MAD-EEG: an EEG dataset for decoding auditory attention to a target instrument in polyphonic music

Giorgia Cantisani, Gabriel Trégoat, Slim Essid, Gael Richard

MAD-EEG: an EEG dataset for decoding auditory attention to a target instrument in polyphonic music

Giorgia Cantisani[1]*, *Gabriel Trégoat*[1], *Slim Essid*[1] , *Gaël Richard*[1]

[1]LTCI, Télécom Paris, Institut Polytechnique de Paris, 75013, Paris, France

firstname.lastname@telecom-paris.fr

Abstract

We present MAD-EEG, a new, freely available dataset for studying EEG-based auditory attention decoding considering the challenging case of subjects attending to a target instrument in polyphonic music. The dataset represents the first music-related EEG dataset of its kind, enabling, in particular, studies on single-trial EEG-based attention decoding, while also opening the path for research on other EEG-based music analysis tasks. MAD-EEG has so far collected 20-channel EEG signals recorded from 8 subjects listening to solo, duo and trio music excerpts and attending to one pre-specified instrument. The proposed experimental setting differs from the ones previously considered as the stimuli are polyphonic and are played to the subject using speakers instead of headphones. The stimuli were designed considering variations in terms of number and type of instruments in the mixture, spatial rendering, music genre and melody that is played. Preliminary results obtained with a state-of-the-art stimulus reconstruction algorithm commonly used for speech stimuli show that the audio representation reconstructed from the EEG response is more correlated with that of the attended source than with the one of the unattended source, proving the dataset to be suitable for such kind of studies.

Index Terms: Auditory attention, Polyphonic music, EEG

1. Introduction

Auditory attention decoding aims at determining which sound source a subject is paying specific attention to. Humans have a remarkable ability to enhance sound sources, tuning out interfering noise as well as *focusing on* specific sound characteristics, such as melodies, rhythms, timbre etc. Attention is then acting as a cognitive filter that allows human beings to better access and process high-level sound information. Previous studies on speech attention decoding [1–6] have shown that the electroencephalographic (EEG) activity tracks dynamic changes in the speech stimulus and can be used successfully to decode selective attention in a multispeaker environment. The natural transposition of this problem in the Music information retrieval (MIR) field is decoding the attention to a particular target instrument while listening to multi-instrumental music. Developing such models would open the path for research on other EEG-driven MIR tasks such as, for instance, selective source separation or enhancement, score following, music generation and transcription or rhythm analysis to mention a few examples. Moreover, from a neuroscientific viewpoint, this would yield a better understanding of musical audio information processing by the human brain.

In the MIR community a few attempts have been made at detecting and extracting music information from the brain's activity while a human subject is listening to music [7–16]. Among them, attended musical source decoding is still a poorly explored topic due to its complexity and the lack of experimental data. In fact, such a study requires data of well-synchronized musical stimuli and corresponding neural responses which can only be acquired in a controlled sensory stimulation experiment.

Many approaches can be used to analyze and understand how EEG signals are affected by specific stimuli [17]. A typical work-flow is to repeat the stimulus several times and then average the EEG responses in order to keep only the stimuli-relevant information and attenuate noise. While this approach was found successful for short stimuli or isolated events, it becomes time-consuming and unpractical with *naturalistic* stimuli, such as real-world music, speech or environmental sounds. Short or highly deviant stimuli generate the so-called *event-related potentials (ERPs)*, which exhibit a characteristic morphology: peaks are observed at a specific time-latency in the average EEG responses. Such characteristics can be re-created in experimental settings through the so-called *oddball paradigm*, where the subject is stimulated with a rare *deviant event* occurring among more frequent *standard events* [18]. Considering audio stimuli, this kind of approach is typically considered to study attention only to particular musical structures such as note onsets, rhythm and pitch patterns or, at least unattended musical deviants among standard and attended events [18]. However, it is then difficult to untangle the attention due to the novelty of the stimulus from the attention to the stimulus itself. Due to this, studying the attention to a particular source in naturalistic stimuli such as music excerpts or speech utterances, is difficult with this kind of approach. This problem has actually been overcome in the case of speech, for which models have been developed to decode attention from *single-trial* EEG responses [1–3, 6].

