EEG-based decoding and recognition of imagined music

Marthe Tibo[1], Simon Geirnaert[1,2], and Alexander Bertrand[1,*]

[1]KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics
[2]KU Leuven, Department of Neurosciences, Research Group ExpORL
[*]alexander.bertrand@kuleuven.be

ABSTRACT

When listening to music, the brain generates a neural response that follows the amplitude envelope of the musical sound. Previous studies have shown that it is possible to decode this envelope-following response from electroencephalography (EEG) data during music perception. However, successful decoding and recognition of imagined music, without the physical presentation of a music stimulus, has not been established to date. During music imagination, the human brain internally replays a musical sound, which naturally leads to the hypothesis that a similar envelope-following response might be generated. In this study, we demonstrate that this response is indeed present during music imagination and that it can be decoded from EEG data. Furthermore, we show that the decoded envelope allows for classification of imagined music in a song recognition task, containing tracks with lyrics as well as purely instrumental tracks. A two-song classifier achieves an accuracy up to 95.3% (vs. 53.3% significance threshold), while a 12-song classifier achieves an accuracy up to 66.7% (vs. 15% significance threshold), where the reported accuracies are medians across 10 subjects. The results of this study demonstrate the feasibility of decoding imagined music, thereby setting the stage for new neuroscientific experiments in this area as well as for new types of brain-computer interfaces based on music imagination.

Introduction

Every day, our brains are exposed to millions of different sounds, ranging from simple tone beeps to highly complex stimuli, like music. When musical sounds are picked up by the human ear, signals travel via the auditory nerve to various brain areas, where a complex chain of processing events occurs. These processes can, for example, make us emotionally react to music[1], bring back memories[2,3], or let us play along to the song on an instrument[4].

During listening to music - termed here as *music perception* - the human brain is capable of processing complex acoustical content. However, the brain has the ability to perform an even more remarkable task. Our minds can imagine complex sounds without the presence of a physical source of sound. An 'earworm', a song which is stuck inside our heads for a long time[5,6], is an example of this so-called music imagination phenomenon.

However, a lot about the specific cortical processing and spatial activation during the perception and imagination of music remains unknown. To shed more light on these functions, the perception and imagination processes can be summarized into models, which define the relation between an auditory stimulus and its neural response. By designing these models, one can reconstruct the amplitude envelope of a sound stimulus from the neural responses to this stimulus or vice versa.

Multiple previous studies have created such models for neural decoding of *perceived* music. Di Liberto et al.[7] reconstructed stimulus envelopes out of EEG responses during music perception experiments with musicians and non-musicians to explore possible neural differences between both groups. Sturm et al.[8] decoded audio power slopes from EEG responses and investigated their relation to behavioral tension data. Gang et al.[9] computed decoders that maximize the correlation between the EEG and music stimulus envelopes using canonical correlation analysis, while Kaneshiro et al.[10] quantified stimulus-response and inter-subject correlations on EEG data, filtered by reliable component analysis. Vlek et al.[11] performed an EEG classification to probe perception and imagery of beat sequences. Moreover, Ofner et al.[12] used a deep generative model to reconstruct short perceived and imagined rhythmical patterns with support from a continuous metronome click.

Furthermore, some studies have focused on classifying songs based on EEG data from music perception experiments. Schaefer et al.[13] have shown that it is possible to distinguish between seven distinct songs during music perception by using a quadratically regularized linear logistic regression classifier on EEG data. Stober et al.[5,14] conducted classification experiments for 12 songs during music perception, using a similarity-constraint encoding approach. Apart from working on perception data, the latter also concentrated on the EEG responses of imagination experiments. However, no consistent patterns were found in the data and an insignificant classification accuracy was obtained.

In this paper, we provide evidence that this EEG-based decoding and song classification is also possible for *imagined* music,

which up to our knowledge has not been established to date. We adopt a linear stimulus reconstruction approach as a prior step to perform song classification, which is borrowed from the neighboring domain of neural decoding of speech. In the field of speech processing, linear decoding techniques achieve good results with regards to the auditory attention decoding problem, such as in the studies of O'Sullivan et al.[15], Fuglsang et al.[16], Wong et al.[17], Das et al.[18], and Geirnaert et al.[19]. While such decoding strategies have been successfully applied for music perception[7–9,13,20–22], they have not been successful in decoding imagined music so far[14]. In this work, the decoder is trained on EEG data of a music perception experiment and then applied to EEG data from a music imagination experiment. For the latter, the decoder is complemented with a delay operator that searches for the (unknown) start of the imagined music, leading to an improved quality of the obtained reconstructions. Finally, these reconstructed envelopes will be labeled by classifiers w.r.t. various properties, obtaining a significant classification accuracy in both perception and imagination experiments. In essence, this study shows a way to successfully classify imagined music (or possibly other stimuli), paving the way for new software tools for analyzing neural activity during music imagination and new brain-computer interface paradigms. In the long term, this could even lead to new futuristic applications such as hands-free and voice-free song selection, song recognition, a 'silent' version of the popular Shazam app, etc.

In the next sections, we first give an overview of the data used in this study, followed by the applied preprocessing steps. Afterwards, the adapted linear decoding model is introduced in detail, together with its modified decoding strategy. Subsequently, we will introduce various classification tasks, which are discussed and analyzed for both the perception and imagination of music, where the former acts as a baseline.

Methodology

Data

Two publicly available data sets are used, containing the EEG responses on musical stimuli for various subjects. Both data sets were recorded using a 64-channel BioSemi Active Two system at a sampling frequency of 512 Hz.

OpenMIIR data set

The OpenMIIR data set, published by Stober et al.[5], contains 12 stimuli with polyphonic music, consisting of 4 nursery rhymes with a singing voice and musical accompaniment, the same 4 nursery rhymes without singing voice, and 4 well-known orchestral works (8 out of 12 stimuli are purely instrumental without singing voice). Inclusion of the same songs, once provided with and once without lyrics, could indicate whether the lyrical content of these songs is an essential factor in its recognition by the used method. The EEG responses to each of these stimuli were recorded in 10 subjects (8 musically trained subjects, 2 nonmusicians) during four different experiments:

- music perception, preceded by cue clicks (with the frequency equal to the beat of the corresponding music stimulus);
- music imagination, preceded by cue clicks;
- two music imagination experiments without cue clicks.

For the last two imagination experiments, one set-up was recorded in the same block as the previous ones, while another was recorded as a standalone trial. These standalone trials were registered for all music stimuli immediately after all recording blocks of the previous three conditions were performed (approximately 30 minutes after starting the first recording block). Five repetitions of each music presentation/imagination are available per subject, summing to 60 trials per subject and experiment. The music stimuli in this data set have an average length of 10.4 s.

Bach data set

This data set, published by Di Liberto et al.[20], contains 10 excerpts from the partitas and sonatas of J.S. Bach and their EEG responses for 20 subjects (10 musicians, 10 nonmusicians). All stimuli are purely instrumental and played from a MIDI file with piano samples. This data set only contains a perception EEG experiment, meaning that these trials will solely be used to investigate the effect of the modified decoding strategy on perception data. Each of these stimuli is presented three times to a subject, leading to a total of 30 trials. The average length of the excerpts is 120 s.

