TOWARDS MUSIC IMAGERY INFORMATION RETRIEVAL: INTRODUCING THE OPENMIIR DATASET OF EEG RECORDINGS FROM MUSIC PERCEPTION AND IMAGINATION

Sebastian Stober, Avital Sternin, Adrian M. Owen and Jessica A. Grahn
Brain and Mind Institute, Department of Psychology, Western University, London, ON, Canada
`{sstober,asternin,adrian.owen,jgrahn}@uwo.ca`

ABSTRACT

Music imagery information retrieval (MIIR) systems may one day be able to recognize a song from only our thoughts. As a step towards such technology, we are presenting a public domain dataset of electroencephalography (EEG) recordings taken during music perception and imagination. We acquired this data during an ongoing study that so far comprises 10 subjects listening to and imagining 12 short music fragments – each 7–16s long – taken from well-known pieces. These stimuli were selected from different genres and systematically vary along musical dimensions such as meter, tempo and the presence of lyrics. This way, various retrieval scenarios can be addressed and the success of classifying based on specific dimensions can be tested. The dataset is aimed to enable music information retrieval researchers interested in these new MIIR challenges to easily test and adapt their existing approaches for music analysis like fingerprinting, beat tracking, or tempo estimation on EEG data.

1. INTRODUCTION

We all imagine music in our everyday lives. Individuals can imagine themselves producing music, imagine listening to others produce music, or simply "hear" the music in their heads. Music imagination is used by musicians to memorize music pieces and anyone who has ever had an "ear-worm" – a tune stuck in their head – has experienced imagining music. Recent research also suggests that it might one day be possible to retrieve a music piece from a database by just thinking of it.

As already motivated in [29], music imagery information retrieval (MIIR) – i.e., retrieving music by imagination – has the potential to overcome the query expressivity bottleneck of current music information retrieval (MIR) systems, which require their users to somehow imitate the desired song through singing, humming, or beat-boxing [31] or to describe it using tags, metadata, or lyrics fragments. Furthermore, music imagery appears to be a very promising means for driving brain-computer in-

terfaces (BCIs) that use electroencephalography (EEG) – a popular non-invasive neuroimaging technique that relies on electrodes placed on the scalp to measure the electrical activity of the brain. For instance, Schaefer et al. [23] argue that *"music is especially suitable to use here as (externally or internally generated) stimulus material, since it unfolds over time, and EEG is especially precise in measuring the timing of a response."* This allows us to exploit temporal characteristics of the signal such as rhythmic information.

Still, EEG data is generally very noisy and thus extracting relevant information can be challenging. This calls for sophisticated signal processing techniques as they have emerged in the field of MIR within the last decade. However, MIR researchers with the potential expertise to analyze music imagery data usually do not have access to the required equipment to acquire the necessary data for MIIR experiments in the first place. [1] In order to remove this substantial hurdle and encourage the MIR community to try their methods in this emerging interdisciplinary field, we are introducing the *OpenMIIR* dataset.

In the following sections, we will review closely related work in Section 2, describe our approach for data acquisition (Section 3) and basic processing (Section 4), and outline further steps in Section 5.

2. RELATED WORK

Retrieval based on brain wave recordings is still a very young and largely unexplored domain. A recent review of neuroimaging methods for MIR that also covers techniques different from EEG is given in [14]. EEG signals have been used to measure emotions induced by music perception [1,16] and to distinguish perceived rhythmic stimuli [28]. It has been shown that oscillatory neural activity in the gamma frequency band (20-60 Hz) is sensitive to accented tones in a rhythmic sequence [27]. Oscillations in the beta band (20-30 Hz) entrain to rhythmic sequences [2, 17] and increase in anticipation of strong tones in a non-isochronous, rhythmic sequence [5,6,13]. The magnitude of steady state evoked potentials (SSEPs), which reflect neural oscillations entrained to the stimulus, changes when subjects hear rhythmic sequences for frequencies related to the metrical structure of the rhythm. This is a sign of entrainment to beat and meter [19,20]. EEG studies have further shown that perturbations

---

[1] For instance, the Biosemi EEG system used here costs several ten-thousand dollars. Consumer-level EEG devices with a much lower price have become available recently but it is still open whether their measuring precision and resolution is sufficient for MIIR research.

of the rhythmic pattern lead to distinguishable event-related potentials (ERPs) [2] [7]. This effect appears to be independent of the listener's level of musical proficiency. Furthermore, Vlek et al. [32] showed that imagined auditory accents imposed on top of a steady metronome click can be recognized from EEG.

