, but is trained to do so on perception instead of imagery data. A tenfold cross-validation regime on the perception data was used to find the optimal regularization parameter, and a classifier using this regularization parameter was retrained on all available perception data. To avoid a structural preference of the retrained classifier for one of the classes in the new test set, calibration was performed by a restricted retraining of the classifier's bias and gain (see the work of Shenoy *et al* (2006)) on a random set of 200 trials of imagery data, while aiming for equal per-class performance. The trials used for calibration were not used for performance evaluation.

2.4. Sequence classification

Next to classification of single beats we explored sequence classification, which is a technique popular in P300 spellers (Hill *et al* 2008). EEG signals were sliced into individual beats, but classifier predictions on the slices were combined into predictions for a specific sequence of beats (see figure 1). The two-, three- and four-beat patterns could each be interpreted as a unique and cyclic sequence of accented and non-accented beats (see figure 2(a)). Given a classifier's prediction for an individual beat, a prediction for each sequence of beats can be made by means of a sequence classification algorithm. The algorithm is informed about the order of accented and non-accented beats in each beat pattern through a 'codebook' (see figure 2(a)). At each new beat, an underlying binary classifier will deliver probabilities for each of the single-beat classes (accented or non-accented). Taking into account the possible codes in the codebook and the probabilities of all previous trials, predictions on the level of beat pattern classes are updated. Assuming trial independence, the probability for the ith row of the codebook matrix C (denoted C_i) given the sequence of data $[X_1, X_2, \ldots]$ is defined by the product of trial predictions as described in equation (1):

$$\Pr(C_i|X_1, X_2, \ldots) = \prod_{n=1,2,\ldots} \Pr(C_{in}|X_n), \tag{1}$$

where $\Pr(C_{in}|X_n)$ represents the trial prediction for the nth element (column index) in the ith row of codebook C, given the data X_n at trial n.

The choice of a binary single-beat distinction, with classes accented and non-accented, was driven by the results of the single-beat classification (see section 2.3). The imagery data were sliced in exactly the same way as for single beats: −50 to 450 ms around stimulus onset. However, there was a difference from the previous procedure described in section 2.3, in that the order of beats as they had occurred in the experiment was preserved, and that data from the first cycle of each beat pattern were included. Preprocessing and classifier training were done as described in section 2.3. Performance of the sequence classifier was evaluated using test sets of (on average 13) imagery sequences left out of the training set. If bad trials, identified by the preprocessing pipeline, occurred somewhere in a sequence, class probabilities were set to chance level for this trial. In this way, bad trials are gracefully ignored, without negatively influencing the performance of the entire sequence. Additionally, sequence classification was performed on the basis of the single-beat classifier trained on perception data and tested on sequences of imagery data.

Since results of the sequencing method described above are limited by the number of available consecutive trials, a sequence simulation algorithm was developed. Using a Monte Carlo approach, this algorithm allows for extrapolation of the sequence classification performance curve by randomly sampling from the available data and applying the classifier. Similar to the sequence classification algorithm for real data (see equation (1)), the simulation algorithm as defined by equation (2) is based on a product of per-trial predictions:

$$\Pr(C_i) = \prod_{n=1,2,\ldots} \Pr(C_{in}|s(C_{in})). \tag{2}$$

The classifier is provided with Monte Carlo sampled data, denoted by $s(C_{in})$, instead of real data for computing per-trial predictions. The Monte Carlo sampling process is defined by equation (3), where $s(c)$ represents a random element taken from the set of trials tr with the true class c:

$$s(c) = \text{rand}(\{tr : \text{class}(tr) \in c\}) \tag{3}$$

The accuracy of the Monte Carlo simulation algorithm is computed as the average performance over 2000 simulated sequences per class.

