Supplement

Name that tune: Decoding music from the listening brain

Rebecca S. Schaefer [a,*], Jason Farquhar [a], Yvonne Blokland [b], Makiko Sadakata [a], Peter Desain [a]

[a] Donders Institute for Brain, Cognition and Behavior: Centre for Cognition, Montessorilaan 3, 6525 HE Radboud University Nijmegen, The Netherlands
[b] Cognitive Artificial Intelligence, Heidelberglaan 6, 3584 CS, Utrecht University, The Netherlands

ARTICLE INFO

ABSTRACT

In the current study we use electroencephalography (EEG) to detect heard music from the brain signal, hypothesizing that the time structure in music makes it especially suitable for decoding perception from EEG signals. While excluding music with vocals, we classified the perception of seven different musical fragments of about three seconds, both individually and cross-participants, using only time domain information (the event-related potential, ERP). The best individual results are 70% correct in a seven-class problem while using single trials, and when using multiple trials we achieve 100% correct after six presentations of the stimulus. When classifying across participants, a maximum rate of 53% was reached, supporting a general representation of each musical fragment over participants. While for some music stimuli the amplitude envelope correlated well with the ERP, this was not true for all stimuli. Aspects of the stimulus that may contribute to the differences between the EEG responses to the pieces of music are discussed.

© 2010 Elsevier Inc. All rights reserved.

Introduction

Recent neuroimaging studies show an increasing ability to decode from brain activity what is being perceived, in both visual (Kay et al., 2008; Haynes, 2009) and auditory (Formisano et al., 2008; Ethofer et al., 2009) modalities. This has however primarily been done using the hemodynamic brain response. We present a new development in this so-called brainreading approach, in that we decode perceived music from the electrophysiological signal from the brain. Using linear discriminant classification we are able to detect up to 100% correctly which musical fragment was presented, out of seven possible choices, with 3 s of EEG data per musical fragment, using the time course sampled at 30 Hz. Taking basic aspects of the stimulus such as the amplitude envelope, as well as ratings of musical content, we find class-dependent differences to in the response to relate to different aspects of the stimuli, using low-level perceptual features but likely cognitive events as well.

Although most of the recent brainreading studies report multivariate decoding of fMRI signals, of course similar research can be done with electroencephalography (EEG), where the number of channels is much lower than the number of voxels, but the number of time samples and trials is usually much larger. Using multivariate decoding methods on EEG signals is quite common in the domain of brain–computer interfacing (BCI) research, where covert mental actions are decoded in real-time with the goal of driving a device with one's brain signals, or thoughts (for a review, see Van Gerven et al., 2009). Although the

detection of heard stimuli is not directly useful for BCI research as it does not relate any intention, classifying heard sounds teaches us more about the neural representations of complex auditory events, such as music and speech perception.

The EEG response to music perception has been studied in the context of specific musical aspects, albeit mostly in the context of more generally occurring components of the event-related potential (ERP). These components, such as the N1/P2 complex (Dekio-Hotta et al., 2009), the P300 oddball response (Halpern et al., 2007; Krohn et al., 2007) and the mismatch negativity (MMN, Brattico et al., 2006; Trainor et al., 2002) have all been shown to be influenced by musical characteristics. ERP responses to musical rhythms have also been investigated, for different metric levels such as the note, beat and bar level (Jongsma et al., 2004), as well as subjective accents (Brochard et al., 2003). Combining a number of these types of musical accents, Palmer et al. (2009) found that different types of accent in a melodic sequence (such as timbre changes, melodic and temporal accents) induce different types of ERP response, showing that listeners' neural responses to musical structure changed systematically as sequential predictability of the melodies and listeners' expectations changed across the melodic context. Similarly to sentences in language, musical phrase endings also often induce a so-called closure-positive shift (Neuhaus et al., 2006). These studies generally use artificially created, usually monophone, isochrone melodies, to isolate the musical aspect under investigation (Schaefer et al., 2009). The response to fully ecological, valid music stimuli has not been studied much. One example is a study by Bhattacharya et al. (2001), who found different patterns of dependencies between EEG channels while listening to different pieces of music, which are higher for listeners with musical training than for non-musicians. Early studies on classifying auditorily presented language

(words or sentences) produced good results (Suppes et al., 1997, 1998), indicating feasibility of classification using EEG signals, but to our knowledge no such effort has been made for musical material.

For the current study, we selected seven musical fragments, aiming to detect the stimulus from the EEG signal. Considering the high temporal resolution of EEG, we hypothesized that the inherent temporal structure of music would make this feasible. We chose natural music stimuli under the assumption that this would produce a more distinct brain response than artificial music. The stimuli were chosen based on their length and perceptual naturalness when played as a loop, as well as maximizing musical differences (such as subjective loudness, western or non-western tonal systems, different time signature, salient melody or background texture). We predicted there to be a relation between sound level and the ERP as the sound intensity dependence of the N1/P2 complex is well-documented and shown to be attributed to auditory cortex activity (Mulert et al., 2005). We are interested to see if there are other musical dimensions that can also help in predicting the class. Thus our main question is whether classification of perceived music is possible, and we explore some aspects of the stimulus to see which relate the most to the brain response.