EEG has also been successfully used to distinguish perceived melodies. In a study by Schaefer et al. [26], 10 participants listened to 7 short melody clips with a length between 3.26s and 4.36s. For single-trial classification, each stimulus was presented 140 times in randomized back-to-back sequences of all stimuli. Using a quadratically regularized linear logistic-regression classifier with 10-fold cross-validation, they were able to successfully classify the ERPs of single trials. Within subjects, the accuracy varied between 25% and 70%. Applying the same classification scheme across participants, they obtained between 35% and 53% accuracy. In a further analysis, they combined all trials from all subjects and stimuli into a grand average ERP. Using singular-value decomposition, they obtained a fronto-central component that explained 23% of the total signal variance. The time courses corresponding to this component showed significant differences between stimuli that were strong enough to allow cross-participant classification. Furthermore, a correlation with the stimulus envelopes of up to 0.48 was observed with the highest value over all stimuli at a time lag of 70–100ms.

FMRI studies [10, 11] have shown that similar brain structures and processes are involved during music perception and imagination. As Hubbard concludes in his recent review of the literature on auditory imagery, *"auditory imagery preserves many structural and temporal properties of auditory stimuli"* and *"involves many of the same brain areas as auditory perception"* [12]. This is also underlined by Schaefer [23, p. 142] whose *"most important conclusion is that there is a substantial amount of overlap between the two tasks* [music perception and imagination]*, and that 'internally' creating a perceptual experience uses functionalities of 'normal' perception."* Thus, brain signals recorded while listening to a music piece could serve as reference data. The data could be used in a retrieval system to detect salient elements expected during imagination. A recent meta-analysis [25] summarized evidence that EEG is capable of detecting brain activity during the imagination of music. Most notably, encouraging preliminary results for recognizing imagined music fragments from EEG recordings were reported in [24] in which 4 out of 8 participants produced imagery that was classifiable (in a binary comparison) with an accuracy between 70% and 90% after 11 trials.

Another closely related field of research is the reconstruction of auditory stimuli from EEG recordings. Deng et al. [3] observed that EEG recorded during listening to natural speech contains traces of the speech amplitude envelope. They used independent component analysis (ICA) and a source localization technique to enhance the strength of this signal and successfully identify heard sentences. Applying their technique to imagined speech, they reported statistically significant single-sentence classification performance for 2 of 8 subjects with better performance when several sentences were combined for

a longer trial duration.

Recently, O'Sullivan et al. [21] proposed a method for decoding attentional selection in a cocktail party environment from single-trial EEG recordings approximately one minute long. In their experiment, 40 subjects were presented with 2 classic works of fiction at the same time – each one to a different ear – for 30 trials. To determine which of the 2 stimuli a subject attended to, they reconstructed both stimulus envelopes from the recorded EEG. To this end, they trained two different decoders per trial using a linear regression approach – one to reconstruct the attended stimulus and the other to reconstruct the unattended one. This resulted in 60 decoders per subject. These decoders where then averaged in a leave-one-out cross-validation scheme. During testing, each decoder would predict the stimulus with the best reconstruction from the EEG using the Pearson correlation of the envelopes as measure of quality. Using subject-specific decoders averaged from 29 training trials, the prediction of the attended stimulus decoder was correct for 89% of the trials whereas the mean accuracy of the unattended stimulus decoder was 78.9%. Alternatively, using a grand-average decoding method that combined the decoders from every other subject and every other trial, they obtained a mean accuracy of 82% and 75% respectively.

3. STUDY DESCRIPTION

This section provides details about the study that was conducted to collect the data released in the OpenMIIR dataset. The study consisted of two portions. We first collected information about the participants using questionnaires and behavioral testing (Section 3.1) and then ran the actual EEG experiment (Section 3.2) with those participants matching our inclusion criteria. The 12 music stimuli used in this experiment are described in Section 3.3.

3.1 Questionnaires and Behavioral Testing

14 participants were recruited using approved posters at the University of Western Ontario. We collected information about the participants' previous music experience, their ability to imagine sounds, and information about musical sophistication using an adapted version of the widely used Goldsmith's Musical Sophistication Index (G-MSI) [18] combined with an adapted clarity of auditory imagination scale [33]. Questions from the perceptual abilities and musical training subscales of the G-MSI were used to identify individual differences in these areas. For the clarity of auditory imagery scale, participants had to self-report their ability to clearly hear sounds in their head. Our version of this scale added five music-related items to five items from the original scale.