2.5. Long-segment classification

To summarize the sequence classification approach, long segments of imagery data were explicitly broken down into individual beats and classified, after which classifier predictions were combined into sequence predictions. Explicit information about the time structure of the mental task, such as the time interval between beats and the order of beats in each possible pattern, is in that case provided to the algorithm. As an alternative, we have investigated how well a classifier can deal with longer data segments, when no such top-down information is provided. We rely on the classifier to learn any time structure beneficial to classification performance. For this, longer segments of data acquired during the imagery phase of the stimulus sequences were sliced and classified (see figure 1). A segment consists of five cycles of each beat pattern, resulting in 5 s of data for the two-beat pattern, 7.5 s for the three-beat and 10 s for the four-beat data after the start of the imagery phase. Since the classifier should work on equally sized data segments, these segments were all truncated to the shortest length of 5 s (corresponding to the five cycles of the two-beat pattern). In addition to the preprocessing, described in section 2.2, the data segments were filtered by a high-pass filter with a 3 dB cutoff at 0.5 Hz and by a low-pass filter with a 3 dB cutoff at 15 Hz. Both filters were of sixth order Butterworth type. The resulting time-domain data on all 64 channels were fed to a regularized (L_2) linear logistic-regression classifier. The three-class problem (with classes 2-, 3-, 4-beat) was addressed with a 1-against-1 style multi-class classifier. Classification performances were obtained by a leave-one-out regime (Bishop 2006), instead of cross-validation with ten outer folds, since this provided a more reliable estimate of the performance with the smaller number of data segments available. The average number of long-segment trials available per class was 44 (between 36 and 48).

2.6. Bit rate comparison

It is difficult to directly compare classification performances of all methods described, because of different numbers of output classes and different duration of the slices of data required for each method. A comparison can more easily be made through bit rate. Using the definition of bit rate by Wolpaw, as described in Kronegg *et al* (2003), the bit rate was computed for each of the classification methods. The bit rate is dependent on three variables: the number of classifications per second, the number of classes and the mean accuracy.

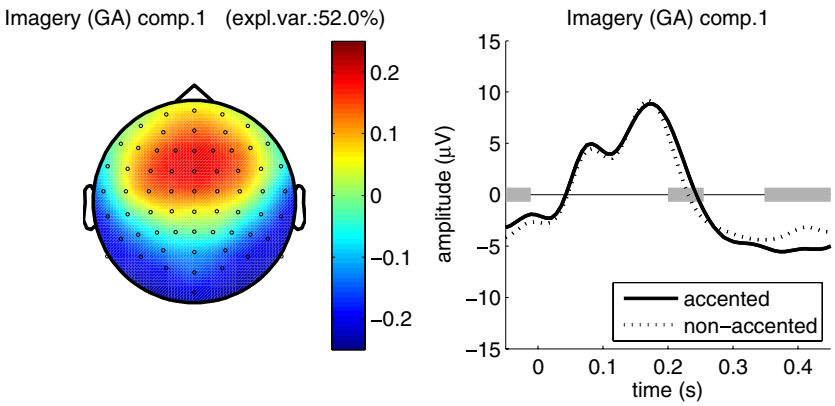

Figure 3. A characteristic of the signal relevant to discrimination of imagery accented and non-accented beats is shown. A grand average ERP over subjects was decomposed by means of PCA and the spatial distribution and corresponding time course are displayed for the strongest component (explaining 52.0% of the variance). Areas in the time course contributing to a significant ($p < 0.05$) difference between accented and non-accented trials are marked by grey bars on the time axis.

3. Results

3.1. Single-beat classification

A grand average ERP over imagery data of all subjects was computed, consisting of the signals of 64 channels as a function of time from −50 to 450 ms relative to each beat. A principal component analysis (PCA) was used to spatially decompose this grand average. Figure 3 shows the spatial distribution and corresponding time course per class of the first component, i.e. the principal component with the largest eigenvalue, explaining 52.0% of the variance. Figure 3 shows a characteristic of the signals for accented and non-accented beats that are later provided to a classifier. Differences between accented and non-accented beats in the time courses of the first component were statistically tested using a cluster randomization test, a non-parametrical statistical test designed to deal with the multiple comparison problem present in EEG data, using biophysiologically motivated constraints to increase the sensitivity of the test (Maris 2004, Maris and Oostenveld 2007). Areas in the time courses contributing to significant differences ($p < 0.05$) are marked in grey on the time axis in figure 3. The main difference between accented and non-accented beats appears to be in the late part of the auditory evoked potential (AEP) (Burkard *et al* 2007) around 200 ms after stimulus onset, and in a negative deflection starting at 250 ms. Similar neurophysiological effects of an identical task are discussed in more detail in Schaefer *et al* (2010a) and in Vlek *et al* (at press). These effects follow a fronto-central distribution on the scalp, similar to scalp distributions observed for distinction of musical stimuli (Schaefer *et al* 2010b).