Method

Participants

Ten subjects volunteered to participate in the study, of which five males, aged 22–53. All had normal hearing, and none had known neurological abnormalities. All participants were right-handed. Musical experience was not a controlled factor; one participant had formal music training (conservatory level) and four others play an instrument regularly (more than once a week). All participants were mainly exposed to Western tonal music as a cultural background, and their musical preferences differed widely.

Stimuli

Stimuli were created out of fragments of recorded music, on the basis of comparable length and musical differences. The original music that the fragments were taken from, and their durations are shown in Table 1, they can also be listened to at http://www.nici.ru.nl/mmm. S1, S4 and S6 are relatively quiet musical fragments, whereas S2, S3 and S7 are relatively intense and energetic, S5 is neutral in this sense. S1, S3, S4, and S5 use Western tonal scales and harmonies, whereas S2 and S7 do not, containing respectively a non-western scale with micro-tuning deviations and non-western harmonies. S1, S2, S3, S4 and S7 have prominent melodies, whereas S5 does not, and mainly provides a musical texture or background. S6 does not include any clear pitch information (only rhythm). All fragments were in 4/4 measures, except for S2. None of the fragments included vocals. The musical fragments were presented back-to-back in randomized sequences, that each consisted of two instances of each fragment, shown schematically in Fig. 1. The randomization ensures a pre-stimulus context that is not class-specific, and averages out any carry-over response from the previous trial. The starting fragment of the sequence was counter-balanced but not used in analysis. Seventy sequences were created, each containing 15 fragments. Additionally, of each stimulus, five sequences were also made in which a single fragment was repeated ten times, to investigate multi-trial classification. All together this resulted in 145 sequences, which were equally divided over five blocks, with the opportunity for the participant to start every sequence with a button press. The total number of trials (before artifact rejection) was 140 for each fragment in the random context, and five repeated sequences of each to use for multi-trial sequence classification. The total duration of the experiment amounted to about 1 h and 45 min (excluding cap fitting time).

Equipment and procedure

The data were recorded using a Biosemi Active-Two system with 64 EEG channels (placed according to the 10–20 system, Jasper, 1958), and eight EMG/reference channels (double mastoids, horizontal and vertical EOG, EMG of the long neck muscle (longus capitis) to control for head nodding and the laryngeal muscle (cricothyroid) to control for subvocalization. The offsets of the active electrodes were kept below 30 mV at the start of the measurement, EEG was sampled at 2048 Hz. The experiment was programmed in StimCat and run on the Brain-Stream platform (www.brainstream.nu), which are both MATLAB code packages soon to be open source. Audio files were edited using Audacity 1.2.5 (http://audacity.sourceforge.net). The instructions and fixation cross were displayed on a 17″ TFT screen with a 600×800 pixel resolution, and stimuli were played through passive speakers (Monacor, type MKS-28/WS) at a comfortable listening level adjusted to the preference of the participant (peaking between 86 and 84 dB SPL). The analyses were performed in MATLAB (Mathworks, Nantick USA).

In order to test if the audio speakers in the EEG cabin somehow affecting the measured signal, we performed the same experiment with a watermelon instead of a brain, using an identical set-up with conductive gel and positioning towards the speakers, see Fig. 2. A similar approach was used in Akhoun et al. (2008) when investigating artifacts in auditory brainstem responses.

Analyses

The data were initially down-sampled to 256 Hz. To remove slow drift, linear de-trending was used and electrodes were re-referenced using a common average. After removal of bad epochs and bad channels (determined by an amplitude of >3.5 times the standard deviation), a spectral filter was applied with high-pass at 1 Hz and low-pass at 14 Hz and further down-sampled to 30 Hz, as initial investigation of the frequency domain revealed that no useful information was found in higher frequencies. As the shortest musical fragment was 3.26 s, this was the epoch duration used for all stimuli. The first fragment of every sequence was not used in the data analysis to avoid state-change effects. For classification, the multi-class problem was split into one-vs.-rest style binary subproblems, for each of which a quadratically regularized linear logistic-regression classifier (Bishop, 1995) was trained. Each binary subproblem classifier, c_i, estimates the posterior probability of its class, $Pr(c_i|X)$, given the data, X, so the class with maximum a-posteriori

Table 1
The musical phrases that were used as stimuli were fragments taken from the recordings listed here.