We also had participants complete a beat tapping and a stimuli familiarity task. Participants listened to each stimulus and were asked to tap along with the music on the table top. The experimenter then rated their tapping ability on a scale from 1 (difficult to assess) to 3 (tapping done properly). After listening to each stimulus participants rated their familiarity with the stimuli on a scale from 1 (unfamiliar) to 3 (very familiar). To participate in the EEG portion of the study, the participants had to receive a score of at least 90% on our beat tapping task.

---

[2] A description of how event-related potentials (ERPs) are computed and some examples are provided in Section 4.
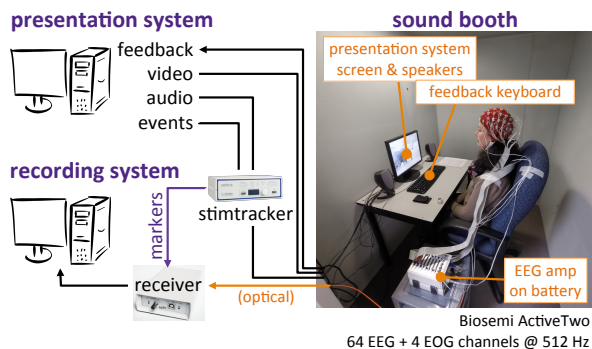
Figure 1. Setup for the EEG experiment. The presentation and recording systems were placed outside to reduce the impact of electrical line noise that could be picked up by the EEG amplifier.

Table 1. Information about the tempo, meter and length of the stimuli (without cue clicks) used in this study.

ID	Name	Meter	Length	Tempo
1	Chim Chim Cheree (lyrics)	3/4	13.3s	212 BPM
2	Take Me Out to the Ballgame (lyrics)	3/4	7.7s	189 BPM
3	Jingle Bells (lyrics)	4/4	9.7s	200 BPM
4	Mary Had a Little Lamb (lyrics)	4/4	11.6s	160 BPM
11	Chim Chim Cheree	3/4	13.5s	212 BPM
12	Take Me Out to the Ballgame	3/4	7.7s	189 BPM
13	Jingle Bells	4/4	9.0s	200 BPM
14	Mary Had a Little Lamb	4/4	12.2s	160 BPM
21	Emperor Waltz	3/4	8.3s	178 BPM
22	Hedwig's Theme (Harry Potter)	3/4	16.0s	166 BPM
23	Imperial March (Star Wars Theme)	4/4	9.2s	104 BPM
24	Eine Kleine Nachtmusik	4/4	6.9s	140 BPM
	mean		10.4s	176 BPM

Participants received scores from 75%–100% with an average score of 96%. Furthermore, they needed to receive a score of at least 80% on our stimuli familiarity task. Participants received scores from 71%–100% with an average score 87%. These requirements resulted in rejecting 4 participants. This left 10 participants (3 male), aged 19–36, with normal hearing and no history of brain injury. These 10 participants had an average tapping score of 98% and an average familiarity score of 92%. Eight participants had formal musical training (1–10 years), and four of those participants played instruments regularly at the time of data collection. After the experiment, we asked participants the method they used to imagine music. The participants were split evenly between imagining themselves producing the music (singing or humming) and simply "hearing the music in [their] head."

3.2 EEG Recording

For the EEG portion of the study, the 10 participants were seated in an audiometric room (Eckel model CL-13) and connected to a BioSemi Active-Two system recording 64+2 EEG channels at 512 Hz as shown in Figure 1. Horizontal and vertical EOG channels were used to record eye movements. We also recorded the left and right mastoid channel as EEG reference signals. Due to an oversight, the mastoid data was not collected for the first 5 subjects. The presented audio was routed through a Cedrus StimTracker connected to the EEG receiver, which allowed a high-precision synchronization (<0.05 ms) of the stimulus onsets with the EEG data. The experiment was programmed and presented using PsychToolbox run in Matlab 2014a. A computer monitor displayed the instructions and fixation cross for the participants to focus on during the trials to reduce eye movements. The stimuli and cue clicks were played through speakers at a comfortable volume that was kept constant across participants. Headphones were not used because pilot participants reported headphones caused them to hear their heartbeat which interfered with the imagination portion of the experiment.