Preprocessing

First, the EEG is re-referenced to the average of the two recorded mastoid channels. For the OpenMIIR data set, however, this step is omitted since the mastoid channels were only recorded for half of the subjects[5]. Instead, we retained the original reference (common mode sense active electrode) on all EEG data (incl. mastoid channels). Afterwards, a multi-channel Wiener filter[23] removes eye blinks and other disturbances from the data. After this step, the clean EEG channels are filtered using a first-order Butterworth filter with a frequency range between 1 and 10 Hz. This choice is inspired by research in speech processing, where it was shown that the envelope and its EEG response are linearly relatable over the aforementioned frequency range[15,17,18]. As a final step, the EEG is downsampled to 64 Hz and standardized.

For each music stimulus, an envelope is extracted based on the procedure described in Biesmans et al.[24]. In short, the signal is split into 15 subbands using a gammatone filter bank, with center frequencies ranging from 1 Hz to 5 kHz. The amplitude

envelope of each subband signal is compressed using a power law with the exponent equal to 0.6. Afterwards, the envelopes are summed across the subbands, resulting in a single aggregated envelope. Finally, the aggregated envelope is filtered between 1–10 Hz using the same bandpass filter as in the EEG preprocessing, subsequently downsampled to 64 Hz and standardized.

Decoding and classification

Stimulus reconstruction

To reconstruct an envelope from the corresponding EEG response, a linear decoding model is used. Consider a music stimulus envelope $s(t)$ and its N-channel EEG response $\mathbf{r}(t) = \begin{bmatrix} r_1(t) & \cdots & r_N(t) \end{bmatrix}^T$, here represented as an N-dimensional vector, where $r_n(t)$ represents the t-th sample in the n-th EEG channel, and where $(\cdot)^T$ denotes the transpose operator. The decoder then linearly combines time-lagged versions of the EEG channels to estimate the envelope $\hat{s}(t)$[15]:

$$\hat{s}(t) = \sum_{n=1}^{N} \sum_{\tau=0}^{L-1} g_n(\tau) r_n(t+\tau). \tag{1}$$

where $g_n(\tau)$ denotes the decoder weight for the n-th channel on the τ-th time lag. In this model, we choose a range of time lags up to 250 ms, which is inspired by the speech decoding literature[15]. At a sampling rate of 64 Hz, this corresponds to up to $L = 17$ sample shifts (incl. the zero shift). However, this model omits an important parameter for music imagination. During these experiments, the exact onset of the imagined music is unknown, resulting in a potential mismatch between the decoded music envelope and the baseline envelope. To counteract this effect, one could choose to broaden the range of time lags. However, this introduces a large number of extra parameters into (1), resulting in overfitting effects and ill-conditioned matrix inversions in the calculation of the decoder (see below). Therefore, we include an extra global time shift ν into the model, independent of the time lags τ, which globally shifts the onset of the EEG signal:

$$\hat{s}_\nu(t) = \sum_{n=1}^{N} \sum_{\tau=0}^{L-1} g_n(\tau) r_n(t+\tau+\nu). \tag{2}$$

Since the time delay ν between the actual music envelope and the response generated by the brain is unknown during (uncued) music imagination, an optimal value for ν has to be estimated in order to maximize the neural decoding performance. We will return to this issue later on (see 'Testing the decoder'). Finally, we note that the decoder is modeled as a linear time-invariant system, which implies that the reconstructed envelope $\hat{s}(t)$ in (2) is equal to a delayed version of the one in (1), i.e., $\hat{s}_\nu(t) = \hat{s}(t+\nu)$. Therefore, in the remainder of this paper, we will omit the subscript ν in $\hat{s}_\nu(t)$ and will simply write $\hat{s}(t+\nu)$ when referring to the delayed version.

Training the decoder

The decoder is trained on the music *perception* trials, both for the perception *and* imagination case. In this case the global delay parameter ν can be set to $\nu = 0$, since the neural activity then entrains to the stimulus with only a short neural path delay, which is covered by the time shifts represented by τ[15]. The decoder weights are optimized such that they minimize the squared error between the original music envelope $s(t)$ and the reconstructed envelope $\hat{s}(t)$ in (1) on a training set of T samples:

$$\min_{\mathbf{g}} \sum_{t=1}^{T} (s(t) - \hat{s}(t))^2 + \lambda \|\mathbf{g}\|_2^2 \tag{3}$$

where $\mathbf{g}$ is an NL-dimensional vector in which all decoder weights $g_n(\tau)$ are stacked (for all channels n and time shifts τ). The second term is a regularization term to counteract overfitting effects[15,17], where λ is a tunable regularization weight. (3) can be rewritten in matrix-vector format as:

$$\min_{\mathbf{g}} \|\mathbf{s} - \mathbf{R}\mathbf{g}\|_2^2 + \lambda \|\mathbf{g}\|_2^2, \tag{4}$$

where $\mathbf{s} = [s(1) \ldots s(T)]^T$ and $\mathbf{R}$ is a $T \times NL$ matrix containing all N EEG channels plus their $L-1$ time-lagged copies in its columns:

$$\mathbf{R} = \begin{bmatrix} \mathbf{r}(1)^T & \mathbf{r}(2)^T & \cdots & \mathbf{r}(L)^T \\ \mathbf{r}(2)^T & \mathbf{r}(3)^T & \cdots & \mathbf{r}(L+1)^T \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{r}(T-1)^T & \mathbf{r}(T)^T & \cdots & 0 \\ \mathbf{r}(T)^T & 0 & \cdots & 0 \end{bmatrix}. \tag{5}$$

The least squares decoder is then found as

$$\mathbf{g} = \left(\mathbf{R}^T\mathbf{R} + \lambda\mathbf{I}\right)^{-1} \left(\mathbf{R}^T\mathbf{s}\right) = \left(\mathbf{C}_{RR} + \lambda\mathbf{I}\right)^{-1} \mathbf{C}_{Rs}, \tag{6}$$

where $\mathbf{I}$ denotes the identity matrix, and $\mathbf{C}_{RR}$ and $\mathbf{C}_{Rs}$ are respectively the estimated autocorrelation matrix of the EEG data and the estimated cross-correlation vector between the EEG data and the stimulus envelope. The decoder is always trained on a per-subject basis. The regularization hyperparameter λ is tuned based on a leave-one-song-out cross-validation procedure on the training set over a grid defined by $\lambda = 1.848^k \times 10^{-6}$ with $k = [0,50]$[17] relative to the autocorrelation matrix $\mathbf{C}_{RR}$. The value of λ that leads to the highest Pearson correlation coefficient $\rho(s(t),\hat{s}(t))$ between the reconstructed and the original envelope (averaged across all folds) is selected as the optimal regularization parameter, after which the decoder is retrained once more on all training data using this parameter.

