Contents lists available at ScienceDirect

Clinical Neurophysiology

Shared mechanisms in perception and imagery of auditory accents

R.J. Vlek [a,*], R.S. Schaefer [a], C.C.A.M. Gielen [b], J.D.R. Farquhar [a], P. Desain [a]

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

ARTICLE INFO

HIGHLIGHTS

- Single-trial detection of imagined auditory accents from EEG.
- Support for shared mechanisms in auditory perception and imagery.
- A machine-learning technique for neuroscience.

ABSTRACT

Objective: An auditory rhythm can be perceived as a sequence of accented (loud) and non-accented (soft) beats or it can be imagined. Subjective rhythmization refers to the induction of accenting patterns during the presentation of identical auditory pulses at an isochronous rate. It can be an automatic process, but it can also be voluntarily controlled. We investigated whether imagined accents can be decoded from brain signals on a single-trial basis, and if there is information shared between perception and imagery in the contrast of accents and non-accents.

Methods: Ten subjects perceived and imagined three different metric patterns (two-, three-, and four-beat) superimposed on a steady metronome while electroencephalography (EEG) measurements were made. Shared information between perception and imagery EEG is investigated by means of principal component analysis and by means of single-trial classification.

Results: Classification of accented from non-accented beats was possible with an average accuracy of 70% for perception and 61% for imagery data. Cross-condition classification yielded significant performance above chance level for a classifier trained on perception and tested on imagery data (up to 66%), and vice versa (up to 60%).

Conclusions: Results show that detection of imagined accents is possible and reveal similarity in brain signatures relevant to distinction of accents from non-accents in perception and imagery.

Significance: Our results support the idea of shared mechanisms in perception and imagery for auditory processing. This is relevant for a number of clinical settings, most notably by elucidating the basic mechanisms of rhythmic auditory cuing paradigms, e.g. as used in motor rehabilitation or therapy for Parkinson's disease. As a novel Brain–Computer Interface (BCI) paradigm, our results imply a reduction of the necessary BCI training in healthy subjects and in patients.

1. Introduction

Our sense for auditory rhythms, such as a pattern where every first beat out of two, three or four beats is accented, is generally well-developed (Michon and Jackson, 1985). These rhythmic 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 this sense for rhythm 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 use of auditory rhythms or cues has also become increasingly popular in clinical environments for rehabilitation purposes. Motor rehabilitation has shown an increase in effect with the addition of external auditory cues, especially for gait rehabilitation (as in Roerdink et al., 2007), but also in bilateral arm training (see Latimer et al., 2010). Apparently the rhythmic processing adds something to the rehabilitation process. This is also seen in other types of time-structured therapies; used in, for example, dyslexia (Overy, 2003), aphasia (Belin et al., 1996), Parkinson's disease (McIntosh et al., 1997; Willems et al., 2007), as well as a number of cognitive functions (Thaut, 2010). Interestingly, it is not necessary for the rhythmic cue to be externally presented, as patients have also been able to increase the efficacy of their rehabilitation while moving to their own internal or imagined rhythm to improve gait (Schauer

* Corresponding author. Tel.: +31 0 24 3615458.
E-mail address: r.vlek@donders.ru.nl (R.J. Vlek).

and Mauritz, 2003). The ease at which such imagined rhythms can take place is nicely demonstrated by the so-called clock illusion or 'tick–tock' effect (Brochard et al., 2003). When a series of isochronous and equal sounding pulses is presented, such as the sound of a clock ('tick–tick–tick–tick...'), the percept of a rhythmic pattern is usually automatically induced, consisting of subjectively added accents on every second beat ('tick–tock–tick–tock...'). The mechanism of the brain inducing these accents is known as subjective rhythmization (Bolton, 1894; Fraise, 1982; London, 2004). As the mechanism of rhythm processing is not fully understood yet, we investigate the electrophysiology of simple rhythm processing, both with externally presented and internally generated accents.

Several studies have looked into the perception of metric patterns and stimulus-induced responses in electroencephalography (EEG). These studies have shown, that both the perception of metric patterns (Snyder and Large, 2005) as well as the expectation of an accent is reflected in EEG-activity (Zanto et al., 2006; Jongsma et al., 2005; Snyder and Large, 2005). Brochard et al. (2003) found that, for loudness deviations in a steady pulse train, subjects automatically exhibited different neuronal responses to deviants in even and odd positions, reflecting binary chunking. In a recent study Snyder and Large (2005) reported that (non-phase-locked) gamma-band activity in EEG can reflect the metric structure of the stimulus and that at an omission of a stimulus this activity 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 describe an effect in the upper beta-band of magnetoencephalography (MEG) measurements at subjectively accented versus non-accented tones. Studies investigating auditory imagery of rhythms or accents with EEG are scarce (but for exceptions, see Desain and Honing, 2003; Schaefer et al., 2011). A recent study by Cebrian et al. (2010) investigated the effect of auditory imagery on the N100 component of the auditory event related potential (ERP) evoked by a target tone following a sequence of imagined tones. They reported a correlation between the N1 (a.k.a. N100) amplitude and the vividness of imagery, converging towards identical N1 amplitudes for perception and extremely vivid imagery.