The EEG experiment was divided into 2 parts with 5 blocks each as illustrated in Figure 2. A single block comprised of all

12 stimuli in randomized order. Between blocks, participants could take breaks at their own pace. We recorded EEG in 4 conditions:

1. Stimulus perception preceded by cue clicks
2. Stimulus imagination preceded by cue clicks
3. Stimulus imagination without cue clicks
4. Stimulus imagination without cue clicks, with feedback

The goal was to use the cue to align trials of the same stimulus collected under conditions 1 and 2. Lining up the trials allows us to directly compare the perception and imagination of music and to identify overlapping features in the data. Conditions 3 and 4 simulate a more realistic query scenario during which the system does not have prior information about the tempo and meter of the imagined stimulus. These two conditions were identical except for the trial context. While the condition 1–3 trials were recorded directly back-to-back within the first part of the experiment, all condition 4 trials were recorded separately in the second part, without any cue clicks or tempo priming by prior presentation of the stimulus. After each condition 4 trial, participants provided feedback by pressing one of two buttons indicating on whether or not they felt they had imagined the stimulus correctly. In total, 240 trials (12 stimuli x 4 conditions x 5 blocks) were recorded per subject. The event markers recorded in the raw EEG comprise:

- Trial labels (as a concatenation of stimulus ID and condition) at the beginning of each trial
- Exact audio onsets for the first cue click of each trial in conditions 1 and 2 (detected by the Stimtracker)
- Subject feedback for the condition 4 trials (separate event IDs for positive and negative feedback)

3.3 Stimuli

Table 1 shows an overview of the stimuli used in the study. This selection represents a tradeoff between exploration and exploitation of the stimulus space. As music has many facets, there are naturally many possible dimensions in which music pieces may vary. Obviously, only a limited subspace could be explored with any given set of stimuli. This had to be balanced against the number of trials that could be recorded for each stimulus (exploitation) within a given time limit of 2 hours for a single recording session (including fitting the EEG equipment).
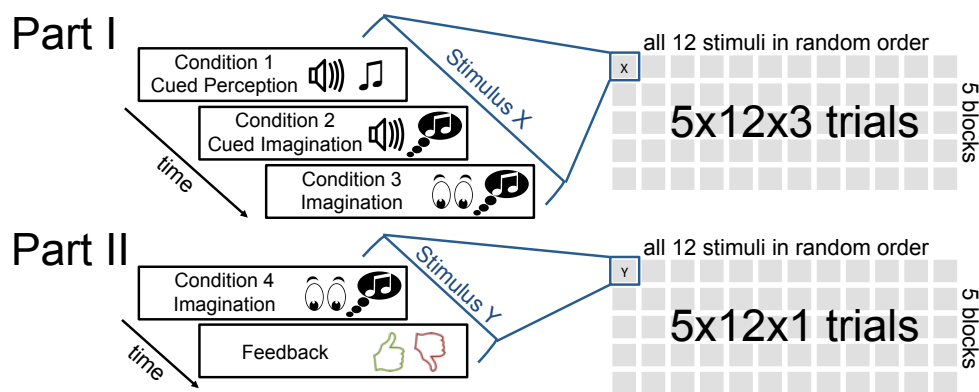
Figure 2. Illustration of the design for the EEG portion of the study.

Based on the findings from related studies (c.f. Section 2), we primarily focused on the rhythm/meter and tempo dimensions. Consequently, the set of stimuli was evenly divided into pieces with 3/4 and 4/4 meter, i.e. two very distinct rhythmic "feels." The tempo spanned a range between 104 and 212 beats per minute (BPM). Furthermore, we were also interested in whether the presence of lyrics would improve the recognizability of the stimuli. Hence, we divided the stimulus set into 3 equally sized groups:

- 4 recordings of songs with lyrics (1–4),
- 4 recordings of the same songs without lyrics (11–14), and
- 4 instrumental pieces (21–24).

The pairs of recordings for the same song with and without lyrics were tempo-matched by pre-selection and subsequent fine adjustment using the time-stretching function of Audacity. [3] Due to minor differences in tempo between pairs of stimuli with and without lyrics, the tempo of the stimuli had to be slightly modified after the first five participants.

All stimuli were considered to be well-known pieces in the North-American cultural context. They were normalized in volume and kept as similar in length as possible with care taken to ensure that they all contained complete musical phrases starting from the beginning of the piece. Each stimulus started with approximately two seconds of clicks (1 or 2 bars) as an auditory cue to the tempo and onset of the music. The clicks began to fade out at the 1s-mark within the cue and stopped at the onset of the music.