Testing the decoder

When decoding imagined music, the global shift parameter ν cannot be set equal to 0 due to an unknown delay between the imagined/decoded music track and the original envelope of the corresponding audio track. This delay parameter ν is different in each trial and is therefore estimated on a per-trial basis. To this end, we apply the pre-trained decoder over a range of values for ν and select the value that yields a reconstructed envelope $\hat{s}(t+\nu)$ that has the highest Pearson correlation coefficient with the original stimulus $s(t)$, i.e.,

$$\nu_{\text{opt}} = \underset{\nu}{\text{argmax}}\ \rho(s(t),\hat{s}(t+\nu)). \tag{7}$$

The shift $\nu = \nu_{\text{opt}}$ then denotes the global shift at which the reconstruction is maximally similar to the real envelope. The decoding performance is quantified by the Pearson correlation coefficient $\rho(s(t),\hat{s}(t+\nu_{\text{opt}}))$ between the reconstructed and original envelope of the test trial. While the global shift parameter ν was introduced to cope with the unknown time offset when decoding music imagination, we also include it here for the testing of decoding music perception to allow for a fair comparison between both cases. This is because the omittance of the additional per-trial optimization in (7) might introduce an unfair disadvantage for the music perception case.

When testing the decoder performance for the perception case, a nested cross-validation scheme on the EEG trials is used, where in each fold, a single music track (incl. all repeated perception trials) is used as the test set, while all repeated perception trials of all other music stimuli of the same subject are used as training set. The delay parameter ν is varied between 0–2 s for the perception experiment in the OpenMIIR data set and 0–1 s for the Bach data set, since only 1 second of post-stimulus EEG is available, thereby limiting the range of ν to at most 1 second.

For the decoding experiments in case of music imagination, the training data consists of the repetitions of all music tracks from the perception experiment, while the test set consists of the trials from the imagination experiment. In other words, we train our decoder on EEG during music perception, while testing it on EEG during music imagination. We test all music imagination conditions in the OpenMIIR data set, i.e., with and without cue clicks, where the delay parameter ν is again varied between 0–2 s. With this range, it is thus assumed that the subjects respond within 2 s of the starting mark.

Classification of songs

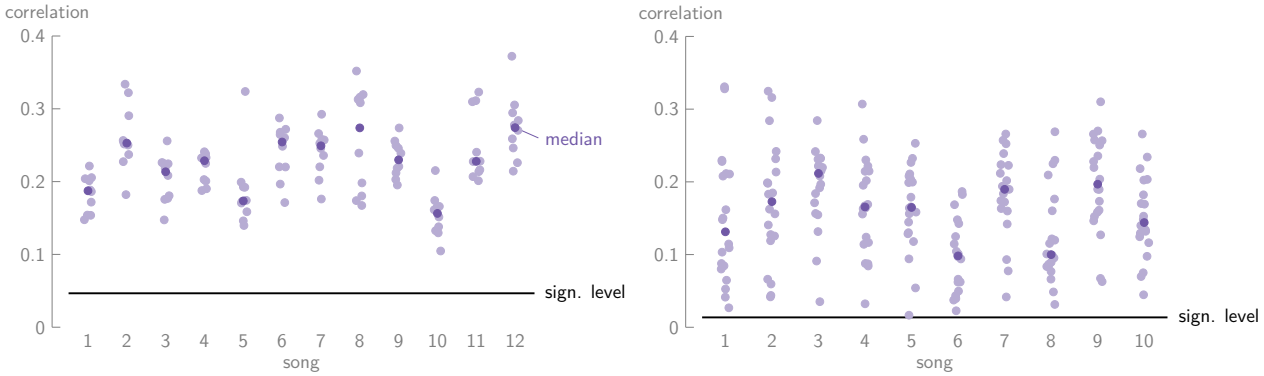
The main goal of this work is to identify with which music track a decoded music envelope corresponds. To this end, for each single trial, the decoded envelope $\hat{s}(t)$ is compared to each music envelope $s_i(t)$ in the database of music tracks or songs ($i = 1,\ldots,12$ for the OpenMIIR data set, and $i = 1,\ldots,10$ for the Bach data set). The maximal correlation coefficient ρ_i is identified for each music track i as

$$\rho_i = \max_{\nu} \rho(s_i(t),\hat{s}(t+\nu)). \tag{8}$$

For each single trial, the perceived/imagined song can then be detected by selecting the music track i for which ρ_i is maximal. It is noted that the quality of the estimated Pearson correlation coefficient depends on the number of available time samples over which it is estimated. If more time samples are available, the estimation noise on ρ_i will be lower, thereby becoming a more discriminative feature for classification. This introduces an important tradeoff between the decision window length or required recording time on the one hand and the accuracy of the classification on the other hand. A similar tradeoff exists in the context of auditory attention decoding on speech[19,25]. In this paper, Pearson correlation coefficients are always computed over the full trial length, unless specified otherwise.

We will consider three different classification experiments; a two-song classifier, a K-song classifier (with K the number of available songs in the respective data set), and a music category classifier, where the latter distinguishes between the three song categories in the OpenMIIR data set[14]. The latter could reveal if (the omission of) lyrical content leads to vastly different reconstructed music envelopes and whether this affects the classification results.

Note that the classification is based on *single-trial* decoding, i.e., multiple trials of the same song are not averaged but decoded and classified individually. Per subject, the accuracy is computed as the percentage of correctly classified reconstructed envelopes across all trials and involved songs. For the two-song classifier, the final accuracy of a subject is then computed by averaging all accuracies across all combinations of song pairs.



(a) OpenMIIR data set

(b) Bach data set

Figure 1. Music perception: Median correlations for each subject (light purple dots) and song in both data sets. Darker colored dots display the median over all subjects.

Periodicity

As an additional analysis, we will also investigate the Pearson correlation coefficient $\rho(s(t), \hat{s}(t+\nu))$ as a function of the time shift ν to explore possible periodical behavior. If such a periodical behavior is present, we will investigate whether it relates to the tempo of the music (typically measured in beats per minute or BPM), which could reveal beat or tempo tracking effects.

However, the observation of periodical behavior of the Pearson correlation coefficient may be an artifact induced by the periodical nature of the original stimulus, which also influences the correlation coefficient $\rho(s(t), \hat{s}(t+\nu))$. To investigate this phenomenon of beat or tempo tracking, the leakage effect from the periodicity of the real envelopes must first be ruled out. To this end, shorter excerpts of the real audio envelopes, equal to the length of two song periods, are cut out and cross-correlated with the corresponding decoded envelopes. If a periodical pattern is observed for more than two song periods in this cross-correlation, this behavior is not only caused by the real envelope, but by a combined oscillatory behavior of both the neural response and the original music envelope. If this periodicity repeats at a frequency similar to the beat of the song, this would mean that the human brain also registers the tempo (BPM) in the neural processing pathway.