The relationship between mental imagery and perception and any similarities between neuronal structures involved in these processes, have been studied for different sensory modalities. Similarly, the relationship between imagery and actual motor activity has been studied. For a comprehensive overview on these relationships, see Kosslyn et al. (2001). Very similar neural activation patterns have been reported for actual and imagined movement tasks, in terms of mu and beta-band desynchronization over sensorimotor cortex (McFarland et al., 2000; Munzert et al., 2009). Strong support has also been reported for shared mechanisms in visual perception and visual imagery tasks. A study by Kosslyn et al. (1995) showed that during visual imagery the primary visual cortex is activated.

Compared to the visual and motor domains, the number of studies focussing on the relationship between auditory perception and imagery is relatively small. Support for shared mechanisms in auditory perception and imagery comes from behavioral (e.g. Farah and Smith, 1983; Halpern et al., 2004) as well as clinical angles (e.g. Kasai et al., 1999; Shinosaki et al., 2003). Zatorre et al. (1996) reported evidence from a positron emission tomography (PET) study for activation of parts of the auditory cortex during perception as well as during imagination of music. A more recent fMRI study by Kraemer et al. (2005) reported activation of secondary and primary auditory cortices during silent gaps in familiar tunes, where subjects reported the experience of continuation of the tune in imagery. However, support for the idea of shared mechanisms is predominantly found in studies concerning timbre or pitch aspects

of music (for an overview, see Halpern, 2001; Hubbard, 2010). Several studies have identified that timbre and pitch aspects are represented in our 'auditory mental image', but interestingly, for loudness aspects this has never been shown (Hubbard, 2010). The idea of shared mechanisms in rhythm processing is supported by results of a recent study by Schaefer et al. (2011), where overlap was found between ERP responses to events in perceived and imagined rhythms.

Recent developments in single-trial multivariate decoding of EEG signals, generally carried out in the context of brain–computer interface research (Dornhege et al., 2007; Gerven et al., 2009), can also be used to uncover patterns of brain activity that were previously not detectable. Single-trial multivariate decoding methods are often specialized in dealing with inter-trial variance and outliers, such that an optimal generalization and detection of the effect is possible. The aim of this study was to investigate whether it is possible to decode auditory accents from brain signals on a single-trial level, in both perceived and imagined auditory accenting patterns. Furthermore, we test the hypothesis that similar brain structures are involved in perception and imagination of accenting patterns, superimposed on a train of auditory beats. In a perception condition, subjects listened to a stimulus where the accents were physically different from the non-accents. In an imagery condition, subjects were listening to identical stimuli without accents, while they were instructed to imagine the accents. As pointed out by Hubbard (2010), a common problem in imagery studies is the lack of control for the process of imagery. In our study a behavioral task at the end of each imagery sequence guarantees a check on imagery processes. Differences between accented and non-accented beats were found with a principal component analysis and by means of single-trial classification, hereby expanding on the work previously reported with different data in a similar experimental design (Schaefer et al., 2011). Classification rates on both imagined and perceived accents are reported and comparison and interpretation is done for the discriminative signal properties in both experimental conditions. We hypothesize that similar brain structures are involved in imagery and perception of auditory accents, and test this hypothesis by classification of data with a cross-condition classification approach. This method allows to search for information shared between the conditions in the contrast of accented and non-accented beats on a single-trial level.

2. Experiment and analysis

2.1. Experimental design and data acquisition

Ten subjects, five females and five males, aged between 22 and 34 years, participated in this study. One subject had a professional musical training, and six participants actively play a musical instrument. 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 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
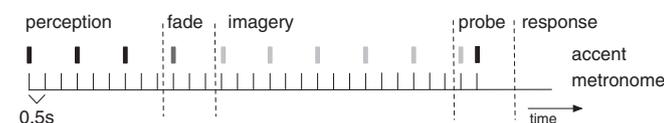
of recording. Simultaneously with the EEG, an electro-oculogram (EOG) was made to exclude eye movements as a possible source of information during EEG classification. Two pairs of auxiliary electrodes were placed. One pair was positioned above and below the left eye to measure eye movements in vertical 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 Fig. 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. 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., 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 Fig. 1, showing a sequence of three-beat patterns.

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 in the range 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 measures, which is indicated as the perception phase in Fig. 1. Subsequently, the pattern was played for one measure during the fade phase followed by five measures containing only the metronome, called 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 artifacts 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. 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 minimize any eye movements and eye 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 $12 \times 4 \times 5 = 240$ 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 artifacts (see Section 2.2).