Stim	Title	Performer	Record label	Year	Length
S1	Tchaikovsky's Nutcracker Suite: March	Kazuchi Ono & Bratislava Radio Symph. Orchestra	Compose Records	1997	3.26 s
S2	Galvanize	The Chemical Brothers	Freestyle Dust	2005	3.45 s
S3	Daft Punk is Playing at my House	LCD Soundsystem	DFA	2005	3.52 s
S4	Agua de Beber	Antonio Carlos Jobim	Verve	1963	3.59 s
S5	Release the Pressure	Leftfield	Columbia	1995	4.36 s
S6	How Insensitive	Richard 'Groove' Holmes	P-Vine Japan	2004	3.87 s
S7	Erkilet Güzeli	Antwerp Gipsy-Ska Orkestra	Evil Penguin Records	2007	3.94 s

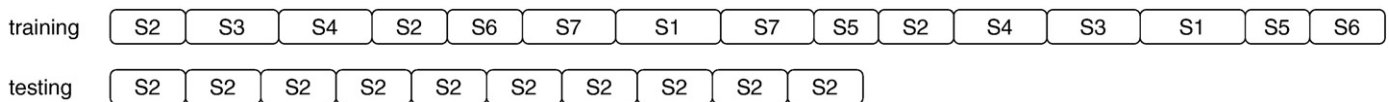| training	S2	S3	S4	S2	S6	S7	S1	S7	S5	S2	S4	S3	S1	S5	S6
testing	S2	S2	S2	S2	S2	S2	S2	S2	S2	S2					

Fig. 1. A schematic representation of the stimulus sequences is shown: randomized sequences were used to collect data to train the classifier with, and multitrial-sequence classification was done on continuous data of single repeating stimuli (in this case, S2).

(MAP) probability was used for prediction. Classification performance was assessed in two ways: individually (using training and test data from a single participant with performance estimated by 10-fold cross-validation) and cross-participant (training the classifier on nine participants and testing it on the one remaining participant). To ensure that the classification results are based upon brain signals and not artifactual sources, the classification was also carried out for the processed signal from the EMG channels (referenced to the linked mastoids). The signals were processed according to standard treatment of EMG signals (Reaz et al., 2006), applying a high-pass filter at 8 Hz, and taking the envelope of the remaining signal. This envelope shows actual muscle activity, and its classification reveals class-specific muscle movement.

We also investigate the result of combining multiple consecutive epochs to improve performance. To do this, the MAP approach used to generate multi-class predictions was extended to include multiple per-epoch classifier predictions,[1] which were tested on continuous data of a single repeating musical fragment. To prevent any effect of a block design in the stimuli, the multi-trial sequences were used for testing only, training of the classifier was done using trials from the randomized sequences.

To visualize the location and time-course of the class-relevant features we first computed grand-average per-class ERPs across all subjects and electrodes by temporally concatenating the stimuli into one long ERP. We then used singular-value-decomposition to decompose these grand-average ERPs into pairs of spatial-patterns and per-class time-courses. In this way we can clearly see both where a particular component is localized and when it responds to different stimulus inputs. The time-courses produced in this way are particularly interesting, as by correlating them with features of the audio stimulus, such as its acoustic envelope or musical content, we can determine which stimulus features generate a strong brain response.

To ascertain the relation between sound volume dependence and the ERP, the correlation between the time course of this main component and the audio envelope was calculated. As an exploratory test, some basic measures of music content were also investigated. This musical content was rated behaviorally by two raters to see if certain events in the music may be related to some aspect of the EEG response. This was done by locating event onsets based on the audio waveform in Praat (http://www.fon.hum.uva.nl/praat/) and scoring each event on 11 dimensions: subjective loudness, beat (the presence of a downbeat) or syncopation (as defined by Longuet-Higgins and Lee, 1982), complexity of harmonic structure, melodic events (as opposed to rhythmic), large interval jumps, novelty (the presence of a new, non-repeating sound event), and the level of expectations answered or violated in the harmony, rhythm (syncopation), timbre, melody or pitch (denoting surprising events in the music). These profiles were then correlated with the same time-course as the envelopes before, only with the mean subtracted to remove general perceptual responses and leave only the stimulus-specific response.

Results

The single trial classification results for both the individual and cross-participant classification are shown in Fig. 3, showing classification rates

far above chance for all participants, both for individual and cross-participant classification. The best individual participant (P4) shows 70% correct classification for the seven-class problem (chance level is 14.3%), the best result for the cross-participant classification is 53% correct (P1 and P6). When comparing the participants' results to each other, T-tests between all the within-participant results show that due to the large number of measuring points, almost all comparisons are significant at $p < 0.01$ and all are significant at $p < 0.05$. The results of the classification on the EMG channels are also shown in Fig. 3, and are all around chance, showing that sub-vocalization or small head movements are not the source of the classification results. The results for the watermelon phantom are also at chance level.

The sequence classification results are shown in Fig. 4, with increasing classification rates for all participants, and up to 100% correct classification achieved after six trials for the best participant (thus using 19.6 s of data). The average multi-trial rate over all participants starts at 50% and goes up to 88% after nine consecutive trials of continuous data.

When decomposing the grand average ERP response over all participants, a fronto-central component emerges that explains 23% of the total variance. The time-course of this component for the different musical stimuli shows the most distinction between the classes, indicating it may be a cleaner response to the stimuli. The second component explains 9% of the variance, but the time-courses do not show big differences between the stimuli. Considering the distribution and the time-course, with a small peak around 300 ms, a possible interpretation of this response may be that it represents the processing of novelty for each new fragment in the randomized sequences changing every 3 to 4.3 s. However other processes such as working memory, prediction formation, memory retrieval and such may also be implicated. The third component appears to add extra detail to the separate stimuli, explaining only 5% of the variance. The components are shown in Fig. 5 and the first 20 components explaining the first 75% of the total variance, and their distributions, are shown in the supplementary online material.