Statistical analyses

The significance of the Pearson correlations is determined by a permutation test in which 10,000 randomly time-permuted reconstructed envelopes are correlated with their corresponding real envelopes[26], where the sample-wise random permutation destroys the correlation structure. The median correlation (across subjects and repetitions) is computed and the 97.5 percentile of these median correlations is used as the significance level. To perform a pairwise comparison of Pearson correlations across experiments or subject groups (musicians vs. nonmusicians), we use a Wilcoxon signed-rank test and Wilcoxon rank-sum test respectively. The 5% significance level for classification accuracies is computed from a binomial distribution, which determines the distribution of the observed accuracy of a random classifier given a number of trials per subject (60 for the OpenMIIR data set, 30 for the Bach data set) and the succes rate of correctly labeling one trial (1/12 for the OpenMIIR data set, 1/10 for the Bach data set)[14,15]. The 95 percentile of this 'random classifier' distribution is then used as the significance level, leading to a significant threshold of 15% and 20% for the OpenMIIR and the Bach data set, respectively. For the song category classifier, a similar procedure leads to a significance threshold of 43.3% (OpenMIIR data set), while for the two-song classifier, thresholds of 53.3% (OpenMIIR data set) and 63.3% (Bach data set) are found.

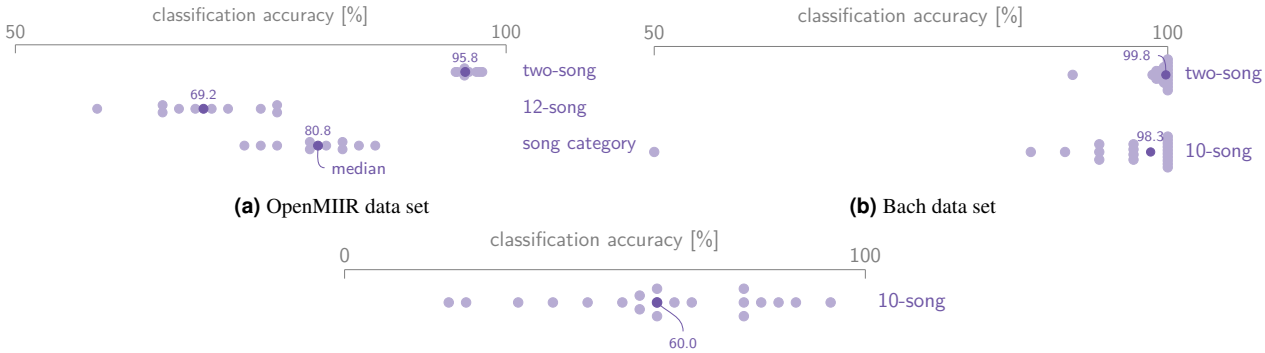
Results and discussion

Music perception

First, we assess the performance of our linear decoding model on the music perception experiments. For both data sets, Figure 1 shows the median (across repeated trials) Pearson correlation between the optimal reconstructed envelopes and their corresponding real envelopes for each subject. Using the permutation test, we see significant median Pearson correlations for the perception experiment over all stimuli (songs) in both data sets. The median over all subjects and songs is equal to 0.21 for the OpenMIIR data set and 0.16 for the Bach data set. These correlations are significantly improved in comparison to the Pearson correlations without the inclusion of a time shift ν into the model (Wilcoxon signed-rank test, $p \ll 10^{-4}$). Note that de Cheveigné et al.[27] found a similar positive influence of a delay parameter on the correlation between the reconstructed and actual envelope in the context of decoding perceived speech. No difference between the correlations of musicians and nonmusicians was found (Wilcoxon rank-sum test, $p > 0.2$).

Two-song classifier

In this task, we investigate whether it is possible to distinguish the reconstructed envelopes of two stimuli from each other, which corresponds to a match-mismatch task. Therefore, we correlate each reconstructed envelope with the real envelopes of two stimuli, of which one of both is matched to the corresponding reconstructed envelope. We select the highest correlation



(a) OpenMIIR data set

(b) Bach data set

(c) Bach data set: 10 s decision window

Figure 2. Music perception: Classification accuracy for each subject (light purple dots) during the two-song, K-song, and song category classification tasks. The darker colored dots represent the median accuracy over all subjects.

out of this duo, which is ideally linked to the matching real envelope. Figure 2 shows the mean accuracy of this classification task over all possible stimulus pairs for each subject. The overall median accuracy over all subjects is equal to 95.8% for the OpenMIIR data set and 99.8% for the Bach data set. The seemingly high results for this classifier prompts us to explore more complex strategies in the next subsections.

K-song classifier

Secondly, we investigate whether the previous strategy can be extended to discern K stimuli from each other, where K is the total number of songs in each data set ($K = 12$ for the OpenMIIR data set, and $K = 10$ for the Bach data set). The accuracy per subject for this K-song classifier is shown in Figure 2. The median classification accuracy over all subjects is 69.2% for the OpenMIIR data set and 98.3% for the Bach data set. The classification accuracy was significantly higher than chance level for all subjects in both data sets (thresholds of 15% and 20% for Open MIIR and Bach, respectively). We observe an overall higher performance for the Bach data set, which can be attributed to the smaller number of classes (12 vs 10) and longer trials leading to more accurate Pearson correlation estimates. To investigate the influence of trial length on the K-song classification task, this set-up was repeated using only the first 10 s of the reconstructed and real stimulus as decision window for classification, comparable to the OpenMIIR data set, leading to a median accuracy of 60.0% (Figure 2c). This implies that the high performance on the Bach data set is attributed to the longer trials over which the correlation is calculated.

Song category classifier

In a third classification task, the stimuli of the OpenMIIR data set will be labeled as one of three music categories: nursery rhymes with lyrics (category 1), the same nursery rhymes without lyrics (category 2), and orchestral music (category 3). The per-subject accuracies are shown in Figure 2a. For all subjects, the accuracy is significantly better than chance (threshold of 43.3%), with a median over all subjects of 80.8%. As expected, the accuracy for each subject is higher than for the K-song classification, but lower than for the two-song classification. This classifier succeeds at classifying the same songs with and without lyrics, suggesting a discriminative difference between the reconstructed envelopes of both music categories.

Music imagination

Secondly, we assess the decoding performance during music imagination. We first test the decoder without applying a global time shift, i.e., $\nu = 0$. It was found that this leads to non-significant and even negative median correlations across subjects and songs for all imagination experiments, based on the previously defined permutation test. Similarly, performing the song category classification task on the imagined music overall leads to low, non-significant accuracies: median accuracies across subjects equal to 29.1% (cued imagination), 33.0% (uncued imagination), and 37.0% (standalone experiment) were found. The significance threshold (43.3%) is surpassed by one subject during cued imagination (45%) and the standalone experiment (47%). For the 12-song classification task, the median accuracies are equal to 5.8% (cued imagination), 6.7% (uncued imagination), and 11.7% (standalone experiment). For the latter, three subject accuracies are equal to the significance threshold (15%), all other accuracies are non-significant. We conclude that decoding imagined music does not work optimally if there is no compensation mechanism for the (unknown) onset mismatch of the imagined music, even during cued experiments. Similar problems were encountered by Stober et al.[14] in a study on the same data, leading to non-significant results for music imagination. Also in this study, this outcome was attributed to an unsuccessful EEG-stimulus synchronization and investigation of data patterns.