2.2. Preprocessing

The raw EEG signal was sliced in chunks of data around the markers indicating the presentation of a metronome tick. This means that the two-, three- and four-beat cycles from the stimulus sequences are split into individual beats. A time window of -50 to 450 ms was chosen around each metronome tick where time 0 ms corresponds to the time of the tick. The time window was chosen to start 50 ms before each metronome onset, to capture possible anticipatory responses to the coming event. These data segments of 500 ms will be called trials. Bad channels were identified for each trial individually 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 exceeding 3500 μV^2 of power in the 50 Hz band (45–55 Hz) or with a maximum derivative exceeding 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 EEG signal, originally sampled at 2048 Hz, was temporally down sampled to a sample 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 as 'bad'. If – according to the four criteria – 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, were also excluded. Trials coming from the first cycle of the perception or imagery phase of a sequence were rejected to avoid possible transient effects. For the remaining trials, bad channels were reconstructed by interpolation from the remaining good channels with a spherical spline interpolation algorithm (Perrin

Table 1
The exact number of trials per subject, after artifact rejection and merging accented and non-accented beats from different beat patterns for each condition. In the perception condition 144 cycles of each pattern were collected per subjected, but for analysis the first cycle is not used. The remaining 96 cycles result in a maximum of $96 \times 3 = 288$ accented and $96 \times 6 = 576$ non-accented trials, as each two-, three-, and four-beat pattern contains one accented beat and one, two or three non-accented beats, respectively. Similarly, the maximum number of trials in imagery is $192 \times 3 = 576$ accented and $192 \times 6 = 1152$ non-accented trials. Due to the randomized position of the probe-accents a small amount of additional trials may be available for the five imagery cycles (e.g. see the additional accented beat available in Fig. 1). If available, these trials were used for analysis and explain why the previously computed maximum is sometimes exceeded.

Subject	Perception		Imagery	
	Accented	Non-accented	Accented	Non-accented
S1	284	564	638	1162
S2	226	452	509	935
S3	276	558	626	1150
S4	260	535	590	1115
S5	276	556	618	1143
S6	249	502	546	1006
S7	243	471	549	977
S8	266	514	601	1061
S9	281	567	632	1165
S10	278	559	620	1149

Fig. 1. The structure of a single sequence in the experiment is shown, in this case for a three-beat pattern. The sequence started with three repetitions of a ternary metric pattern (perception phase), followed by one repetition (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 repetitions (imagery phase). At the end, an accented beat was presented to test whether the subject maintained the correct rhythm.

et al., 1989). The exact number of trials after artifact rejection and merging from different beat patterns is shown in Table 1 for each condition. For these trials the average number of bad channels that had to be reconstructed by interpolation was 1.5, with a maximum of 12 bad channels per trial. The interpolation step assures a stable number of good channels for the classifiers to work with, while avoiding rejection of channels throughout the whole data-set when channels are only occasionally bad. The remaining trials were re-referenced to a common average reference (CAR) and linearly de-trended. Trials were further processed using Fieldtrip (http://www.ru.nl/fcdonders/fieldtrip/) functions. A high pass filter with a 3-dB cut-off at 0.5 Hz and low pass filter with a 3-dB cut-off at 15 Hz were applied. Both filters were of a sixth order Butterworth type. A more common low-pass cut-off at 40 Hz was also tried, but only had a negative effect on classification performance (see Section 2.4), presumably due to additional noise without additional information about the classes.

2.3. Principal component analysis

For a better view on the neurophysiological response to perceived and imagined accents a principal component analysis (PCA) was performed (see Dien and Frishkoff, 2005, for an example of PCA on ERP data). Grand average ERPs over all subjects were computed, as well as ERPs per subject (the number of trials averaged to obtain these ERPs is shown in Table 1). These ERPs were decomposed per experimental condition by means of PCA, resulting in spatial distributions and time-courses for each component, ordered according to magnitude of the eigenvalue (i.e. amount of variance in the data explained by the component). For the time-course of each component, differences between accented and non-accented beats 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 physiologically motivated constraints to increase the sensitivity of the test (Maris, 2004; Maris and Oostenveld, 2007). A main advantage of the PCA is that its output is easy to visualize and interpret. However, this approach works with averaged ERPs and averaging is a fairly coarse way to obtain a generalization of the neurophysiological response to specific mental tasks. Alternatively, a different approach was pursued that has the potential to generalize more accurately over the individual trials.

2.4. Classification

A machine-learning paradigm was used to classify EEG data corresponding to accented and non-accented beats for the perception and imagery conditions, respectively. Classification of the trials was done using a regularized (L_2) Logistic Regression algorithm (Bishop, 2006). Given the experimental design, many more trials were 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. To measure the performance of the classifier, 10-fold cross-validation (Bishop, 2006) was performed. The protocol for 10-fold cross-validation was performed multiple times, each with a different regularization parameter, and the best performance is reported.[1] Since test-sets

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. For classification of perception data, 713 trials were used on average for training the classifier, and 79 for testing. For classification of imagery data, training sets contained on average 1601 trials, using 178 trials for testing. All available time-points per trial within the range of −50 to 450 ms around stimulus onset on all available EEG channels (64) were used as features for classification.