For the binary single-beat approach to the decoding of subjective rhythmization from brain signals, results are shown in figure 4, where scaling of the *y*-axis ranges from chance level to maximum performance (100% correct). This scaling convention is used consistently for all performance figures. With the binary classification approach (figure 4(a)), distinguishing perceived accented from non-accented beats, an average performance of 68.8% (SD = 5.6%) is achieved. The best subject (S6) reached 74.3%. Of more interest to BCI is the classification of imagined accents, where the stimulus does not carry class information. Here an average classification rate of 60.4% (SD = 4.2%) was achieved over subjects, while the best subject (S4) reached an accuracy of 66.8%. All subjects performed significantly ($p < 0.001$) above the chance level of 0.5.

Alternatively, with the multi-class approach applied to imagery data (figure 4(b)), we distinguish between nine beat classes. An average performance of 15.9% (SD = 2.8%) was achieved, while a performance of 20.6% was obtained for the best subject (S7). All subjects performed significantly ($p < 0.05$) above the chance level of 1/9. Given the difference in chance level between the binary (1/2) and nine-class (1/9) approach, a comparison of performances is made by means of bit rates. Performance of the binary classification translates into an average bit rate of 4.4 bits min^{-1} (in the range of 1.0–10.0 bits min^{-1}) over all subjects, while the multi-class approach yields an average rate of 2.8 bits min^{-1} (in the range of 0.4–6.5 bits min^{-1}). Based on this finding, it was decided to construct the sequence classification algorithm (see section 2.4) from a binary classifier. Its performance will be described in the following section (section 3.2).

Results of the cross-condition classifier can also be seen in figure 4(a). An average performance of 58.2% (SD = 4.8%) over subjects and 66.4% for the best subject (S4) is achieved. Results are significantly ($p < 0.01$) above chance for seven out of ten subjects.

Beyond visual inspection of the subjects' behaviour during the experiment, a separate classification of the EOG channels was performed. In this way, we were able to verify that potentially remaining eye-artefacts (not completely removed by our decorrelation pre-processing step) were not turning into artefactual sources of classifier performance. None of the subjects performed significantly ($p < 0.01$) above chance when using only EOG channels.

3.2. Sequence classification

Scaling up the recognition of single beats to the level where sequences of beats can be decoded, we obtained the following

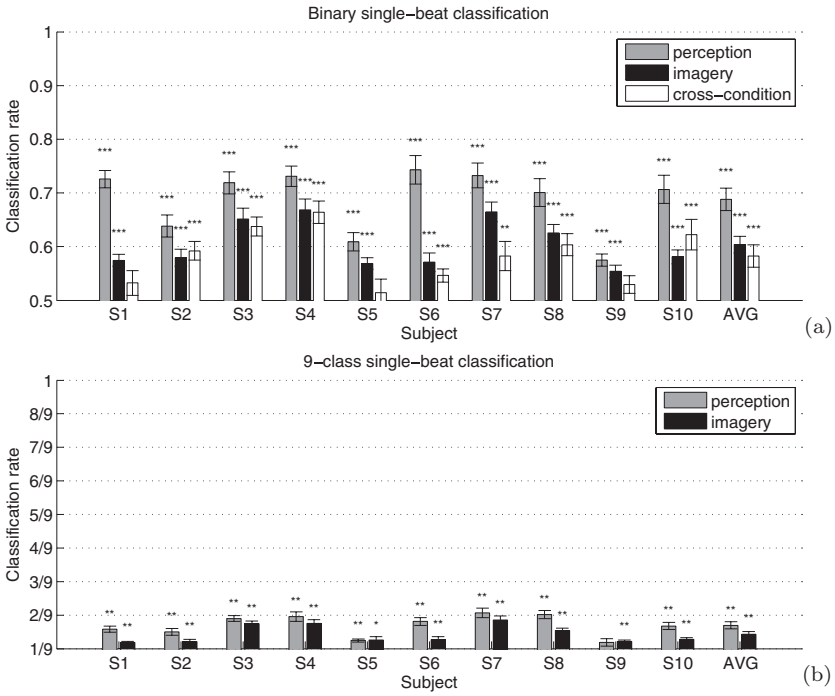

Figure 4. Panel (a) shows balanced classification rates for the binary approach to single beat classification of perception data, imagery data and with the method of training a classifier with perception, and testing it on imagery data (cross-condition). Panel (b) shows performances of the nine-class approach for the perception and imagery condition. The average over subjects is labelled AVG. The error bars indicate the standard error within the tenfold cross-validation. The significance level is indicated by * for $p < 0.05$, ** for $p < 0.01$ or *** for $p < 0.001$, based on a *t*-test.