3.4 Data and Code Sharing

With the explicit consent of all participants and the approval of the ethics board at the University of Western Ontario, the data collected in this study are released as OpenMIIR dataset [4] under the *Open Data Commons Public Domain Dedication and License (PDDL).* [5] This comprises the anonymized answers from the questionnaires, the behavioral scores, the subjects' feedback for the trials in condition 4 and the raw EEG and EOG data of all trials at the original sample rate of 512 Hz. This amounts to approximately 700 MB of data per subject.

Raw data are shared in the FIF format used by MNE [9], which can easily be converted to the MAT format of Matlab.

Additionally, the Matlab code and the stimuli for running the study are made available as well as the python code for cleaning and processing the raw EEG data as described in Section 4. The python code uses the libraries MNE-Python [8] and deepthought [6], which are both published as open-source under the 3-clause BSD license. [7]

This approach ensures accessibility and reproducibility. Researchers have the possibility to just apply their methods on the already pre-processed data or change any step in the pre-processing pipeline according to their needs. No proprietary software is required for working with the data. The wiki on the dataset website can be used to share code, ideas and results related to the dataset.

4. BASIC EEG PROCESSING

This section describes basic EEG processing techniques that may serve as a basis for the application of more sophisticated analysis methods. More examples are linked in the wiki on the dataset website.

4.1 EEG Data Cleaning

EEG recordings are usually very noisy. They contain artifacts caused by muscle activity such as eye blinking as well as possible drifts in the impedance of the individual electrodes over the course of a recording. Furthermore, the recording equipment is very sensitive and easily picks up interferences such as electrical line noise from the surroundings. The following common-practice pre-processing steps were applied to remove unwanted artifacts.

The raw EEG and EOG data were processed using the MNE-Python toolbox. The data was first visually inspected for artifacts. For one subject (P05), we identified several episodes of strong movement artifacts during trials. Hence, these particular data need to be treated with care when used for analysis – possibly picking only specific trials without artifacts. The bad

---

[3] http://web.audacityteam.org/
[4] https://github.com/sstober/openmiir
[5] http://opendatacommons.org/licenses/pddl

[6] https://github.com/sstober/deepthought
[7] http://opensource.org/licenses/BSD-3-Clause

trials might however still be used for testing the robustness of analysis techniques.

For recordings with additional mastoid channels, the EEG data was re-referenced by subtracting the mean mastoid signal [30]. We then removed and interpolated bad EEG channels identified by manual visual inspection. For interpolation, the spherical splines method described in [22] was applied. The number of bad channels in a single recording session varied between 0 and 3. The data were then filtered with an fft-bandpass, keeping a frequency range between 0.5 and 30 Hz. This also removed any slow signal drift in the EEG. Afterwards, we down-sampled to a sampling rate of 64 Hz. To remove artifacts caused by eye blinks, we computed independent components using extended Infomax ICA [15] and semi-automatically removed components that had a high correlation with the EOG channels. Finally, the 64 EEG channels were reconstructed from the remaining independent components without reducing dimensionality.

4.2 Grand Average Trial ERPs

A common approach to EEG analysis is through the use of event-related potentials (ERPs). An ERP is an electrophysiological response that occurs as a direct result of a stimulus. Raw EEG data is full of unwanted signals. In order to extract the signal of interest from the noise, participants are presented with the same stimulus many times. The brain's response to the stimulus remains constant while the noise changes. The consistent brain response becomes apparent when the signals from the multiple stimulus presentations are averaged together and the random noise is averaged to zero. In order to identify common brain response patterns across subjects, grand average ERPs are computed by averaging the ERPs of different subjects.

The size and the timing of peaks in the ERP waveform provide information about the brain processes that occur in response to the presented stimulus. By performing a principle component analysis (PCA), information regarding the spatial features of these processes can be obtained.