In addition to the separate classification of imagery and perception data, 'cross-conditional classification' was performed. For this approach a classifier was trained on imagery data and tested on perception. A 10-fold cross-validation regime on the imagery data was used to find the optimal regularization parameter, and a classifier using this regularization parameter was retrained on all available imagery data. To avoid a structural preference of the retrained classifier to one of the classes in the new (perception) 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 perception data, while aiming for equal per-class performance. The trials used for calibration were not used for performance evaluation. In a similar fashion, cross-conditional classification was also performed in the other direction, training a classifier on perception and testing it on imagery data.

3. Results

For a better understanding of the classification performance on this data-set, a characteristic of the signal relevant to discrimination of accented and non-accented beats is presented first. A PCA was performed on grand average ERPs and ERPs of a representative subject (S4), and the resulting spatial distributions and time-courses corresponding to the first PCA component, i.e. the principal component with the largest eigenvalue, are shown in Fig. 2. For the grand average, the first PCA component for perception (Fig. 2a) explains 53.6% of the variance, while for imagery (Fig. 2b) this is 52.0%. For subject S4 the first PCA component for perception (Fig. 2c) explains 96.2% of the variance and explains 93.1% for the imagery data (Fig. 2d). Areas in the time-courses contributing to significant differences ($p < 0.05$) are marked in gray on the x-axes in Fig. 2.

When looking at the topographical distributions of the first PCA component in the grand average data, a positive fronto-central distribution with a negative occipital counter-part can be observed for both the perception (Fig. 2a) and imagery (Fig. 2b) condition. Similar topographical distributions can be observed for subject S4, where perception (Fig. 2c) shows a striking resemblance to the imagery (Fig. 2d) condition. The distributions appeared to be relatively invariant over subjects, resulting in great similarity in the grand average distributions and single subject distributions, here illustrated by showing data for S4.

In the time-courses of the PCA components, a significant ($p < 0.001$) difference between accented and non-accented beats was found around 180 ms for both perception and imagery (Fig. 2). In the grand average data, the second positive peak at this latency, presumably the P2 component of the AEP, shows a larger amplitude for accented beats than for non-accented beats in the perception condition. In the imagery condition of the grand average, this effect does not become significant until 200 ms. In the data of subject S4 the effect appears to be reversed, showing a significantly ($p < 0.01$) larger positive deflection around 180 ms for non-accented beats than for accent ones in both imagery and perception condition. This illustrates the variability over subjects. For a more detailed decomposition of grand average data from a similar experimental design, and interpretation of responses to ac-

---

[1] This approach to the selection of a regularization parameter has a potential of over-fitting, but was chosen for computational reasons. As a check, the reported performances were compared with those obtained with a double-nested cross-validation protocol with 10 outer and 5 inner folds, which is robust against this type of over-fitting. As performances did not significantly differ, it can be concluded that in this particular case the regularization parameter selection procedure does not lead to over-fitting.
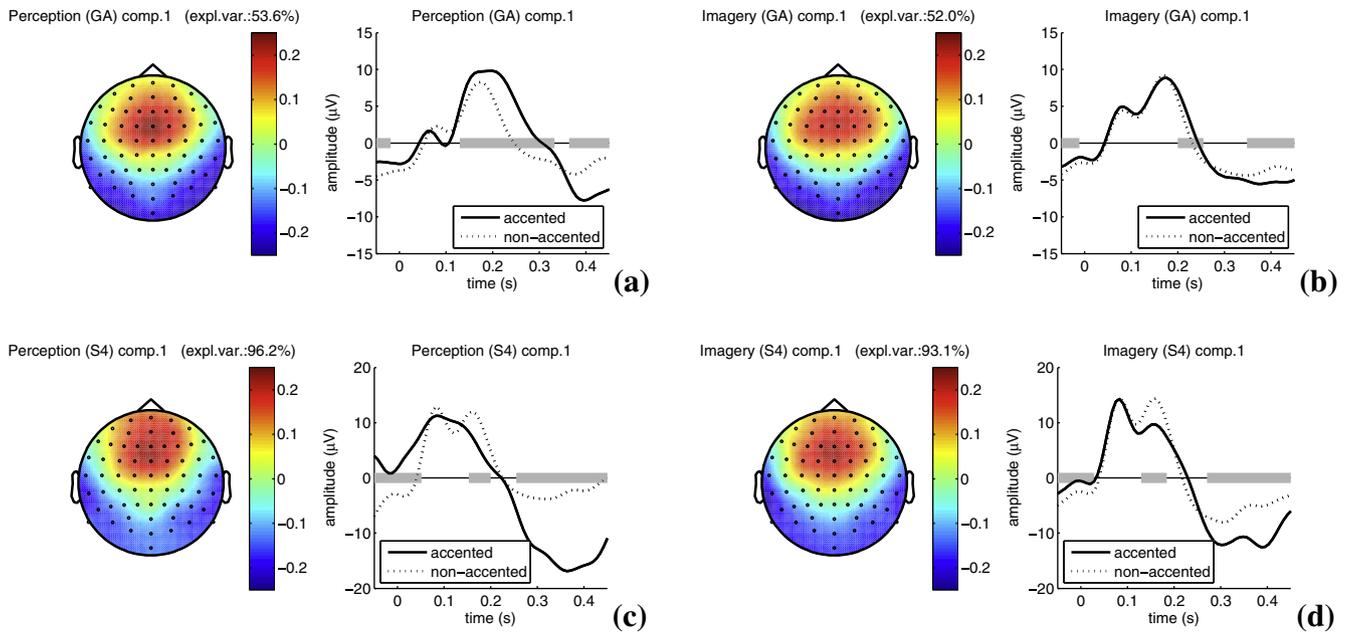
Fig. 2. A decomposition by means of PCA was made per experimental condition of grand average ERPs and ERPs from a representative subject (S4). The number of trials averaged to obtain these ERPs is given in Table 1. Panels (a) and (b) show the topographical distribution and time-course of the first PCA component of the decomposed grand average ERPs for perception and imagery, respectively. Panels (c) and (d) show the same for the decomposed ERPs of a single representative subject (S4). Areas contributing to a significant ($p < 0.05$) difference between accented and non-accented trials are marked by gray bars on the time axis. Similarity can be observed between the grand average and the single subject data, but more importantly, there is a large similarity for the perception and imagery conditions.