results. Classification of two-, three- and four-beat sequences resulted in performances shown in figure 5(a), yielding an average accuracy of 48.8% (SD = 8.7%) over subjects and 63.2% for the best subject (S7). A more detailed view of the sequence classification results for one of the best (S3) and worst (S9) subjects can be found in figure 5(b). This figure shows how classification accuracy on real data (both solid lines) increases with each additional trial. Note that it is not until the third trial that the performance starts to increase above chance level. Sequencing on the basis of a binary classifier, trained with perception data, resulted in an average accuracy of 43.7% (SD = 5.2%) and 50.0% for the best subject (S3).

Figure 5(b) also shows the result of the Monte Carlo simulation algorithm for sequence classification for one of the best (S3) and worst (S9) subjects. After ten trials, the performance of the simulation algorithm deviates on average 5.4% from the performance on the real data. The Monte Carlo simulation approach will be used in section 3.4 for extrapolation and exploration of other coding types.

3.3. Long-segment classification

In contrast to the sequencing approach, where top-down information about the time structure in the neural correlates is provided to the algorithms, the classification method for long segments is not provided with such information. Results of the classification of long segments of data are visualized in figure 6. Seven out of ten subjects perform significantly ($p < 0.05$) above the chance level of 1/3. Over all ten subjects an average accuracy of 44.9% (SD = 7.6%) is achieved, with a maximum of 60.7% for the best subject (S7).

3.4. Bit rate comparison

For comparison of the reported results on imagery data, classification rates were translated into bit rates, visualized in figure 7. With an average bit rate of 4.4 bits min^{-1} over subjects, the single-beat classification paradigm clearly performs the best. From the classification methods capable of dealing with longer segments or sequences of data, the sequence classification method achieved the highest bit rate, with an average of 1.1 bits min^{-1}, compared to 0.7 bits min^{-1} for the long-segment classification method.

Using the Monte Carlo approach, simulations were performed on sequence classification. First, we use the simulation for extrapolation of the sequences that were otherwise restricted to the length of ten beats. Second, we will illustrate the effect of different types of coding. These conditions were not present in the experimental design, but simulating them may help in predicting the future possibilities of the subjective accenting paradigm. Simulations were performed per subject, since the Monte Carlo method is data driven and thus subject specific. By computing the average bit rate curve over subjects, corresponding to simulated sequence classification of the two-, three- and four-beat patterns over 40 trials, we found (figure 8, solid line) that the optimum bit rate is reached within five trials.

Furthermore, simulation allows us to illustrate the expected performance with patterns other than the standard