Specific datasets were assembled for these studies but this kind of data is still not available for music stimuli. In this context, we acquired a new dataset, named MAD-EEG, which is suitable also for studying auditory attention to a target instrument in polyphonic music using both single-trial and averaging-based decoding techniques. The dataset is freely available online to stimulate research in this area.

2. Related works

There are a few publicly available music-related EEG datasets. Stanford University researchers have assembled a number of such datasets (`NMED-H` [19], `NMED-T` [20] and `NMED-RP` [21]) containing EEG and behavioural responses to different kinds of naturalistic music stimuli. However, they were acquired while the user was focusing on the entire stimulus or on its rhythm and not on a particular instrument, thus they are not suitable for the auditory attention decoding task. The `DEAP` database [22] used music videos as stimuli in order to study human affective states and the `OpenMIIR` dataset [23] contains EEG

This project has received funding from the European Unions Horizon 2020 research and innovation program under the Marie Skodowska-Curie grant agreement No. 765068

recordings collected during music perception and imagination. These two datasets were designed for a different purpose, thus the subjects were not asked to pay attention to anything in particular. The only dataset where participants were asked to attend to a target instrument while listening to polyphonic music, is the `music BCI` dataset [18]. It followed a multi-streamed oddball paradigm where each of 3 instruments was playing a repetitive musical pattern, interspersed with a randomly occurring deviant pattern which yields clean P300 ERPs. However, this dataset was specifically designed for ERP-based attention decoding studies. Our goal is rather to focus on real world music compositions which are not specifically designed to evoke ERPs and be able to study the continuous EEG response to a given stimulus.

In contrast with the above-listed works and taking inspiration from the speech datasets, we have performed EEG recordings of subjects while they were listening to realistic polyphonic music and attending to a particular instrument in the mixture. The main novelty of our contribution is the design of the experimental protocol and its implementation to collect a music-related EEG dataset specifically developed for attention decoding purposes. This will allow researchers to study the responses to naturalistic music stimuli using both single-trial and averaging-based attention decoding techniques.

3. Dataset creation

Surface electroencephalographic (EEG) signals were recorded from 8 subjects while they were listening to polyphonic music stimuli. For each audio stimulus, which consists of a mixture containing from two to three instruments, the subjects were asked to attend to a particular instrument. Each subject thus listened to a total of 78 stimuli presented in a random order, each one consisting of 4 repetitions of the same roughly 6-second long music excerpt. This corresponds to a total of approximately 30-32 minutes of 20-channel EEG recordings. Each subject listened to 14 solos, 40 duets and 24 trios.

It is worth noting that this setting is completely different from the ones previously proposed. The experimental protocol usually applied for attention decoding experiments like the ones of [3, 5, 18], considers two monaural sources each played to a different ear through headphones. Instead, in our recording sessions, the stimuli were reproduced using speakers and the audio was rendered in different spatial configurations.

3.1. Stimuli

The stimuli consist of polyphonic music mixes created starting from a selection of music excerpts played by single instruments. For pop excerpts, they were single instrument tracks of a real composition, while for classical music the different instruments were combined in order to get realistic duets and trios. The sound volume was then peak-normalized for all mixes, so as to avoid bias that could result from varying loudness audio.

Attention to speech is mostly semantic while attention to an instrument could stem from multiple factors (e.g. timbre, melody, etc). Moreover, the more instruments in the mixture, the more difficult is supposed to be the attention task when the instrument is not in the foreground. Thus, different configurations were considered in the choice of the musical stimuli in order to test the influence of such factors on attention decoding:

- Two musical *genres*: pop and classical. Pop excerpts were carefully chosen with sharp rhythmical and harmonic patterns to contrast with the classical ones.

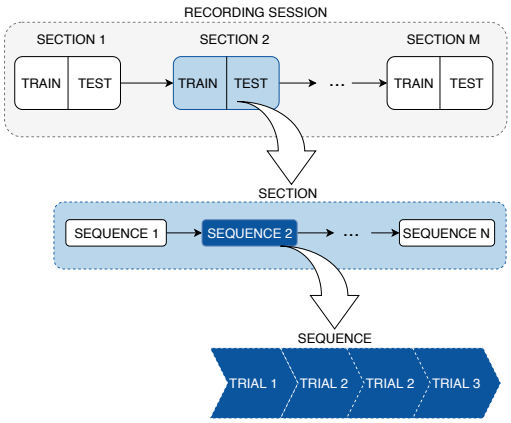

Figure 1: *A recording session is divided in sections. Each section is associated with a given musical piece in the dataset and consists of a training and a test phase, where a series of stimuli sequences is played. Each stimulus sequence consists of 4 trials where the same stimulus is listened to repetitively.*