cented and non-accented beats with respect to their function in the overarching rhythm patterns, see Schaefer et al. (2011).

In both the grand average data and subject S4 data, time-courses also show a significantly ($p < 0.001$) larger negative deflection after approximately 350 ms for accented beats, than for non-accented beats (Fig. 2). This effect was found both for the perception and imagery conditions, but occurs slightly later in the grand average than in the data of subject S4. It persists until the end of the used time-window and possibly beyond, which may be important when interpreting the early effect between −50 to 50 ms. This early effect is presumably the same long-latency effect, carrying over from the trials where a non-accented beat is preceded by an accented beat. By dividing the non-accented beats into two groups, depending on whether they were preceded by an accented beat, we found a significant difference ($p < 0.05$) for both imagery and perception between these groups in the time interval between −50 and 50 ms for both the grand average and the single subject S4. This supports the idea that the strong negative deflection for accented beats starting around 350 ms persists beyond our time-window.

Next, a classifier was trained to distinguish accented and non-accented beats, based on single-trial 64 channel time-domain data. Its performance per subject is visualized in Fig. 3. For the perception condition, the best subject (S4) reaches a classification rate of 74.9%, while on average subjects reach 69.6% (SD = 5.1%). For the imagery condition, the best subject (S3) reaches 65.7% correct classification, while the average over subjects is 60.8% (SD = 3.9%). These results illustrate that not only perceived accents, but also subjective or imagery accents, not present in the stimulus, can be decoded from brain signals at the level of single trials.

Since the long latency (>350 ms) difference between accented and non-accented beats persists beyond the used time window, this could potentially influence classification performance. To check this, the data was processed and classified in exactly the same way as described, but the 500 ms time-window for a trial was now shifted forward 100 ms (originally starting at −50 ms and now at 50 ms). This way we avoid that the long-latency effect of the accented trials is leaking into the non-accented trials. Classifier performance did not significantly differ from the results
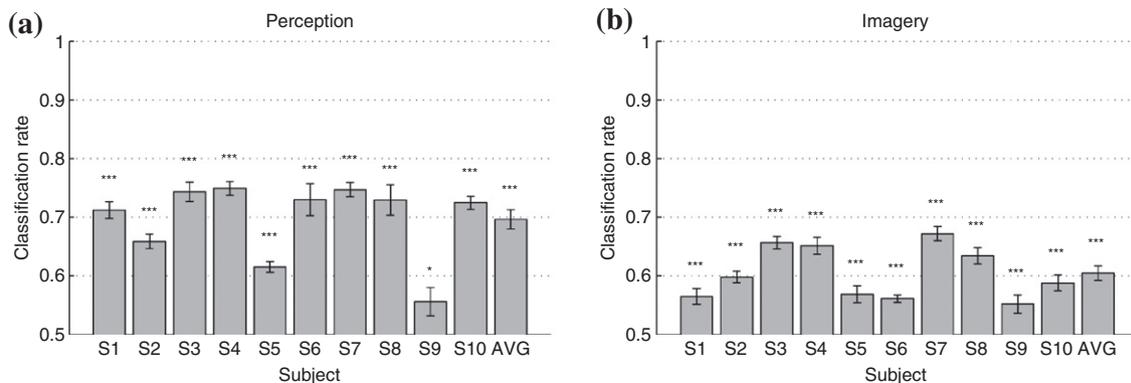
Fig. 3. Balanced classifier performances for the classification of accented versus non-accented beats per subject and as subject-average (AVG) for both the perception (a) and imagery condition (b). The error bars indicate the standard error within the 10-fold cross-validation. Significance levels from chance level (0.5) are indicated by ∗ for $p < 0.05$ and ∗∗∗ for $p < 0.001$.
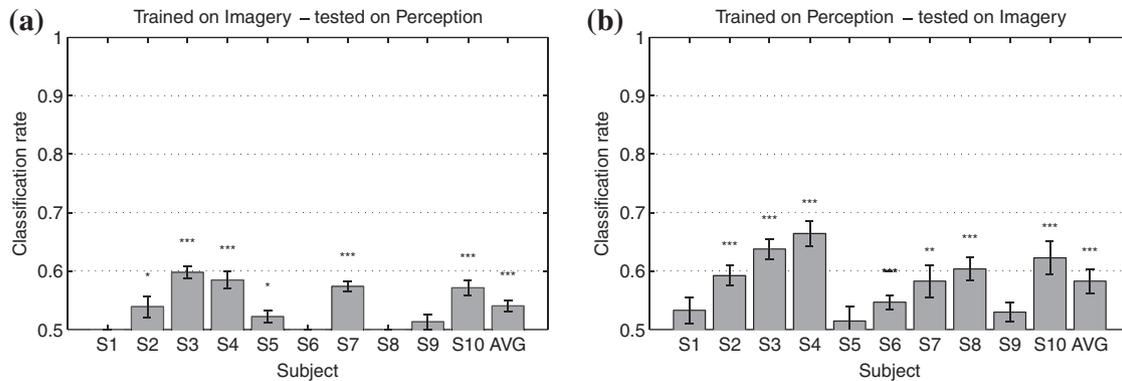
Fig. 4. Balanced classifier performances for the classification of accented versus non-accented beats are shown for all subjects and as subject-average (AVG) for a classifier that is trained on imagery and tested on perception (a), and the other way around (b). The error bars indicate the standard error within the 10-fold cross-validation. Significance levels from chance level (0.5) are indicated by ∗ for $p < 0.05$, ∗∗ for $p < 0.01$ and ∗∗∗ for $p < 0.001$.