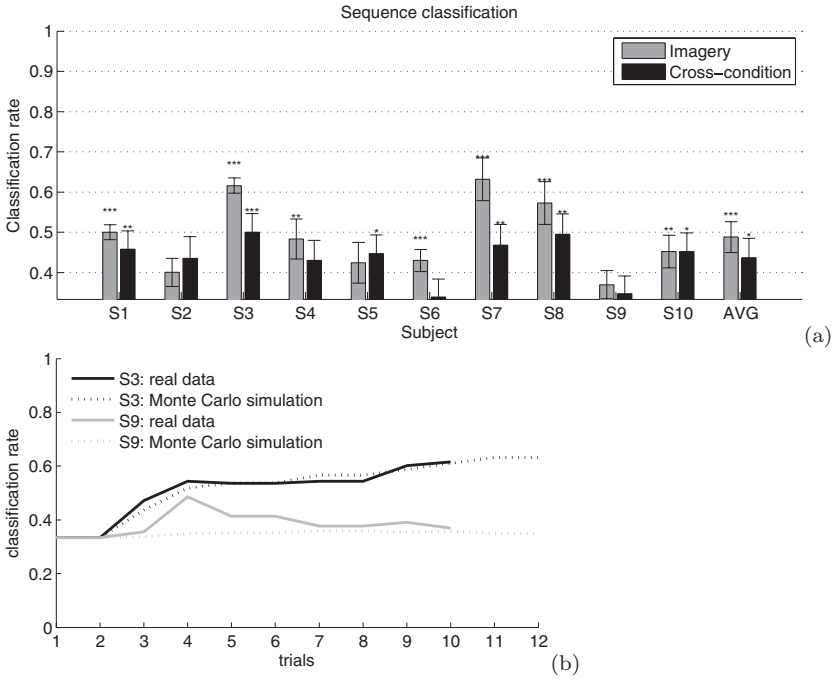

Figure 5. Panel (a) shows performances of the sequence classification of two-, three- and four-beat patterns based on ten trials (5 s) of data. For a classifier trained and tested on imagery data, the error bars indicate the standard error within the tenfold cross-validation. For the cross-condition classification, a classifier was trained on the perception data, allowing all imagery data to be used for performance evaluation. For this regime, error bars thus indicate the standard error over all imagery trials. The significance level above the chance level of 1/3 is indicated by * for $p < 0.05$, ** for $p < 0.01$ or *** for $p < 0.001$, based on a *t*-test. The average over subjects is labelled AVG. Panel (b) gives an overview of sequence classification and Monte Carlo simulation performances, as a function of the number of trials. Performances of one of the best (S3) and worst (S9) subjects are shown.

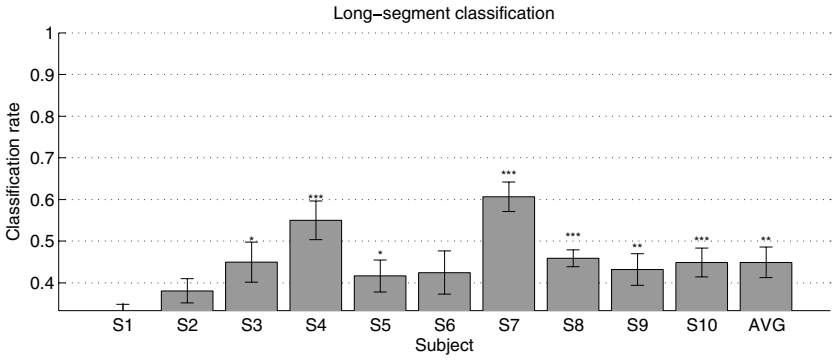

Figure 6. Performance of the classifier on long segments of time-domain data is shown. Subject S1 is performing below chance and is not visible with current scaling. The average over subjects is labelled AVG. The significance level above chance (1/3) is indicated by * for $p < 0.05$, ** for $p < 0.01$ or *** for p < 0.001, based on a *t*-test.

two-, three- and four-beat. A positive shift in the average bit rate curve can be observed when using all possible phase-shifted versions in addition to the three standard patterns (for instance a two-beat pattern started at the non-accented beat, as a separate class). In a BCI system this would allow the user to encode his or her intention with nine instead of three classes. A higher bit rate is expected when selecting more optimal codes, such as cyclic codes of four beats with a Hamming distance of at least two bits to each other (being 1001, 1010, 0101, 0110, 1100, 0011 and excluding 0000 and 1111). Such a type of coding would result in a six-class output of the BCI system, achieving an optimal average bit rate of almost 3 bits min^{-1}.

4. Discussion

We have shown that it is possible to decode subjective accents voluntarily imposed on an auditory metronome from measured brain activity. A binary approach to classification of single beats, distinguishing accented and non-accented beats, gave a higher bit rate than a nine-class approach. Although time frequency features have also been considered, features in the time domain were used and reported in this study, as they seemed to contain more class-relevant information. Also the gamma-band activity described by Snyder and Large (2005) was not found consistently and strongly enough to pursue further investigation for BCI application. The results of

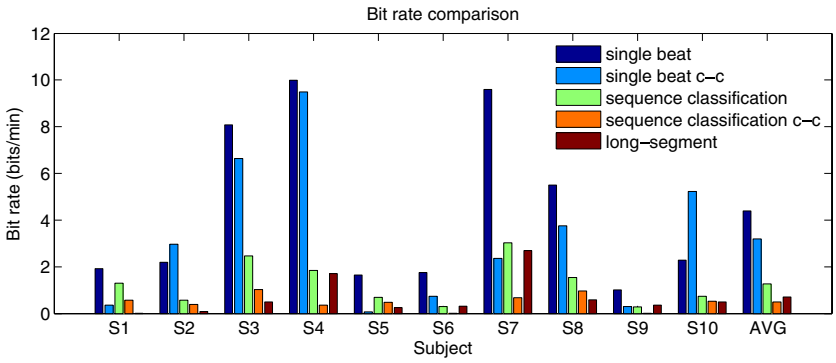

Figure 7. Performance of the different classification approaches for imagery data, described in sections 2.3, 2.4 and 2.5, was translated into bit rate and visualized for all ten subjects as well as the average over subjects. Cross-condition classification is abbreviated as 'c–c'. The average over subjects is labelled AVG.