Therefore, we repeat the same process using an optimal time shift $\nu = \nu_{\text{opt}}$ for each of the three imagination experiments. Figure 3 shows the median (across repeated trials per subject) Pearson correlation between the optimally reconstructed envelopes and their corresponding real envelopes during the cued imagination experiment. The median correlation over all subjects and songs is equal to 0.20 for the cued music imagination, 0.19 for the uncued music imagination, and 0.21 for the standalone experiment. Interestingly, the obtained correlations are all significant, meaning that the envelope reconstruction benefits from the inclusion of the time shift ν, confirming the hypothesis of a possible mismatch between the trial and imagination onsets.

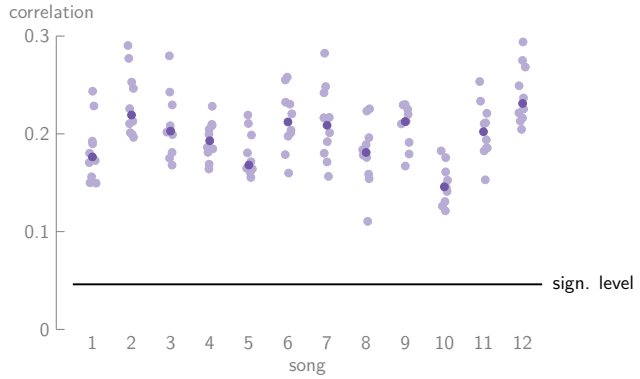


Figure 3. Music imagination: Median correlations (across repeated trials) for each subject (lighter dots) and song during cued music imagination for the model with optimal time shift. Darker colored dots display the median over all subjects.

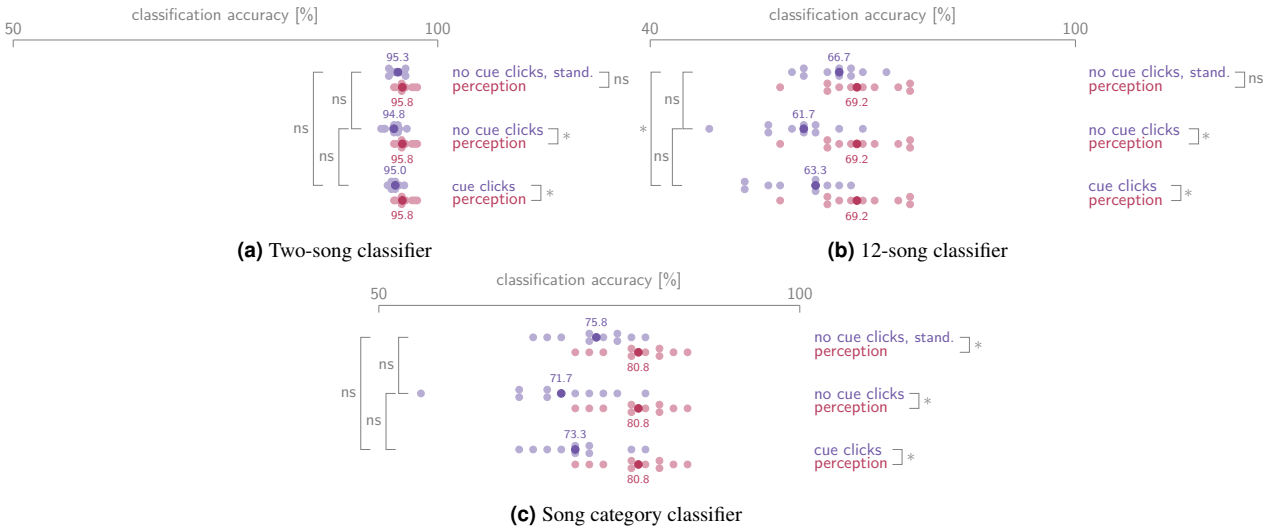


Figure 4. Music imagination: Classification accuracy for each subject (light purple dots) during the two-song, 12-song, and song category classification tasks, compared to the perception case (red dots). The darker colored dots represent the median accuracy over all subjects. * refers to a significant difference on the 0.05-level (Wilcoxon signed-rank test), ns to a non-significant difference.

Vlek et al.[11] showed similarities in the brain signatures during music perception and imagination, concluding that both have similar processing mechanisms, which is used in the current decoding strategy.

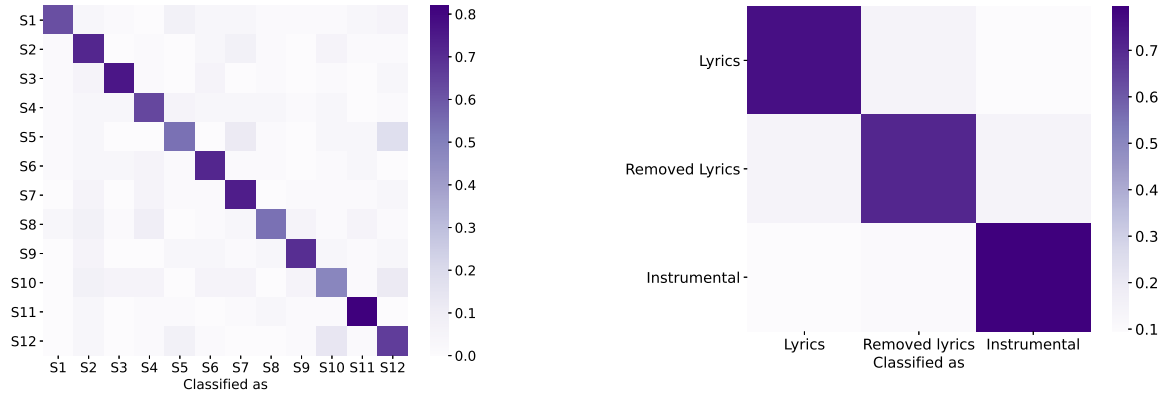
However, as expected, these correlation values are lower than for the perception experiment in the OpenMIIR data set (Wilcoxon signed-rank test, $p < 10^{-3}$), indicating possible distortions in the imagined stimuli (e.g., tempo inconsistencies), which lead to a reduced similarity between the reconstructed and real envelopes during testing. Moreover, this observation could be signaling differences between the neural processing of perceived and imagined music. No differences were found between musicians and non-musicians (Wilcoxon rank-sum test, $p > 0.2$).

Two-song classifier

When applying the two-song classifier to our imagination data, we see a high mean accuracy over all possible song categories for every subject in Figure 4a. The overall median over all subjects is 95.0% for the cued, 94.8% for the uncued, and 95.3% for the standalone imagination experiment. Comparing these accuracies to the results during music perception, the latter significantly outperforms the music imagination experiments (Wilcoxon signed rank test, $p < 0.0039$), except for the standalone imagination experiment ($p = 0.055$). This standalone condition, which is recorded last for all subjects, might be influenced by better memorization of the music stimulus across the experiments. No significant differences were found between the three imagination experiments.