achieved with the early time-window starting at −50 ms. This means that the classifier's performance was not unjustifiably boosted by our initial choice of the time window.

The preprocessing step, where EOG channels were de-correlated from the EEG, appeared to be a useful method for removing most of the eye artifacts in the EEG data. However, the method cannot guarantee that artifacts are completely removed. To make sure that eye artifacts possibly remaining are not providing the classifier with class-relevant information, classification was performed on the EOG channels alone. None of the subjects achieved a performance significantly different from chance level in this case.

The similarity between PCA components for the perception and imagery data in Fig. 2 suggests that there is shared processing for perception and imagination. To investigate this further, cross-condition classification has been carried out. Fig. 4 shows balanced classification performances for all subjects when a classifier was trained on perception (on average 792 trials) and tested on imagery data, and the other way around (where the training set on average consisted of 1779 trials). For a classifier trained on imagery and tested on perception data performance was 59.8% for the best subject (S3) and an average over subjects of 54.9% (SD = 3.7%). Classification rates are significantly ($p < 0.05$) different from chance level for 6 out of 10 subjects. The classifier trained on perception and tested on imagery data yielded a performance of 66.4% for the best subject (S4), and an average performance of 58.8% (SD = 5.8%) over all subjects. Classification rates are significantly ($p < 0.05$) different from chance level for 7 out of 10 subjects. For some subjects cross-conditional classification did not yield performances significantly different from chance level. This predominantly concerns subjects that also performed worse in imagery and perception (e.g. S9), which seems to point to a more general problem of signal strength.

4. Discussion

The results of this study show that it is possible to decode perceived and imagined accents from brain signals on a single-trial level with a mean accuracy of 69.6% for perception and 60.8% for imagery, using 500 ms of data. We described a method for cross-condition classification and showed that it is possible to classify imagery data significantly above chance with a classifier trained on perception data, and vice versa. The fact that cross-condition classification is possible, indicates that trials from the perception and imagery condition contain shared information about the presence or absence of accents. Independent classification of EOG signals did not result in a performance significantly above chance for any of the subjects, which allows us to conclude that possible eye-artifacts were not unjustifiably boosting classifier performance.

According to the AEP literature (Mulert et al., 2005; Burkard et al., 2007; Mayhew et al., 2010), the N1 and P2 components in the AEP may be influenced by intensity differences as are present between accented and non-accented stimuli in the perception condition. Considering that these stimuli also differ in harmonic content, the effect of timbre (or harmonic content) on P2 amplitude, reported by Meyer et al. (2006), may also be relevant. The N1 and P2 can be identified in the decomposed ERPs in Fig. 2 at about 100 and 180 ms, respectively. Although no significant difference in the N1 component is observed, possibly due to relatively small intensity differences between accented and non-accented beats, the P2 is significantly different between perceived accents and non-accents (see Fig. 2a and c). Surprisingly, the P2 in the imagery condition is also significantly modulated (see Fig. 2b and d). This could reflect the imagined difference in timbre between accented and non-accented beats. Although some form of imagery process was guaranteed by the behavioral task in our experimental design, it is hard to tell what strategy subjects exactly used. Therefore, it is not unlikely that the inter-subject variability observed for the P2 modulation reflects the differences in imagery strategies.