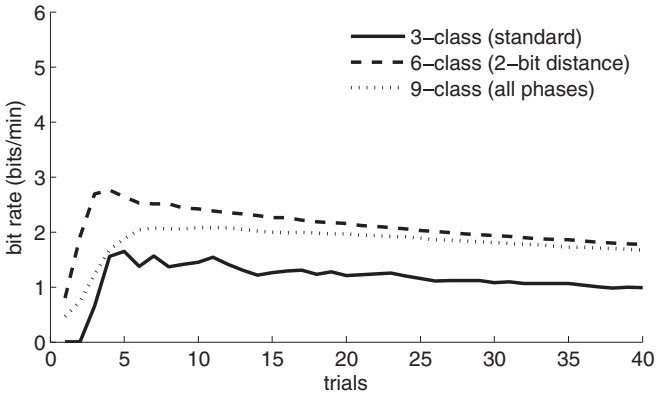

Figure 8. Using a Monte Carlo approach, simulations were performed with the sequence classification algorithm. Performances were translated into bit rates and averaged over subjects. The bit rate curve of the standard three-class situation shows how the optimum bit rate is met within five trials. Furthermore, the simulations illustrate how different types of coding are beneficial to the expected bit rates, such as nine-class coding using all phase-shifted versions in addition to the standard beat patterns, or six-class coding using cyclic patterns of four beats with a Hamming distance of two bits.

single-beat classification translate into an average bit rate of 4.4 bits min^{-1}. This is better than the average bit rate of 2 bits min^{-1} reported for an auditory P300 BCI (Klobassa *et al* 2009) and close to bit rates of approximately 5 bits min^{-1} reported for an auditory BCI based on the concept of auditory stream segregation (Kanoh *et al* 2010). It has to be considered that translation of the classification performance to bit rate in the present study merely provides an estimate of the potential speed of an online BCI relying on this paradigm. The translation does not take into account the influence on bit rate of factors such as the inter-trial interval required for the subject to switch between different subjective accenting patterns or the time required to establish these patterns. These factors will be investigated in future research. Independent classification of EOG signals did not result in a performance significantly above chance for any of the subjects and allows us to conclude that possible eye-artefacts were not boosting classifier performance.

We have also shown that it is possible to decode subjective accents with a cross-condition classification method. This approach does not outperform conventional classification of subjective accents but is nonetheless considered useful to BCI. Using perception of auditory accents during a BCI training regime takes away confusion for the user about what mental task to perform. This simplification of the training regime will contribute to correct execution of the (now well-defined) mental task. Once a classifier is trained on data gathered during this training regime, feedback could be used to guide the subject to an optimal performance of the less well-defined task of imagined accenting.

Although single imagery accented and non-accented beats were sliced and classified successfully in this study, it has to be taken into account that these trials originate from data of the overarching beat patterns. The brain response leading to successful classification of a single accent may not exist without the presence of the overarching beat pattern. In other words, a subject may not be able to voluntarily decide per beat whether it will be accented or non-accented. Since this is crucial for BCI application of the single-beat approach, the assumption that subjects can do this needs to be validated in further research.

In the case that this assumption were not valid, an alternative is available through classification of sequences of beats. This method allows for the combination of any number trials required to meet the desired accuracy of the BCI system. In a BCI application beat patterns instead of single beats could then be used to encode the user's intention. We have shown that these patterns can be successfully decoded with a sequence classification algorithm. With this algorithm, the achieved bit rates were considerably lower than for classification of single beats. This can be explained by suboptimal coding of the beat patterns. As shown in the codebook (figure 2(a)), the first two beats of all beat patterns are identical. Inherently, probabilities for the two-, three- and four-beat patterns are equal for these trials, when using a distinction of only accented and non-accented classes. A similar ambiguity between beat patterns occurs at other points in the sequence. This effect is reflected by the plateaus in the generally rising performance functions (see figure 5(b)). While the theoretical maximum bit rate of the single-beat classification approach at maximum confidence is $\frac{1}{0.5/60} = 120$ bits min^{-1}, the theoretical maximum of

the sequence classification of beat patterns has a ceiling of $\frac{\log_2(3)}{4 \cdot 0.5/60} \approx 47.5$ bits min^{-1}.