As proposed in [26], we computed grand average ERPs by aggregating over all trials (excluding the cue clicks) of the same stimulus from all subjects except P05 (due to the movement artifacts). In their experiment, Schaefer et al. [26] used very short stimuli allowing each stimulus to be repeated many times. They averaged across hundreds of short (3.26s) trials, concatenated the obtained grand average ERPs and then applied PCA, which resulted in clearly defined spatial components. We had fewer repetitions of our stimuli. Therefore, to preserve as much data as possible, we used the full length of the trials as opposed to the first 3.26 seconds. We then concatenated the grand average ERPs and applied a PCA, which resulted in principal components with poorly defined spatial features as shown in Figure 3 (A and B). As an alternative, we performed a PCA on the concatenated raw trials without first calculating an average across trials. This approach produced clearly defined spatial components shown in Figure 3 (C and D). Components 2 to 4 are similar to those described in [26]. Except for their (arbitrary) polarity, the components are very similar across the two conditions, which may be indicative of similar processes being involved in both perception and imagination of music as
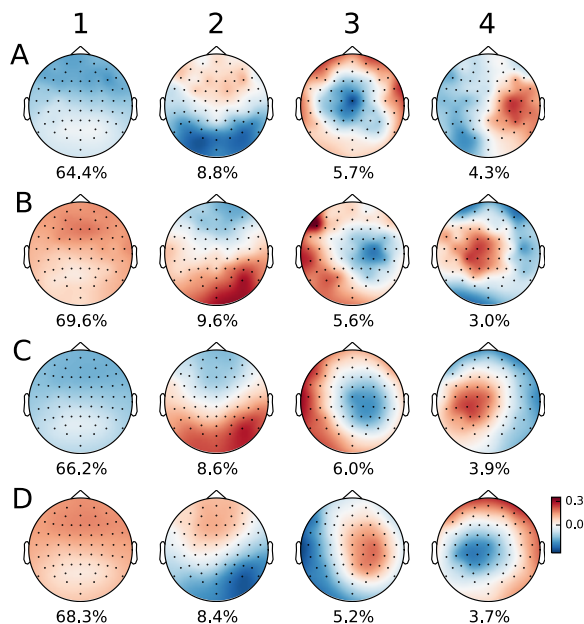
Figure 3. Topographic visualization of the top 4 principle components with percentage of the explained signal variance. Channel positions in the 64-channel EEG layout are shown as dots. Colors are interpolated based on the channel weights. The PCA was computed on **A**: the grand average ERPs of all perception trials, **B**: the grand average ERPs of all cued imagination trials, **C**: the concatenated perception trials, **D**: the concatenated cued imagination trials.

described in [11, 25].

Schaefer et al. [26] were able to use the unique time course of the component responsible for the most variance to differentiate between stimuli. Analyzing the signals corresponding to the principle components, we have not yet been able to reproduce a significant stimulus classification accuracy. This could be caused by our much smaller number of trials, which are also substantially longer than those used by [26]. Furthermore, the cross-correlation between the stimulus envelopes and the component waveforms were much lower (often below 0.1) than reported in [26].
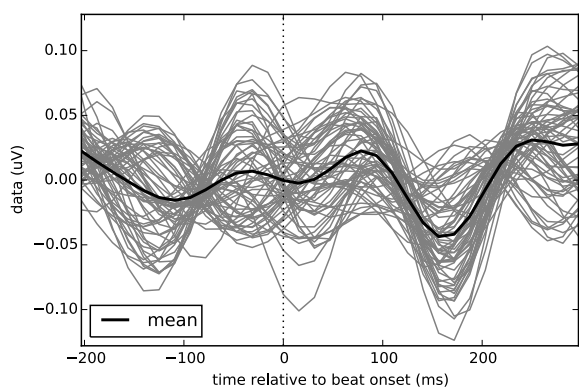
4.3 Grand Average Beat ERPs

In the previous section, we computed ERPs based on the trial onsets. Similarly, it is also possible to analyze beat events. Using the dynamic beat tracker [4] provided by the librosa [8] library, we obtained beat annotations for all beats within the audio stimuli. To this end, the beat tracker was initialized with the known tempo of each stimulus. The quality of the automatic annotations was verified through sonification.

Knowing the beat positions allows to analyze the respective EEG segments in the perception condition. For this analysis, the EEG data was additionally filtered with a low-pass at 8 Hz to remove alpha band activity (8–12 Hz). Figure 4 shows