- Two musical *compositions* per genre.
- Two *themes* per musical piece, that is, for the same piece, there are two different excerpts exhibiting exactly the same instruments but playing different parts of the score.
- Two *ensemble types*: duets and trios.
- Two *spatial renderings*: monophonic and stereo.
- *Musical instruments*: combinations of Flute, Oboe, French Horn, Bassoon and Cello for classical excerpts, along with Voice, Guitar, Bass and Drums for pop ones.

3.2. Recording protocol

One important aspect we had to consider is that the duration of each stimulus had to be long enough to allow the study of attention on a single-trial basis while targeting realistic music excerpts. On the other hand, the duration of the experiment had to remain reasonably short in order to control the cognitive load on the subject. Too long experiments would indeed result in an unsatisfactory level of concentration throughout the session. Consequently, we limited the duration of a stimulus to around 6 seconds. Then, during the experiment, each stimulus was heard by the subject 4 consecutive times, referred to as *trials*, corresponding to around 24 seconds of EEG recordings, which is long enough for studying single-trial methods, while still making it possible to consider EEG-signal averaging techniques.

For each subject the *recording session* was divided in *sections* (see Figure 1). In each section a series of stimuli *sequences* is played. The played stimuli are randomly chosen from the stimuli dataset. Each section is actually composed of a *training* and a *test* phase. During the training phase, single instrument tracks of a given piece are played separately (solo), in a random order. Then, during the test phase, all the corresponding duo and trio variants of the same piece are played, also in a random order, but with a potentially different spatial rendering and considering a different theme of the same musical piece. This means that in some trials the subjects hear ensembles either playing the same part of the score that was played during the training phase (solo of one of the instruments of the ensemble) or a different one. Thus, a subject may hear a solo version of Theme 1 during training and a duet/trio version of

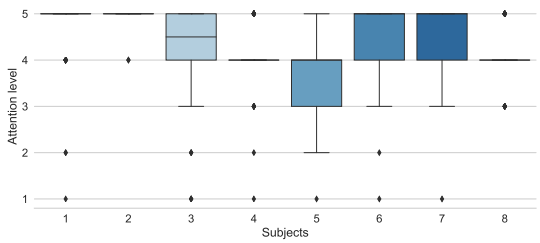

Figure 2: *Statistics of the level of attention reported by each subject for each stimulus. The scale goes from 0 to 5, where 0 represents no attention and 5 the maximum level.*

Theme 2 during testing. This is meant to allow studies on the generalization ability of an attention decoding system when the pitch contour varies between training and testing. Therefore, for each instrument solo of a given piece, between 2 and 6 mixtures where the same instrument is attended are available, but the theme and spatial rendering may be different.

A section is presented to the user through a slide-show video showing instructions, displayed as white text on a black background, asking the participant to attend to a particular instrument and visually fix a cross at the center of the screen while the stimulus is played. A "beep" precedes each stimulus launch.

3.2.1. Attention self-assessment

Right after each section, the subjects were asked to self-assess the level of attention they paid to each stimulus on a discrete scale ranging from 1 to 5. The level of attention was generally high, with the exception of only a few stimuli (see Figure 2) which can be used to evaluate how the performance of an attention decoding system changes with the attention self-assessed by the subjects. At the end of the session the participants were also asked to indicate the level of fatigue and stress experienced (low, medium or high), and if they had comments or remarks on the whole process. In general, the fatigue/stress experienced is reasonable ensuring that the quality of the collected data is good since the subjects were not overloaded.

3.2.2. EEG signal artifacts

Blinking, heartbeat and other involuntary movements greatly modify the EEG recordings while being totally independent of the stimuli and can therefore bias the interpretation of the signals recorded. Thus, subjects were instructed to maintain, for the duration of each trial, visual fixation on a cross at the center of the screen and to minimize eye blinking and all other motor activities. Moreover, during breaks at the beginning, middle and end of the experiment, a series of instruction videos were used to ask the participants to perform different gestures (shake their cheeks, blink their eyes ...) each of which has a particular influence on the EEG. This portion of EEG signals are also available within the dataset and can be used by those who are interested in studying artifact removal techniques, possibly to use them on the music-related portions of this EEG dataset.