It has been mentioned that application of the single-beat classification approach for BCI requires certain assumptions to be true, while the sequence classification approach is free of such assumptions, but a tradeoff in bit rate is inherent. By means of the Monte Carlo simulations, we were able to illustrate that alternative coding types can be found that boost the expected bit rate of the BCI system. These alternatives also have their own assumptions on what mental task subjects are able to perform. In both the nine- and six-class coding types, the cyclic nature of the patterns persists, but it is assumed that subjects can voluntarily imagine a phase-shifted pattern or imagine a more complex pattern consisting of multiple accented beats. Even in the standard three-class coding, optimization may be possible by dynamically stopping the sequence classification process when the desired confidence of a prediction is reached. However, for rapid communication, this requires the subject to be able to stop imagining a pattern on demand and switch to the next one. For BCI applications of these types of coding, validation is first required. However, given the expected increase in the bit rate, we believe this may be a fruitful direction for future research.

Comparison of the performance of the Monte Carlo simulation algorithm and sequence classification on real data shows a relatively good fit. However, with this simulation algorithm trial independence is assumed, which may not be completely realistic given the brain's properties. This is a possible reason for the Monte Carlo results to slightly deviate from classification results on the real data.

For most subjects in this study, the sequence classification method yields slightly higher bit rates than the long segment classification method (see figure 7). This seems to suggest that exploitation of the time structure present in the neural correlates of the mental task is beneficial to classification performance. Finally, we conclude that the successful single-trial classification of both single beats and sequences of beats suggests that subjective rhythmization is a feasible paradigm for an auditory BCI. The paradigm can be made easier for the subject by using a cross-condition classification approach, such that the user's task during the training part of the BCI merely consists of the perception of rhythmic stimuli.

Acknowledgments

The authors gratefully acknowledge the support of the BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science.

References

Bishop C (ed.) 2006 *Pattern Recognition and Machine Learning* (Berlin: Springer)

Brochard R, Abecasis D, Potter D, Ragot R and Drake C 2003 The 'ticktock' of our internal clock: direct brain evidence of subjective accents in isochronous sequences *Psychol. Sci.* **14** 362–6

Burkard R, Don M and Eggermont J (eds) 2007 *Auditory Evoked Potentials* (Baltimore, MD: Lippincott Williams and Wilkins)

Desain P and Honing H 2003 Single trial ERP allows detection of perceived and imagined rhythm *Proc. RENCON Workshop: Int. Joint Conf. on Artificial Intelligence (IJCAI)* pp 1–4

Dornhege G, Millán J d R, Hinterberger T, McFarland D and Müller K-R (eds) 2007 *Toward Brain–Computer Interfacing* (Cambridge, MA: MIT Press)

Farwell L and Donchin E 1988 Talking off the top of your head: toward a mental prosthesis utilizing event-related potentials *Electroencephalogr. Clin. Neurophysiol.* **70** 510–23

Fraise P 1982 *The Psychology of Music* (New York: Academic) pp 149–80 (chapter: rhythm and tempo)

Furdea A, Halder S, Krusienski D, Bross D, Nijboer F, Birbaumer N and Kuebler A 2009 An auditory oddball (p300) spelling system for brain–computer interfaces *Psychophysiology* **46** 617–25

Gerven M V *et al* 2009 The brain–computer interface cycle *J. Neural Eng.* **6** 1–10

Hill J, Farquhar J, Martens S, Biessmann F and Schölkopf B 2008 Effects of stimulus type and of error-correcting code design on BCI speller performance *Advances in Neural Information Processing Systems* (Red Hook, NY: Curran) vol 21 pp 665–72

Hill N, Lal T, Bierig K, Birbaumer N and Schölkopf B 2004 Attentional modulation of auditory event-related potentials in a brain–computer interface *IEEE Int. Workshop on Biomedical Circuits and Systems* pp 1–4

Iversen J, Repp B and Patel A 2009 Top–down control of rhythm perception modulates early auditory responses *Ann. NY Acad. Sci.* **1169** 58–73

Jongsma M, Eichele T, Quiroga R Q, Jenks K, Desain P, Honing H and van Rijn C 2005 Expectancy effects on omission evoked potentials in musicians and non-musicians *Psychophysiology* **42** 191–201

Kanoh S, Miyamoto K and Yoshinobu T 2010 A brain–computer interface (BCI) system based on auditory stream segregation *J. Biomech. Sci. Eng.* **5** 32–40