The differences in average classification rate per stimulus, see Fig. 6, show that all stimuli are detectable at the $p < 0.01$ significance level, although some musical fragments elicited a somewhat stronger classifiable response than others. For the within-subjects classification,

Fig. 2. To check for artifacts in the set-up, the experiment was also carried out with a watermelon phantom.

---

[1] Assuming epoch independence, the MAP probability of class c_i given the sequence of data, $[X_1, X_2, \ldots]$, is given by the product of epoch predictions, $Pr(c_i | X_1, X_2, \ldots) = \prod_{j=1,2,\ldots} Pr(c_i | X_j)$.
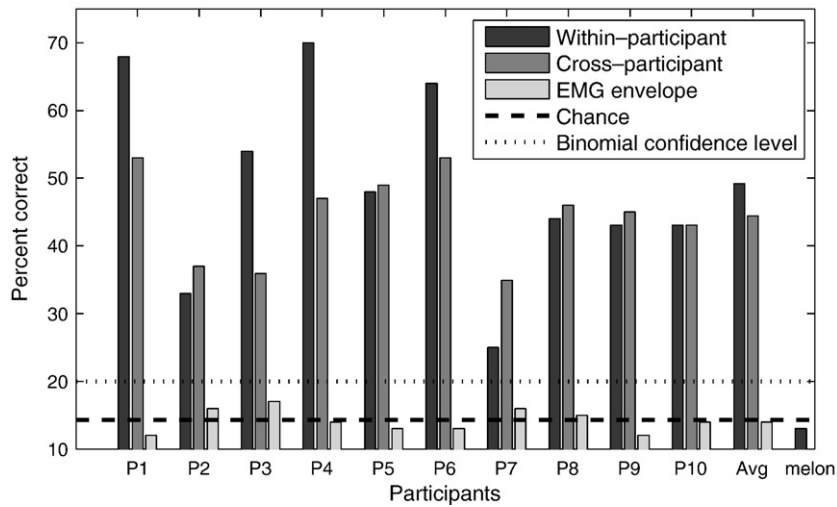
Fig. 3. The classification results are shown for both the individual and the cross-participant classification. The results of the EMG envelope and the melon phantom are also shown (cross-participant classification not performed). Chance level is 14.3% for a seven-class problem (broken line), the $p < 0.01$ significance level is 20% (dotted line) showing all rates to be significantly higher than chance.

the results are significantly better for S6 and S1 than for the others ($p < 0.01$). The clear advantage of the cross-participant classification is very likely due to the increased amount of training data used in this case by merging all the data. Here, S1 has the strongest classifiable response and S7 is notably weaker than the others.

The comparison of the ERP of Component 1 from Fig. 5 to the envelope of the sound in the stimulus is shown in the left panel of Fig. 7, showing that the correlation differs considerably per stimulus (from −.07 to .48). Although the correlations vary over stimuli, closer examination of the stimuli for which the correlations are above .10 shows that for the music with fewer events and more space in between, the envelope predicts the ERP response quite well, whereas this is not true for stimuli that have a high melodic event density (S2 and S7, which are two of the less easily detected stimuli). By finding the time lag at which the correlation between the envelope and the ERP is maximal, we get an impression of the response time of the ERP to the sound intensity. When summing the correlation at different time-lags for every stimulus (shown in the bottom of Fig. 7.1) the highest correlation over all stimuli is lagged at about 70–100 ms, which is an indication of the processing time and may likely relate to the N1 response.

The results of the music comparison are less clear, and of an exploratory nature. The correlations for four of the rated dimensions are shown in the right panel of Fig. 7. The 'Beat'-profile (Fig. 7.2A) shows a common positive correlation around 100 ms, which may indicate an influence on the N1/P2 complex. The differences per stimulus show this to be context-dependent: for instance S2 and S7, which have a weaker metric accent than the other stimuli, do not show a strong response to the meter. Another profile identified syncopation ('Exp beat', Fig. 7.2B), which shows a relatively early correlation (≈ 100 ms) for some stimuli (not all stimuli contain syncopations), and again S2 and S7 emerge, together with S3 and S6 which also contain strong syncopated accents. An early correlation with melodic events (Fig. 7.2C) shows the pitch to have influence as well, and Melody Jumps, or large (>4 semitones) intervals, Fig. 7.2D) indicates that the processing of the size of the interval may be a slower process, shown for those stimuli with a clear melody (thus excluding S5 and S6). As S5 was chosen to investigate a fragment with only a musical texture or background, it is interesting to see that this was also a stimulus that was relatively less easy to detect.