An important effect observed for the distinction of accented and non-accented beats, is a late (>300 ms) effect, yielding a larger negative deflection for accented, than for non-accented beats. Both timing and scalp distribution of this effect are consistent with the findings of Schaefer et al. (2011). Similar effects were also observed in a different musical task (Pearce et al., 2010), but a clear explanation of the underlying mechanism is not available yet. A connection to mechanisms for selective attention, with resulting 'processing negativity' (Naatanen, 1982) or contingent negative variation (CNV) (Chen et al., 2009) effects, is not unlikely, but highly speculative (mainly because both effects have usually been reported to occur at different latencies).

As shown in Fig. 2, signals relevant to the discrimination of accents from non-accents have an almost identical topographical distribution over the scalp and reveal much resemblance in the time-course of the perception and imagery condition. In combination with positive results of cross-condition classification, this suggests that similar brain areas are involved in both perception and imagery. This is supported by the finding of shared brain areas for musical perception and imagery as reported by Zatorre et al. (1996) and Kraemer et al. (2005) and described by Halpern (2001) and Hubbard (2010). The finding of shared mechanisms for rhythm processing seems to complement existing literature on shared mechanisms for timbre and pitch.

Comparing the PCA analysis and single-trial classification, we conclude that both methods led to the same conclusion regarding information shared between perception and imagery data.

Although the single-trial classification procedure potentially allows for a much better generalization of an effect than PCA on the averaged ERP, it is also more susceptible to noise and variance. Issues like non-stationarity in brain responses, in both background activity and responses related to the task, as well as noise sources outside the brain (from the environment or other parts of the body), cause performances to deviate from 100% correct classification.

Independently of the neurophysiological or cognitive implications, our results will certainly be beneficial for research and applications in brain–computer interfacing. Since subjective accents can be voluntarily added, it allows users to encode their intentions in their brain signals. The method described in this paper can be utilized to decode these intentions (for application of this approach, see Vlek et al., in press). The fact that a classifier can be trained on perceptual data and later applied to imagery data adds extra convenience, because this takes much load off the subject (listening is much easier than imagining). A brain–computer interface based on this auditory paradigm, may be able to augment or restore (communicative) abilities of patients (e.g. those suffering from amyotrophic lateral sclerosis or spinal cord injury). The commonalities between perception and imagery also have implications for cue based rehabilitation tasks. Based on these results, we might expect similar brain activations for self-generated rhythms in movement as in externally cued rhythms. Although a number of relevant aspects of cuing, such as tempo and stimulus complexity, were not investigated here, there is support for the notion that we can also deliberately generate the brain activity that would normally result from hearing a rhythmic pattern. The practical implications of these findings need to be further investigated, but considering the broad nature of functions apparently effected by rhythmic processing, the current results show promise in multiple clinical directions.

Acknowledgements

References

Belin P, Eeckhout PV, Zilbovicius M, Remy P, Francois C, Guillaume S, et al. Recovery from nonfluent aphasia after melodic intonation therapy: a PET study. Neurology 1996;47:1504–11.

Bishop C, editor. Pattern recognition and machine learning. Springer; 2006.

Bolton T. Rhythm. Am J Psychol 1894;6:145–238.

Brochard R, Abecasis D, Potter D, Ragot R, Drake C. The ticktock of our internal clock: direct brain evidence of subjective accents in isochronous sequences. Psychol Sci 2003;14(4):362–6.

Burkard R, Don M, Eggermont J, editors. Auditory evoked potentials. Lippincott Williams and Wilkins; 2007.

Chen Y, Huang X, Yang B, Jackson T, Peng C, Yuan H, et al. An event-related potential study of temporal information encoding and decision making. Neuroreport 2009;21(2):152–5.

Desain P, Honing H. Single trial ERP allows detection of perceived and imagined rhythm. In: Proceedings of the RENCON workshop international joint conference on artificial intelligence (IJCAI); 2003.

Dien J, Frishkoff G. Event-related potentials: a methods handbook. In: Principal components analysis of event-related potential datasets. MIT Press; 2005.

Dornhege G, del R. Millán J, Hinterberger T, McFarland D, Müller K-R, editors. Toward brain–computer interfacing. MIT Press; 2007.

Farah M, Smith A. Perceptual interference and facilitation with auditory imagery. Percept Psychophys 1983;33(5):475–8.

Fraise P. The psychology of music. In: Rhythm and tempo. Academic Press; 1982. p. 149–80.

Gerven MV, Farquhar J, Schaefer R, Vlek R, Geuze J, Nijholt A, et al. The brain–computer interface cycle. J Neural Eng 2009;6.

Halpern A. Cerebral substrates of musical imagery. Ann N Y Acad Sci 2001;930:179–92.

Halpern A, Zatorre R, Bouffard M, Johnson J. Behavioral and neural correlates of perceived and imagined musical timbre. Neuropsychologia 2004;42:1281–92.

Hubbard T. Auditory imagery: empirical findings. Psychol Bull 2010;136(2):302–29.