---

[8] https://github.com/bmcfee/librosa
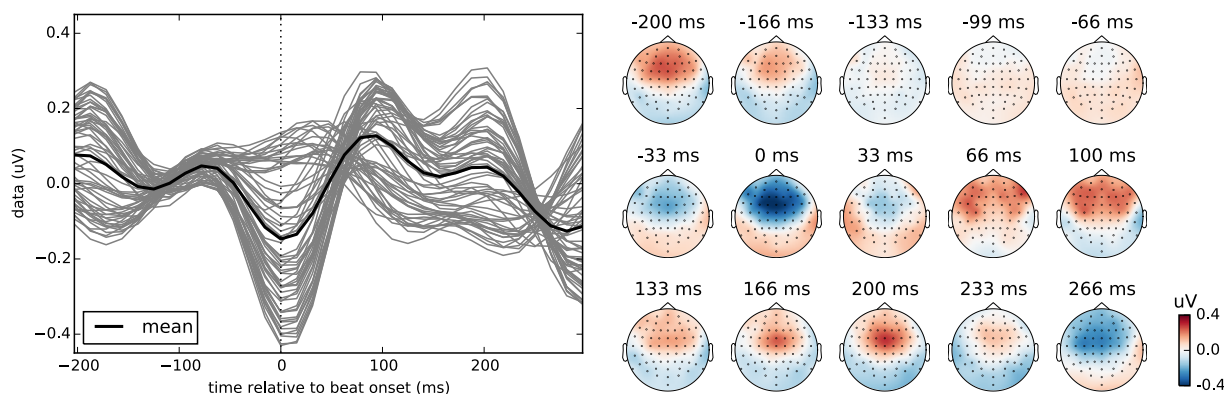
Figure 4. Grand average beat ERP for the perception trials (16515 beats). All times are relative to the beat onset. Left: Individual channels and mean over time. Right: Topographic visualization for discrete time points (equally spaced at 1/30s interval).

Figure 5. Grand average beat ERP for the cued imagination trials (16515 beats). All times are relative to the beat onset. Note the difference in amplitude compared to Figure 4.

the grand average ERP for all beats except the cue clicks [9] in all perception trials of all subjects except P05. Here we considered epochs, i.e., EEG segments of interest, from 200 ms before until 300 ms after each beat marker. Before averaging into the ERP, we applied a baseline correction of each epoch by subtracting the signal mean computed from the 200 ms sub-segment before the beat marker.

The ERP has a negative dip that coincides with the beat onset time at 0 ms. Any auditory processing related to the beat would occur much later. A possible explanation is that the dip is caused by the anticipation of the beat. However, this requires further investigation. There might be potential to use this effect as the basis for an MIIR beat or tempo tracker. For comparison, the respective grand average ERP for the cued imagination trials is shown in Figure 5. This ERP looks very different from the one for the perception conditions. Most notably the amplitude scale is very low. This outcome was probably caused by the imprecise time locking. In order to compute meaningful ERPs, the precise event times (beat onsets) need to be known. However, small tempo variations during imagination are very likely and thus the beat onsets are most likely not exact.

---

[9] Cue clicks were excluded because these isolated auditory events illicit a different brain response than beats embedded into a stream of music.

5. CONCLUSIONS AND OUTLOOK

We have introduced OpenMIIR – an open EEG dataset intended to enable MIR researchers to venture into the domain of music imagery and develop novel methods without the need for special EEG equipment. We plan to add new EEG recordings with further subjects to the dataset and possibly adapt the experimental settings as we learn more about the problem. In our first experiments using this dataset, we were able to partly reproduce the identification of overlapping components between music perception and imagination as reported earlier.

Will it one day be possible to just think of a song and the music player will start its playback? If this could be achieved, it would require the intense interdisciplinary collaboration between MIR researchers and neuroscientists. We hope that the OpenMIIR dataset will facilitate such a collaboration and contribute to new developments in this emerging field for research.

6. REFERENCES

[1] R. Cabredo, R. S. Legaspi, P. S. Inventado, and M. Numao. An Emotion Model for Music Using Brain Waves. In *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR'12)*, pages 265–270, 2012.

[2] L. K. Cirelli, D. Bosnyak, F. C. Manning, C. Spinelli, C. Marie, T. Fujioka, A. Ghahremani, and L. J. Trainor. Beat-induced fluctuations in auditory cortical beta-band activity: Using EEG to measure age-related changes. *Frontiers in Psychology*, 5(Jul):1–9, 2014.

[3] S. Deng, R. Srinivasan, and M. D'Zmura. Cortical signatures of heard and imagined speech envelopes. Technical report, DTIC, 2013.

[4] D. P. W. Ellis. Beat Tracking by Dynamic Programming. *Journal of New Music Research*, 36(1):51–60, 2007.

[5] T. Fujioka, L. J. Trainor, E. W. Large, and B. Ross. Beta and gamma rhythms in human auditory cortex during musical beat processing. *Annals of the New York Academy of Sciences*, 1169:89–92, 2009.

[6] T. Fujioka, L. J. Trainor, E. W. Large, and B. Ross. Internalized Timing of Isochronous Sounds Is Represented in Neuromagnetic Beta Oscillations. *Journal of Neuroscience*, 32(5):1791–1802, 2012.