Discussion

The current study shows that it is possible to detect perceived music from the single trial EEG signal. The results do not stem from a mechanical or muscle artifact, as tested with a watermelon phantom and glottal and neck EMG respectively. Also, the maximum correlation of the EEG signal with the auditory envelope at 100 ms is evidence that some
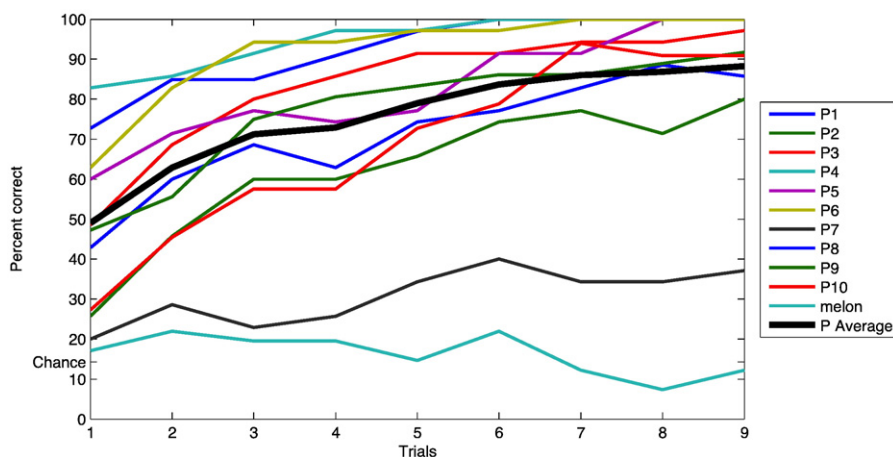

Fig. 4. Here, the multi-trial classification results are shown per participant and the melon phantom, as well as the average over participants (thick black line). Chance level is 14.3% for a seven-class problem (mark on *y*-axis).

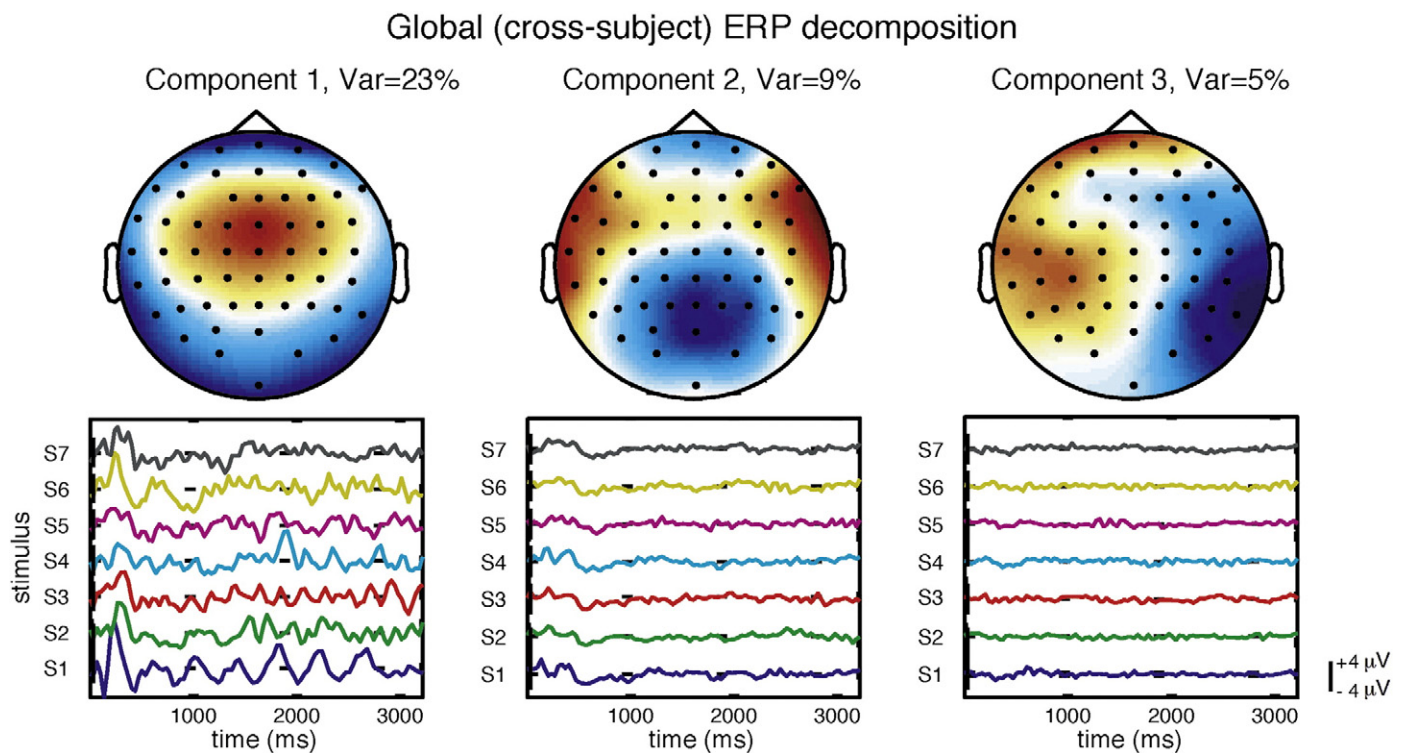## Global (cross-subject) ERP decomposition

Fig. 5. The first three components contributing to the variance of the grand average ERP (all participants, all stimuli). The time-courses for the seven stimuli are plotted below, showing the first component to distinguish most between the musical fragments. The bar on the bottom right shows the scale of the y-axis, with the bar on each time-course running from −4 to 4 μV.