12-song classifier

In Figure 4b, the 12-song classification accuracy is shown for each subject. For each of the three imagination experiments, the results are significant (threshold of 15%). The median accuracy over all subjects is 63.3% for the cued imagination, 61.7% for the uncued imagination, and 66.7% for the standalone experiment. Comparing these accuracies to the results during music perception, the latter significantly outperforms the accuracies obtained for the music imagination experiments (Wilcoxon signed-rank test, $p < 0.016$), except for the standalone imagination experiment ($p = 0.23$), similar to what was observed in



(a) 12-song classifier

(b) Song category classifier

Figure 5. Music imagination: Confusion matrices for the 12-song classifier and the song category classifier during the standalone imagination experiment. Each row of the matrix is normalized to sum to 1. We see a good overall classification, even for the same songs with/without lyrics or with similar tempos (Table 1).

the two-song classification task. Furthermore, a significant difference between the cued and standalone imagination trials is found. Figure 5a shows the confusion matrix of the 12-song classification task during the standalone imagination experiment. Looking at Table 1, even highly similar stimuli with (almost) equal tempos are significantly distinguished. This indicates that the classification is based on more than only the beat tempo. Furthermore, it seems that there are no more errors between the same songs with and without lyrics (and thus with the same tempo) compared to other songs.

Song category classifier

In Figure 4c, the accuracy values of the song category classifier per subject are displayed, which are significant for all imagination conditions (threshold of 43.3%). The median accuracy over all subjects is 73.3% for the cued imagination, 71.7% for the uncued imagination, and 75.8% for the standalone experiment. Comparing these accuracies to the results during music perception, the latter significantly outperforms all music imagination experiments (Wilcoxon signed-rank test, $p < 0.04$). Between the imagination experiments, no significant differences were found. Figure 5b shows the confusion matrix for the song category classifier during the standalone imagination experiment. Similar to the 12-song classification task, these results show that even for highly similar stimuli (nursery rhymes with and without lyrics) that have exactly the same tempo (Table 1), the song classifier can determine the right category based on brain responses. In this confusion matrix, songs with and without lyrics are clearly distinguished from each other and from instrumental tracks, showing that the addition or omission of lyrics also changes the reconstructed music envelopes.

Category	**Stimulus ID (Tempo [BPM])**			
Lyrics	S1 (212)	S2 (189)	S3 (200)	S4 (160)
Removed lyrics	S5 (212)	S6 (189)	S7 (200)	S8 (160)
Instrumental	S9 (178)	S10 (166)	S11 (104)	S12 (140)

Table 1. Music stimulus properties (OpenMIIR data set).

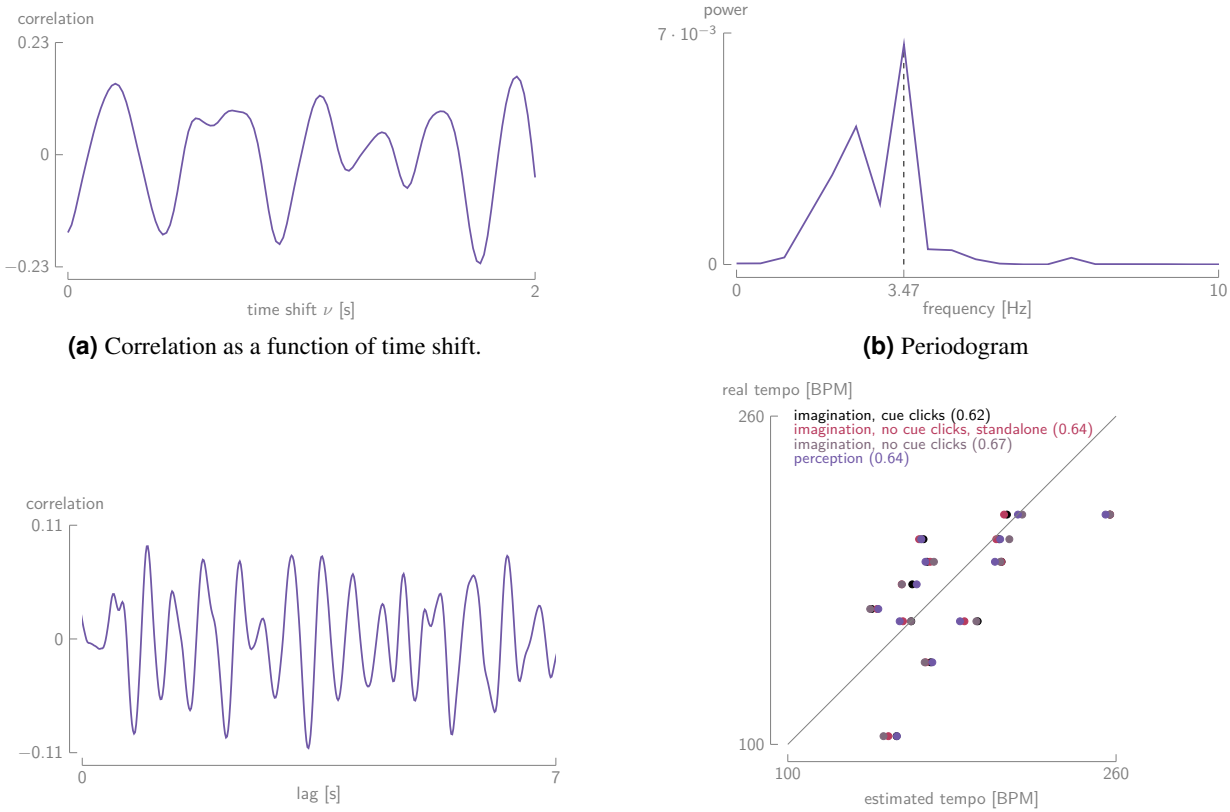
Periodicity

When plotting the Pearson correlations $\rho(s(t), \hat{s}(t+\nu))$ as a function of ν per trial, a seemingly periodical or oscillatory behavior arises, as demonstrated in Figure 6a. The periodogram in Figure 6b sheds more light on the frequency of this periodical pattern, with a peak around the beat frequency of the stimulus. This suggests that the neural response has a periodical behavior, possibly following the tempo of the music.

To rule out the aforementioned effect of the original stimulus on this periodical behavior, we investigate the cross-correlation between the decoded envelope and a two-period excerpt of the same stimulus, as explained in the methodology section. In Figure 6c, we see a similar pattern of minima and maxima extending over more than two stimulus periods, confirming that the exhibited periodical pattern is also present in the neural activity generated by the imagined music.

To further investigate this phenomenon, the median distance between two peaks in these cross-correlations is extracted, leading to an estimated beat frequency per trial. The median beat frequency over all trials and subjects is subsequently correlated with the real tempi of the stimuli. Figure 6d shows that these tempo estimates are highly similar to the real tempi of the corresponding stimuli. Using a t-test, significant correlations with the real tempo are found for the three imagination experiments of the OpenMIIR data set (cue clicks: correlation = 0.62, $p = 0.032$, no cue clicks: correlation = 0.67, $p = 0.018$, standalone: correlation = 0.64, $p = 0.024$).