3.3. Participants

A total of 8 subjects (7 males, 1 female), aged 23-54 years (mean age 28), all but one right-handed, participated to the experiment. Subjects reported no history of hearing impairments or neurological disorder. Before starting the acquisitions, all the participants were informed about the modalities and purposes of the experiment, to which they agreed by signing an informed consent in accordance with the Declaration of Helsinki. The participants were hired within our laboratory and they took part to the experiments as volunteers: 2 were PhD students, 5 were Master students and 1 was a sound engineer.

All of them were non-professional musicians with different levels/years of playing experience. However, they all defined themselves as beginners. Five out of them play the Guitar, one the Bass, one the Drums and one is a multi-instrumentalist playing Drums, Guitar and Bass. They all regularly practice with their instruments, study music theory and listen to music. Thus, all of them can be considered to be very familiar with the instruments present in the pop mixtures. When asked how familiar they were with the classical music instruments in the dataset, not everyone was familiar with the Bassoon, the French Horn and the Oboe. However, participants were properly trained before the experiment in recognizing the instruments they were not familiar with, using other musical excerpts than the ones used during the recordings.

3.4. EEG recording equipment

A B-Alert X24[1] headset was used to record the surface EEG, EOG (Electrooculogram), EMG (Electromyogram) and ECG (Electrocardiogram) of the participants, as well as their head motion acceleration, thanks to an integrated inertial measurement unit, at a sampling frequency $f_s = 256Hz$. The headset consists of a wireless digital acquisition unit connected to an electrode strip. The strip used has electrodes F1, F2, F3, F4, Fz, C1, C2, C3, C4, Cz, CPz, P1, P2, P3, P4, Pz, POz, O1, O2 and Oz, placed according to the 10-20 montage system. Active electrodes were referenced to left mastoid in an unipolar setting. The impedance of all the electrodes was kept below $40k\Omega$.
A custom software interface running on a 64-bit operating system was used to automatize the whole acquisition process and save the necessary information to synchronize the stimuli and the EEG responses. An External Sync Unit (ESU) receives data from the EEG headset via Bluetooth and passes it over to the acquisition software along with timestamps associated to each EEG signal sample. This ESU can also receive custom experimenter's auxiliary data and record it along the EEG data. We use this feature of the ESU to accurately record stimulus playback start times. Thus, the beginning of the stimulus playback is detected in real-time through a Python script monitoring the playback PC soundcard output. These playback start-events are then sent through the PC's serial port to the ESU so they can be marked as timestamps for the stimuli. This is done, to detect the exact time instant when each stimulus starts, within a 10-ms tolerance-window. The EEG and the stimuli timestamps are thus saved by the EEG recording software and can be subsequently used offline for synchronization.

3.5. Audio playback

Music stimuli were presented using a Hi-Fi audio playback system (JMlab chorus lcr700 speakers and Yamaha DSP-AX2 natural Sound AV amplifier). The listener was seated at the center of the room, 2 meters from a full HD TV (165 cm) screen and 2.8 meters from the two speakers. The speakers were positioned $\pm 45^o$ along the azimuth direction relative to the listener. The spatial rendering was implemented by merely using conventional stereo panning. This means, that for each instrument in the mixture we define the left and right channels as follows:

[1] https://www.advancedbrainmonitoring.com/xseries/x24/

$\begin{bmatrix} L \\ R \end{bmatrix} = \begin{bmatrix} \alpha \\ 1-\alpha \end{bmatrix} s_i(n)$, where $\alpha \in [0,1]$ and $s_i(n)$ is the mono-channel audio track of the single instrument i.

The volume was set in such a way to be comfortable for the participants and was kept constant during the whole duration of the experiments across all sessions.

3.6. Data preparation and release

A number of pre-processing stages were undertaken in order to release the dataset. Firstly, the EEG data was visually inspected to detect anomalies and only valid recording takes are being released (e.g. subject 5 has EEG responses to 53 stimuli instead of 78). Also, the 50 Hz power-line interference was removed using a notch filter and EOG/ECG artifacts were detected and removed using independent component analysis (ICA).

All EEG recordings (raw and pre-processed), audio stimuli and behavioural data of the subjects are available on the companion website.[2] All the data was anonymized.