processing is involved. The range of results is quite broad for the different participants, with individual single trial seven-class rates varying from 25% to 70%. This seems to involve a response that is general enough to be able to achieve 35% to 53% while classifying across participants, showing detection to be possible based on a general template and without training the classifier on subject-specific data. This shows that, contrary to most tasks investigated in BCI research, there are no subjects with unclassifiable responses (what is referred to as illiteracy for a certain task). This is likely due to early perceptual processing, which may be more common and less individualized then a BCI task, such as imagined movement. Although S1 was the stimulus that was most easily detectable in the cross-participant classification, all stimuli were detectable based on a single trial. For the within-participant classification, the range of binary classification rates was 58% to 64%, and the cross-participants rates varied from 65% to 79%. The stimuli that were on average least detectable also showed the smallest correlation with the audio envelope or had the least musical information.

To better describe the information that is used by the classifier, we decomposed the averaged ERP response for all the stimuli concatenated, over all participants. In this way we can isolate the class-specific responses better by discarding components that do not add any information based on their time-courses, and see which channels are most important based on the component weight distribution. The decomposed ERP shows a main component with a fronto-central distribution that explains 23% of the data and shows the most distinction between the classes. Comparison of the time-course of this component with the audio envelope shows that the correlation between them is quite high for some stimuli, but quite low for some others. The most obvious difference between these stimuli appears to be the event density in the music, however other factors could also be of influence, such as the bandwidth of the signal, as was reported by Atcherson et al. (2009). Exploring a possible relation with music structure we found the strongest correlation lags between the EEG and different musical events differed over stimuli as well, supporting the notion that different processes are contributing to the brain signal

that is useful for the detection of the representation. Considering that all stimuli were detectable, including stimuli of which the amplitude envelope and the ERP hardly correlate, the unique response to a piece of music suggests a combination of both bottom-up and top-down processes. This is supported by the finding by Brechmann et al. (2002) that while responses in the primary auditory areas show a clear dependence on stimulus intensity, the auditory association areas do not show this close relation. Recent results from concurrent EEG and fMRI reported by Mayhew et al. (2010), namely that auditory stimulation induces activity in bilateral secondary auditory cortices as well as the right pre- and post-central gyri, anterior cingulate
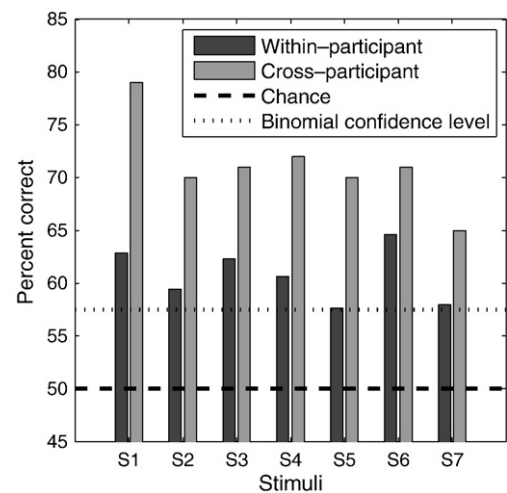

Fig. 6. The single trial classification rates per stimulus in percentages (binary, 1 versus all others) are shown here averaged over participants, both for individual and cross-participant classification. Chance level is 50%, the p<0.01 significance level is 57.5% (dotted line).
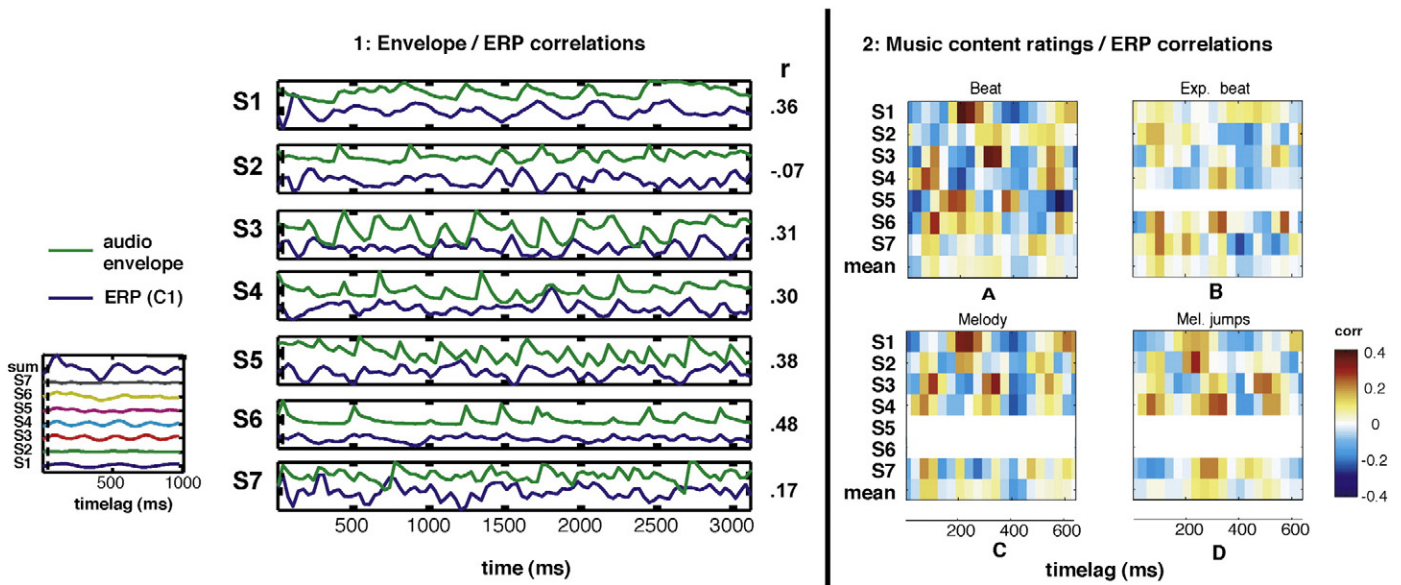
Fig. 7. In panel 1 (left), the audio envelopes are shown together with the ERP of component 1 for every stimulus, showing the correlation between the two on the right. The y-axes are scaled individually, all between 2 and 4 μV. The plot on the left below shows the total correlation per stimulus at different time-lags, with the sum as the first line. The summed correlation is maximal at a latency of 100 ms, and the plots on the top of panel 1 have been lagged at this interval to visualize the correlation. In panel 2 (right), the correlations are shown for every stimulus (S1–S7 and the mean) for four of the scored musical dimensions, at different time-lags (0–600 ms). The first two (A and B) relate to the time-structure: 'beat' describes the rhythmic structure (the rhythmic pulse), 'E-beat' a rhythmic surprise (syncopation). The other two plots (C and D) concern the melody: 'Melody' is identifies melodic events and 'Mel jumps' denotes large pitch jumps.