These observations of periodicity during imagined music are in line with previous observations during music perception. Previously, Nozaradan et al.[28] revealed a neural entrainment in EEG responses during music perception while subjects imagined additional beats onto the rhythm - which were also present in the EEG response. Spectral analyses during music perception by



(a) Correlation as a function of time shift.

(b) Periodogram

(c) Cross-correlation reconstructions - two periods envelope.

(d) Real vs. estimated tempo

Figure 6. (a) The correlation as a function of time shift ν for a representative trial and subject of the OpenMIIR data set, for the standalone imagination experiment (no cue clicks). (b) Periodogram of the correlation shown in subfigure (a) , showing a peak (3.47 Hz) close to the beat frequency of the stimulus (3.33 Hz). (c) The cross-correlation of the reconstruction with two periods of the envelopes, for the same representative trial, shows a similar periodical pattern. (d) Plot of the real vs. estimated tempo for each of the twelve songs per perception/imagination experiment. One dot represents the tempo of one song in one of the four experiments. The black line depicts the bisecting diagonal (corresponding to correlation $\rho = 1$) as a reference.

Kaneshiro et al.[10] and Gang et al.[9] revealed distinct peaks at frequencies related to the beat frequency - i.e., the beat itself and its harmonics. Moreover, Tal et al.[29] conducted an MEG experiment using drum patterns as stimuli to uncover that neural entrainment, in the case of music perception, is at least partially internally produced instead of stimulus-driven. A phase-locking to the beat is created in the brain even if some beats are missing. A similar periodical pattern is now also observed for entirely imagined music, as shown in Figure 6.

Conclusions and outlook

In this study, we have demonstrated the possibility to decode and recognize imagined music. Our methodology employs a linear neural decoding model, as traditionally used for speech and perceived music[7,15,17,21,24,27]. We have shown that a similar linear neural decoding paradigm can also be used to decode imagined music if a time shift parameter ν is introduced, which controls for the unknown onset of the imagined song, and which can be automatically extracted from the data. The three classification tasks - a two-song, K-song, and song category classifier - are based on the correlation between the real music envelopes and the EEG-based reconstructed versions obtained via the linear decoding model. While the standard decoding model (with $\nu = 0$) results in non-significant classification accuracies, the results improve drastically when using the time-shifted version of the model, resulting in significant classification accuracies. Surprisingly, we have demonstrated that all imagined music tracks in a data set can be distinguished from each other with a relatively high accuracy based on single-trial EEG data. This is also true for highly similar stimuli with (almost) equal tempos. Furthermore, even the same songs, once with and once without lyrics, could correctly be identified with a high accuracy. This implies that the speech-like modulations of the envelope can be discriminated from the purely instrumental modulations in the envelope. This is compatible with speech decoding studies where it was indeed shown that a similar linear decoding technique is able to reconstruct the envelope of (perceived) speech[15,17,24,26]. The possibility to discriminate between the same *imagined* song with and without lyrics might then suggest that also imagined speech could possibly be decoded with similar techniques, which could be an interesting question for future research.

The classification accuracies for imagined music were slightly lower than for music perception on the same music stimuli (63.3%, 61.7%, and 66.7% vs. 69.2% respectively on the 12-song classification task). We hypothesize that this is caused by mismatches between the original music envelope and the imagined music, for example, due to tempo variation. Other (more advanced) similarity metrics, robust to such tempo variations, such as dynamic time warping, may overcome this problem. Besides, these slightly lower results could also be due to differences in the neural processing mechanism between music imagination and perception - note that the perception trials were used to train the decoder applied to the imagination trials. Classification accuracies could possibly be further improved by using more advanced features. For example, since the estimated beat or tempo tracking in the neural responses correlates to the tempo of the original song (see Figure 6d), one could include the estimated tempo as an additional feature.

Moreover, investigating this beat tracking phenomenon teaches us more about the processing of auditory stimuli by the brain. In the case of music perception, previous literature shows that this neural entrainment is at least partially internally produced instead of stimulus-driven, thereby creating regular beat patterns in the brain even if some beats are missing in the perceived music[29]. Finding a beat pattern in the case of music imagination, as shown in this study, further broadens this observation to an entirely internally generated song.

However, further research is necessary to bring this approach closer to practical setting. First, for each subject, the experiments were recorded in a single continuous session. If this would not be the case, a retest would usually require calibration and retraining of the decoders due to a different fitting of the EEG cap. Moreover, 64 channels are used for the recordings in both data sets, which is too bulky for practical brain-computer interface (BCI) applications.

A better understanding of the process of music imagination and its decoding possibilities could increase our knowledge of the (musical) brain and its highly complex workings. Furthermore, with the upcoming interest in BCI technology, the results of this study open the door for several futuristic applications. For example, one can imagine a hands-free and silent BCI which allows one to select and play a song from a pre-defined playlist by merely thinking of it. Nevertheless, although the classification results obtained in this study are still far away from a practically viable accuracy, it provides an important initial step, demonstrating that imagined music recognition is possible, which paves the way for new research endeavors to further explore and improve this new paradigm.

References

1. Mori, K. & Iwanaga, M. Two types of peak emotional responses to music: The psychophysiology of chills and tears. *Sci. Rep.* **7**, 46063–46063 (2017).
2. Cuddy, L. L. & Duffin, J. Music, memory, and alzheimer's disease: is music recognition spared in dementia, and how can it be assessed? *Med. Hypotheses* **64**, 229–235 (2005).
3. Schulkind, M. D., Hennis, L. K. & Rubin, D. C. Music, emotion, and autobiographical memory: They're playing your song. *Mem. Cogn.* **27**, 948–955 (2013).
4. Lilliestam, L. On playing by ear. *Pop. Music.* **15**, 195–216 (1996).
5. Stober, S., Sternin, A., Owen, A. M. & Grahn, J. A. Towards music imagery information retrieval: Introducing the OpenMIIR dataset of EEG recordings from music perception and imagination. In *Proceedings of the 16th ISMIR Conference, 2015*, 763–769.
6. Jakubowski, K., Finkel, S., Stewart, L. & Müllensiefen, D. Dissecting an earworm: Melodic features and song popularity predict involuntary musical imagery. *Psychol. Aesthet. Creat. Arts* **11**, 122–135 (2017).
7. Di Liberto, G., Pelofi, C., Shamma, S. & de Cheveigné, A. Musical expertise enhances the cortical tracking of the acoustic envelope during naturalistic music listening. *Acoust. Sci. Technol.* **41**, 361–364 (2020).
8. Sturm, I., Dahne, S., Blankertz, B. & Curio, G. Multi-Variate EEG Analysis as a Novel Tool to Examine Brain Responses to Naturalistic Music Stimuli. *Plos One* **10** (2015).
9. Gang, N., Kaneshiro, B., Berger, J. & Dmochowski, J. P. Decoding neurally relevant musical features using canonical correlation analysis. In *Proceedings of the 18th ISMIR Conference, 2017*, 131–138.
10. Kaneshiro, B., Nguyen, D. T., Norcia, A. M., Dmochowski, J. P. & Berger, J. Natural music evokes correlated eeg responses reflecting temporal structure and beat. *NeuroImage* **214**, 116559–116559 (2020).
11. Vlek, R. J., Schaefer, R. S., Gielen, C. C. A. M., Farquhar, J. D. R. & Desain, P. Shared mechanisms in perception and imagery of auditory accents. *Clin. Neurophysiol.* **1526-1532** (2011).
12. Ofner, A. & Stober, S. Shared generative representation of auditory concepts and EEG to reconstruct perceived and imagined music. *19th ISMIR Conf. Paris, France, 2018* .