Kelso J (ed.) 1982 *Human Motor Behavior* (Hillsdale, NJ: Lawrence Erlbaum Associates)

Klobassa D, Vaughan T, Brunner P, Schwartz N, Wolpaw J, Neuper C and Sellers E 2009 Toward a high-throughput auditory p300-based brain–computer interface *Clin. Neurophysiol.* **120** 1252–61

Kronegg J, Alecu T and Pun T 2003 Information theoretic bit-rate optimization for average trial protocol brain computer interfaces *HCI International, 10th Int. Conf. on Human–Computer Interaction (Crete)* (Hillsdale, NJ: Lawrence Erlbaum Associates) pp 1–5

London J 2004 *Hearing in Time: Psychological Aspects of Musical Meter* (Oxford: Oxford University Press)

Maris E 2004 Randomization test for ERP topographies and whole spatiotemporal data matrices *Psychophysiology* **41** 142–51

Maris E and Oostenveld R 2007 Nonparametric testing of EEG- and MEG-data *J. Neurosci. Methods* **164** 177–90

Michon J and Jackson J 1985 *Time, Mind and Behavior* (Berlin: Springer)

Müller-Putz G, Scherer R, Brauneis C and Pfurtscheller G 2005 Steady-state visual evoked potential (SSVEP)-based communication: impact of harmonic frequency components *J. Neural Eng.* **2** 123–30

Nijboer F, Furdea A, Gunst I, Mellinger J, McFarland D, Birbaumer N and Kübler A 2008 An auditory brain–computer interface (BCI) *J. Neurosci. Methods* **167** 43–50

Oostenveld R 2010 FieldTrip:, a Matlab software toolbox for MEG and EEG analysis http://fieldtrip.fcdonders.nl/

Perrin F, Pernier J, Bertrand O and Echallier J 1989 Spherical splines for scalp potential and current mapping *Electroencephalogr. Clin. Neurophysiol.* **72** 184–7

Pfurtscheller G, Brunner C, Schlögl A and da Silva F L 2006 Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks *NeuroImage* **31** 153–9

Pfurtscheller G, Neuper C, Flotzinger D and Pregenzer M 1997 EEG-based discrimination between imagination of right and left hand movement *Electroencephalogr. Clin. Neurophysiol.* **103** 642–51

Pham M, Hinterberger T, Neumann N, Kübler A, Hofmayer N, Grether A, Wilhelm B, Vatine J and Birbaumer N 2005 An auditory brain–computer interface based on the self-regulation of slow cortical potentials *Neurorehabil. Neural Repair* **19** 206–18

Regan D 1977 Steady-state evoked potentials *J. Opt. Soc. Am.* **67** 1475–89

Schaefer R, Farquhar J, Blokland Y, Sadakata M and Desain P 2010 Name that tune: decoding music from the listening brain *NeuroImage* at press doi:10.1016/j.neuroimage.2010.05.084

Schaefer R, Vlek R and Desain P 2010 Decomposing rhythm processing: electroencephalography of perceived and self-imposed rhythmic patterns *Psychol. Res.* doi:10.1007/s00426-010-0293-4

Schlögl A, Keinrath C, Zimmermann D, Scherer R, Leeb R and Pfurtscheller G 2007 A fully automated correction method of EOG artifacts in EEG recordings *Clin. Neurophysiol.* **118** 98–104

Schreuder M, Blankertz B and Tangermann M 2010 A new auditory multi-class brain–computer interface paradigm: spatial hearing as an informative cue *PLoS ONE* **5** 1–14

Shenoy P, Krauledat M, Blankertz B, Rao R and Müller K-R 2006 Towards adaptive classification for BCI *J. Neural Eng.* **3** R13–23

Snyder J and Large E 2005 Gamma-band activity reflects the metric structure of rhythmic tone sequences *Cogn. Brain Res.* **24** 117–26

Vatikiotis-Bateson E and Kelso J 1993 Rhythm type and articulatory dynamics in English, French and Japanese *J. Phonetics* **21** 231–65

Vlek R, Schaefer R, Gielen C, Farquhar J and Desain P Shared mechanisms in perception and imagery of auditory accents *Clin. Neurophysiol.* at press

Zanto T, Snyder J and Large E 2006 Neural correlates of rhythmic expectancy *Adv. Cogn. Psychol.* **2** 221–31