cortex (AAC) and supplementary motor cortex, points at other processing taking place, over and above registering loudness.

With the music perception literature in mind, in which many interpersonal differences in the EEG response have been identified that appear to be related to personal music preference (Caldwell and Riby, 2007), cultural background (e.g. Nan et al., 2006) as well as musical training (Koelsch et al., 1999; Van Zuijen et al., 2005; Vuust et al., 2005; Fujioka et al., 2004), the finding of a common ERP representation in the grand average, strong enough to allow cross-participant classification, was not necessarily expected, and we assume that low-level perceptual mechanisms are the main source of this common representation. However, as many brain responses to musical dimensions have also been shown to be independent of musical training, we believe this common representation to be augmented by implicit musical expectations (for a review on predictive processing of auditory information, see Winkler et al., 2009). This may explain certain aspects of the ERP that are not directly related to the audio envelope, such as the peak seen in the ERP of S4 at about 1800 ms (coinciding with a downbeat in the absence of a percussive event) or the ERP response to the latter half of S2, where a non-western tuning system was used (see Fig. 7.1, fourth and second row). Of course more standardized experiments need to be carried out to confirm these findings.

To summarize, we found it was possible to detect perceived music from the single-trial ERP, with 70% correct classification for our best participant, out of seven stimuli and using 3 s of data. It was also possible to detect the music stimuli across participants, supporting the notion of a universal representation for each stimulus. This shows the promise of EEG, next to fMRI, for brainreading paradigms, and the possibilities offered by music as an inherently time-based domain. Interesting steps for the future would be to investigate perception of spoken language, and the interaction between linguistic concepts and audio information.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2010.05.084.

References

Akhoun, I., Moulin, A., Jeanvoine, A., Ménard, M., Buret, F., Vollaire, C., Scoretti, R., Veuillet, E., Berger-Vachon, C., Collet, L., Thai-Van, H., 2008. Speech auditory brainstem response (speech ABR) characteristics depending on recording conditions, and hearing status: an experimental parametric study. J. Neurosci. Methods 175 (2), 196–205.
Atcherson, S.R., Gould, H.J., Mendel, M.I., Ethington, C.A., 2009. Auditory N1 component to gaps in continuous narrowband noises. Ear Hear. 300 (6), 687–695.
Bhattacharya, J., Petsche, H., Pereda, E., 2001. Interdependencies in the spontaneous EEG while listening to music. Int. J. Psychophysiol. 42, 287–301.
Bishop, C., 1995. Neural Networks for Pattern Recognition. Oxford University Press.
Brattico, E., Tervaniemi, M., Näätänen, R., Peretz, I., 2006. Musical scale properties are automatically processed in the human auditory cortex. Brain Res. 1117, 162–174.
Brechmann, A., Baumgart, F., Scheich, H., 2002. Sound-level dependent representation of frequency modulations in human auditory cortex: a low-noise fMRI study. J. Neurophysiol. 87, 423–433.
Brochard, R., Abecasis, D., Potter, D., Ragot, R., Drake, C., 2003. The "ticktock" of our internal clock: direct brain evidence of subjective accents in isochronous sequences. Psychol. Sci. 140 (4), 362–366.
Caldwell, G.N., Riby, L.M., 2007. The effects of music exposure and own genre preference on conscious and unconscious cognitive processes: a pilot ERP study. Conscious. Cogn. 16, 992–996.
Dekio-Hotta, S., Kojima, T., Karino, S., Yamasoba, T., Dekio, I., Ito, Y.M., Satake, H., Kaga, K., 2009. N1 component reflects difference of terminal chords in three-chord sequences. NeuroReport 200 (3), 251–256.
Ethofer, T., Van De Ville, D., Scherer, K., Vuilleumier, P., 2009. Decoding of emotional information in voice-sensitive codes. Curr. Biol. 19, 1028–1033.
Formisano, E., De Martino, F., Bonte, M., Goebel, R., 2008. "who" is saying "what"? brain-based decoding of human voice and speech. Science 322, 970–973.
Fujioka, T., Ross, B., Kakigi, R., Pantev, C., Trainor, L.J., 2004. Musical training enhances automatic encoding of melodic contour and interval structure. J. Cogn. Neurosci. 160 (6), 1010–1021.
Van Gerven, M., Farquhar, J., Schaefer, R.S., Vlek, R.J., Geuze, J., Nijholt, A., Ramsey, N., Haselager, P., Vuurpijl, L., Gielen, S., Desain, P., 2009. The brain–computer interface cycle. J. Neural Eng. 60 (4), 041001.
Halpern, A.R., Martin, J.S., Reed, T.D., 2007. An erp study of major–minor classification in melodies. Music Percept. 250 (3), 181–191.