13. Schaefer, R. S., Farquhar, J., Blokland, Y., Sadakata, M. & Desain, P. Name that tune: Decoding music from the listening brain. *NeuroImage* **56**, 843–849 (2011).

14. Stober, S. Toward Studying Music Cognition with Information Retrieval Techniques: Lessons Learned from the OpenMIIR Initiative. *Front. Psychol.* **8** (2017).

15. O'Sullivan, J. A. *et al.* Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cereb. Cortex* **25**, 1697–1706 (2015).

16. Fuglsang, S. A., Dau, T. & Hjortkjaer, J. Noise-robust cortical tracking of attended speech in real-world acoustic scenes. *NeuroImage* **156**, 435–444 (2017).

17. Wong, D. D. E. *et al.* A Comparison of Regularization Methods in Forward and Backward Models for Auditory Attention Decoding. *Front. Neurosci.* **12:531** (2018).

18. Das, N., Bertrand, A. & Francart, T. EEG-based auditory attention detection: boundary conditions for background noise and speaker positions. *J. Neural Eng.* **15**, 066017 (2018).

19. Geirnaert, S. *et al.* Neuro-Steered Hearing Devices: Decoding Auditory Attention From the Brain. *arXiv* (2020).

20. Di Liberto, G. M. *et al.* Cortical encoding of melodic expectations in human temporal cortex. *eLife* **9** (2020).

21. Cantisani, G., Essid, S. & Richard, G. EEG-Based Decoding of Auditory Attention to a Target Instrument in Polyphonic Music. In *2019 IEEE WASPAA*, vol. 2019-, 80–84 (IEEE, 2019).

22. Treder, M. S., Purwins, H., Miklody, D., Sturm, I. & Blankertz, B. Decoding auditory attention to instruments in polyphonic music using single-trial EEG classification. *J. Neural Eng.* **11**, 10 (2014).

23. Somers, B., Francart, T. & Bertrand, A. A generic EEG artifact removal algorithm based on the multi-channel Wiener filter. *J. Neural Eng.* **15** (2018).

24. Biesmans, W., Das, N., Francart, T. & Bertrand, A. Auditory-Inspired Speech Envelope Extraction Methods for Improved EEG-Based Auditory Attention Detection in a Cocktail Party Scenario. *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**, 402–412 (2017).

25. Geirnaert, S., Francart, T. & Bertrand, A. An Interpretable Performance Metric for Auditory Attention Decoding Algorithms in a Context of Neuro-Steered Gain Control. *IEEE Trans. Neural Syst. Rehabil. Eng.* **28**, 307–317 (2020).

26. Verschueren, E., Somers, B. & Francart, T. Neural envelope tracking as a measure of speech understanding in cochlear implant users. *Hear. Res.* **373**, 23–31 (2019).

27. de Cheveigné, A. *et al.* Decoding the auditory brain with canonical component analysis. *NeuroImage* **172**, 206–216 (2018).

28. Nozaradan, S., Peretz, I., Missal, M. & Mouraux, A. Tagging the neuronal entrainment to beat and meter. *J. Neurosci.* **31**, 10234–10240 (2011).

29. Tal, I. *et al.* Neural Entrainment to the Beat: The "Missing-Pulse" Phenomenon. *J. Neurosci.* **37**, 6331–6341 (2017).

30. Tibo, M., Geirnaert, S. & Betrand, A. Decoding EEG responses during perception and imagination of music. www.scriptiebank.be/scriptie/2020/decoding-eeg-responses-during-perception-and-imagination-music (2020).

Author contributions statement

M.T., S.G., and A.B. conceived the experiment(s), M.T. conducted the experiment(s), analyzed the results, and wrote the first manuscript draft. M.T. and S.G. produced the images of this manuscript. All authors reviewed and edited the manuscript. Part of this work is based on the M.Sc. thesis project of M.T.[30].

Additional information

Funding This research is funded by an Aspirant Grant from the Research Foundation - Flanders (FWO) (for S. Geirnaert), the KU Leuven Special Research Fund C14/16/057, FWO project nr. G0A4918N, the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 802895), and the Flemish Government (AI Research Program). The scientific responsibility is assumed by its authors. A. Bertrand and S. Geirnaert are also affiliated to Leuven.AI - KU Leuven institute for AI, B-3000, Leuven, Belgium.

Accession codes

Competing interests The authors declare no competing interests.

Figure and table legends

Figure 1. Music perception: Median correlations for each subject (light purple dots) and song in both data sets. Darker colored dots display the median over all subjects.

Figure 2. Music perception: Classification accuracy for each subject (light purple dots) during the two-song, K-song, and song category classification tasks. The darker colored dots represent the median accuracy over all subjects.

Figure 3. Music imagination: Median correlations (across repeated trials) for each subject (lighter dots) and song during cued music imagination for the model with optimal time shift. Darker colored dots display the median over all subjects.

Figure 4. Music imagination: Classification accuracy for each subject (light purple dots) during the two-song, 12-song, and song category classification tasks, compared to the perception case (red dots). The darker colored dots represent the median accuracy over all subjects. * refers to a significant difference on the 0.05-level (Wilcoxon signed-rank test), ns to a non-significant difference.

Figure 5. Music imagination: Confusion matrices for the 12-song classifier and the song category classifier during the standalone imagination experiment. Each row of the matrix is normalized to sum to 1. We see a good overall classification, even for the same songs with/without lyrics or with similar tempos (Table 1).

Table 1. Music stimulus properties (OpenMIIR data set).

Figure 6. (a) The correlation as a function of time shift ν for a representative trial and subject of the OpenMIIR data set, for the standalone imagination experiment (no cue clicks). (b) Periodogram of the correlation shown in subfigure (a), showing a peak (3.47 Hz) close to the beat frequency of the stimulus (3.33 Hz). (c) The cross-correlation of the reconstruction with two periods of the envelopes, for the same representative trial, shows a similar periodical pattern. (d) Plot of the real vs. estimated tempo for each of the twelve songs per perception/imagination experiment. One dot represents the tempo of one song in one of the four experiments. The black line depicts the bisecting diagonal (corresponding to correlation $\rho = 1$) as a reference.