[7] E. Geiser, E. Ziegler, L. Jancke, and M. Meyer. Early electrophysiological correlates of meter and rhythm processing in music perception. *Cortex*, 45(1):93–102, 2009.

[8] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7, 2013.

[9] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, and M. S. Hämäläinen. MNE software for processing MEG and EEG data. *NeuroImage*, 86(0):446 – 460, 2014.

[10] A. R. Halpern, R. J. Zatorre, M. Bouffard, and J. A. Johnson. Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia*, 42(9):1281–92, 2004.

[11] S. Herholz, A. Halpern, and R. Zatorre. Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of cognitive neuroscience*, 24(6):1382–97, 2012.

[12] T. L. Hubbard. Auditory imagery: empirical findings. *Psychological Bulletin*, 136(2):302–329, 2010.

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

[14] B. Kaneshiro and J. P. Dmochowski. Neuroimaging methods for music information retrieval: Current findings and future prospects. In *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR'15)*, 2015.

[15] T.-W. Lee, M. Girolami, and T. J. Sejnowski. Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources. *Neural Computation*, 11(2):417–441, 1999.

[16] Y.-P. Lin, T.-P. Jung, and J.-H. Chen. EEG dynamics during music appreciation. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'09)*, pages 5316–5319, 2009.

[17] H. Merchant, J. Grahn, L. J. Trainor, M. Rohrmeier, and W. T. Fitch. Finding a beat: a neural perspective across humans and non-human primates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2015.

[18] D. Müllensiefen, B. Gingras, J. Musil, and L. Stewart. The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLoS ONE*, 9(2), 2014.

[19] S. Nozaradan, I. Peretz, M. Missal, and A. Mouraux. Tagging the neuronal entrainment to beat and meter. *The Journal of Neuroscience*, 31(28):10234–10240, 2011.

[20] S. Nozaradan, I. Peretz, and A. Mouraux. Selective Neuronal Entrainment to the Beat and Meter Embedded in a Musical Rhythm. *The Journal of Neuroscience*, 32(49):17572–17581, 2012.

[21] J. A. O'Sullivan, A. J. Power, N. Mesgarani, S. Rajaram, J. J. Foxe, B. G. Shinn-Cunningham, M. Slaney, S. A. Shamma, and E. C. Lalor. Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cerebral Cortex*, (25):1697–1706, 2015.

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

[23] R. Schaefer. *Measuring the mind's ear EEG of music imagery*. PhD thesis, Radboud University Nijmegen, 2011.

[24] R. Schaefer, Y. Blokland, J. Farquhar, and P. Desain. Single trial classification of perceived and imagined music from EEG. In *Proceedings of the 2009 Berlin BCI Workshop*. 2009.

[25] R. S. Schaefer, P. Desain, and J. Farquhar. Shared processing of perception and imagery of music in decomposed EEG. *NeuroImage*, 70:317–326, 2013.

[26] R. S. Schaefer, J. Farquhar, Y. Blokland, M. Sadakata, and P. Desain. Name that tune: Decoding music from the listening brain. *NeuroImage*, 56(2):843–849, 2011.

[27] J. S. Snyder and E. W. Large. Gamma-band activity reflects the metric structure of rhythmic tone sequences. *Cognitive Brain Research*, 24:117–126, 2005.

[28] S. Stober, D. J. Cameron, and J. A. Grahn. Using convolutional neural networks to recognize rhythm stimuli from electroencephalography recordings. In *Advances in Neural Information Processing Systems 27 (NIPS'14)*, pages 1449–1457, 2014.

[29] S. Stober and J. Thompson. Music imagery information retrieval: Bringing the song on your mind back to your ears. In *13th International Conference on Music Information Retrieval (ISMIR'12) - Late-Breaking & Demo Papers*, 2012.

[30] M. Teplan. Fundamentals of EEG measurement. *Measurement science review*, 2(2):1–11, 2002.

[31] G. Tzanetakis, A. Kapur, and M. Benning. Query-by-Beat-Boxing: Music Retrieval For The DJ. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR'04)*, pages 170–177, 2004.

[32] R. J. Vlek, R. S. Schaefer, C. C. A. M. Gielen, J. D. R. Farquhar, and P. Desain. Shared mechanisms in perception and imagery of auditory accents. *Clinical Neurophysiology*, 122(8):1526–1532, 2011.

[33] J. Willander and S. Baraldi. Development of a new clarity of auditory imagery scale. *Behaviour Research Methods*, 42(3):785–590, 2010.