Haynes, J.-D., 2009. Decoding visual consciousness from human brain signals. Trends Cogn. Sci. 130 (5), 194–202.

Jasper, H.H., 1958. The ten–twenty electrode system of the international federation. Electroencephalogr. Clin. Neurophysiol. 10, 371–375.

Jongsma, M.L.A., Desain, P., Honing, H., 2004. Rhythmic context influences the auditory evoked potentials of musicians and nonmusicians. Biol. Psychol. 660 (2), 129–152.

Kay, K., Naselaris, T., Prenger, R.J., Gallant, J.L., 2008. Identifying natural images from human brain activity. Nature 452, 352–355.

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

Krohn, K.I., Brattico, E., Valimaki, V., Tervaniemi, M., 2007. Neural representations of the hierarchical scale pitch structure. Music Percept. 240 (3), 281–296.

Longuet-Higgins, H.C., Lee, C.S., 1982. The perception of musical rhythms. Perception 11, 115–128.

Mayhew, S.D., Dirkx, S.G., Niazy, R.K., Ianetti, G.D., Wise, R.G., 2010. EEG signatures of auditory activity correlate with simultaneously recorded fMRI responses in humans. Neuroimage 490, 849–864.

Mulert, C., Jäger, L., Propp, S., Karch, S., Störmann, S., Pogarell, O., Möller, H., Juckel, G., Hegerl, U., 2005. Sound level dependence of the primary auditory cortex: simultaneous measurement with 61-channel EEG and fMRI. Neuroimage 28, 49–58.

Nan, Y., Knösche, T.R., Friederici, A.D., 2006. The perception of musical phrase structure: a cross-cultural erp study. Brain Res. 1094, 179–191.

Neuhaus, C., Knösche, T.R., Friederici, A.D., 2006. Effects of musical expertise and boundary markers on phrase perception in music. J. Cogn. Neurosci. 180 (3), 472–493.

Palmer, C., Jewett, L.R., Steinhauer, K., 2009. Effects of context on electrophysiological response to musical accents. Ann. NY Acad. Sci. 1169, 470–480.

Reaz, M.B.I., Hussain, M.S., Mohd-Yasin, F., 2006. Techniques of EMG signal analysis: detection, processing, classification and applications. Biol. Proced. Online 80 (1), 11–35.

Schaefer, R.S., Desain, P., Suppes, P., 2009. Structural decomposition of EEG signatures of melodic processing. Biol. Psychol. 82, 253–259.

Suppes, P., Lu, Z.H., Han, B., 1997. Brain-wave recognition of words. Proc. Natl Acad. Sci. 94, 14965–14969.

Suppes, P., Han, B., Lu, Z.H., 1998. Brain-wave recognition of sentences. Proc. Natl Acad. Sci. 95, 15861–15866.

Trainor, L.J., McDonald, K.L., Alain, C., 2002. Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. J. Cogn. Neurosci. 140 (3), 430–442.

Vuust, P., Pallesen, K.J., Bailey, C., van Zuijen, T., Gjedde, A., Roepstorff, A., Østergaard, L., 2005. To musicians, the message is in the meter: preattentive neuronal responses to incongruent rhythm are left-lateralized in musicians. Neuroimage 24, 560–564.

Winkler, I., Denham, S.L., Nelken, I., 2009. Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends Cogn. Sci. 130 (12), 532–540.

Van Zuijen, T.L., Sussman, E., Winkler, I., Näätänen, R., Tervaniemi, M., 2005. Auditory organization of sound sequences by a temporal or numerical regularity—a mismatch negativity study comparing musicians and nonmusicians. Cogn. Brain Res. 23, 270–276.