Iversen J, Repp B, Patel A. Top-down control of rhythm perception modulates early auditory responses. Ann N Y Acad Sci 2009;1169:58–73.

Jongsma M, Eichele T, Quiroga RQ, Jenks K, Desain P, Honing H, et al. Expectancy effects on omission evoked potentials in musicians and non-musicians. Psychophysiology 2005;42(2):191–201.

Kasai K, Asada T, Yumoto M, Takeya J, Matsuda H. Evidence for functional abnormality in the right auditory cortex during musical hallucinations. Lancet 1999;354(9191):1703–4.

Kelso J, editor. Human motor behavior. Lawrence Erlbaum Associates; 1982.

Kosslyn S, Ganis G, Thompson L. Neural foundations of imagery. Nat Rev Neurosci 2001;2:635–42.

Kosslyn S, Thompson W, Alpert IKN. Topographical representations of mental images in primary visual cortex. Nature 1995;378:496–8.

Kraemer D, Macrae C, Green A, Kelley W. Musical imagery: sound of silence activates auditory cortex. Nature 2005;434(158).

Latimer C, Keeling J, Lin B, Henderson M, Hale L. The impact of bilateral therapy on upper limb function after chronic stroke: a systematic review. Disabil Rehabil 2010;32:1221–31.

London J. Hearing in time: psychological aspects of musical meter. Oxford University Press; 2004.

Maris E. Randomization test for ERP topographies and whole spatiotemporal data matrices. Psychophysiology 2004;41(1):142–51.

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

Mayhew S, Dirckx S, Niazy R, Iannetti G, Wise R. EEG signatures of auditory activity correlate with simultaneously recorded fMRI responses in humans. Neuroimage 2010;49:849–64.

McFarland D, Miner L, Vaughan T, Wolpaw J. Mu and beta rhythm topographies during motor imagery and actual movement. Brain Topogr 2000;12(3):177–85.

McIntosh G, Brown S, Rice R, Thaut M. Rhythmic auditory-motor facilitation of gait patterns in patients with Parkinson's disease. J Neurol Neurosurg Psychiatry 1997;62:22–6.

Meyer M, Baumann S, Jancke L. Electrical brain imaging reveals spatio-temporal dynamics of timbre perception in humans. Neuroimage 2006;32:1510–23.

Michon J, Jackson J. Time, mind and behavior. Berlin: Springer; 1985.

Mulert C, Jäger L, Propp S, Karch S, Störmann S, Pogarell O, et al. Sound level dependence of the primary auditory cortex: simultaneous measurement with 61-channel EEG and fMRI. Neuroimage 2005;28(1):49–58.

Munzert J, Lorey B, Zentgraf K. Cognitive motor processes: the role of motor imagery in the study of motor representations. Brain Res Rev 2009;60:306–26.

Naatanen R. Processing negativity: an evoked-potential reflection of selective attention. Psychol Bull 1982;92(3):605–40.

Navarro Cebrian A, Janata P. Electrophysiological correlates of accurate mental image formation in auditory perception and imagery tasks. Brain Res 2010;1342:39–54.

Overy K. Dyslexia and music. from timing deficits to musical intervention. Ann N Y Acad Sci 2003;999:497–505.

Pearce M, Ruiz MH, Kapasi S, Wiggins G, Bhattacharya J. Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. Neuroimage 2010;50:302–13.

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

Roerdink M, Lamoth C, Kwakkel G, van Wieringen P, Beek P. Gait coordination after stroke: benefits of acoustically paced treadmill walking. Phys Ther 2007;87:1009–22.

Schaefer R, Vlek R, Desain P. Decomposing rhythm processing: electroencephalography of perceived and self-imposed rhythmic patterns. Psychol Res 2011;75:95–106.

Schauer M, Mauritz K. Musical motor feedback (MMF) in walking hemiparetic stroke patients: randomized trials of gait improvement. Clin Rehabil 2003;17:713–22.

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

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

Shinosaki K, Yamamoto M, Ukai S, Kawaguchi S, Ogawa A, Ishii R, et al. Desynchronization in the right auditory cortex during musical hallucinations: a MEG study. Psychogeriatrics 2003;3(2):88–92.

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

Thaut M. Neurologic music therapy in cognitive rehabilitation. Music Percept 2010;27:281–5.

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

Vlek R, Schaefer R, Gielen C, Farquhar J, Desain P. Sequenced subjective accents for brain–computer interfaces. J Neural Eng, doi:10.1088/1741-2560/8/3/036002.

Willems A, Nieuwboer A, Chavret F, Desloovere K, Dom R, Rochester L, et al. Turning in parkinson's disease patients and controls: the effect of auditory cues. Mov Disord 2007;22:1871–8.

Zanto T, Snyder J, Large E. Neural correlates of rhythmic expectancy. Adv Cogn Psychol 2006;2(2–3):221–31.

Zatorre R, Halpern A, Perry D, Meyer E, Evans A. Hearing in the minds ear: a PET investigation of musical imagery and perception. J Cogn Neurosci 1996;8:29–46.