4. Validation experiment

4.1. Stimulus reconstruction

We validated our dataset exploiting a stimulus reconstruction approach commonly used for speech attention decoding [2–6, 24, 25], which allows for mapping the multichannel EEG recording to the spectrogram of the speech stimulus. In particular, we estimated subject-specific reconstruction filters for each instrument by training a simple linear regression model on solos with their EEG response as targets for the regression. Filters were obtained using a normalized reverse correlation to minimize the mean square error of the reconstructed spectrogram [24]. A Shrinkage regularization of the covariance matrix was used to prevent overfitting [5, 26]. The filters are assumed to be causal, considering time lags between 0 and 250 ms.

We used the pre-processed EEG time signals as responses and the audio spectrograms were computed with a hop-length equal to the ratio between the audio and the EEG sampling frequency and STFT window size equal to twice the hop-length in order to time-align the EEG signal with the audio spectrogram.

4.2. Results

Once we have obtained the reconstructed stimulus representation from the EEG, we evaluate how much it is correlated with the attended source and with the unattended source. We here consider only the cases of duets, where in the mixture we have two competitive instruments and the subject is attending only to one of the two. The Pearson's correlation coefficient r is chosen to measure the linear relationship between the two signals. This evaluation procedure is commonly used for attention decoding studies, such as in [2, 3]. In particular, we computed $r_{attended}$ between the reconstructed stimulus representation and the attended instrument one; and $r_{unattended}$ between the same reconstructed stimulus representation and the unattended instrument one. This was done for the set $\mathcal{D}$ of all the duet recordings that all the subjects had to listen to, yielding a set of pairs $\{(r^i_{attended}, r^i_{unattended})\}_{i\in\mathcal{D}}$. Considering the subset $\{r^i_{attended}\}_{i\in\mathcal{D}}$ and $\{r^i_{unattended}\}_{i\in\mathcal{D}}$, we can compute two distributions, one for the attended instruments and the other for the unattended instruments in duets. These distributions are reported in Figure 3 and can be considered different with a confidence level of 99.9% ($p < 10^{-27}$ using a Wilcoxon test). Moreover, the correlation ranges are comparable with the ones reported by [2] for speech spectrogram reconstruction. Also in Figure 3, we can see the scatter plot where each data point $r_{attended}$ versus $r_{unattended}$ for the same prediction obtained with our model from the EEG. The plots show that $r_{attended} > r_{unattended}$ in more than 78% of the tests. We can reject the hypothesis that this result was generated randomly with $p < 10^{-15}$, using a randomization test over 10000 repetitions [27, 28]. Thus, one can interpret the plot as follows: when both the $r_{attended}$ and $r_{unattended}$ coefficients are very low and similar, the quality of the reconstructed stimulus is low, so it is difficult to decode which one is the attended instrument. In fact, the majority of the cases where $r_{unattended} > r_{attended}$ are concentrated below $r = 0.2$. On the contrary, we have high $r_{unattended}$ only in few cases but the corresponding $r_{attended}$ is almost always very similar. Here the model is accounting for effects which are probably more related to the whole mixtures than individual instruments. When $r_{attended}$ is high, usually the corresponding $r_{unattended}$ is low, meaning that the model is discriminating well between the two instruments.

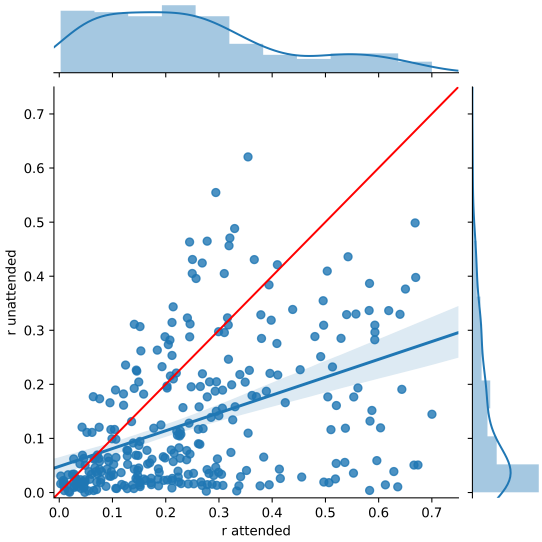

Figure 3: *Pearson's correlation coefficients with the attended instrument representation and with the unattended one across all duet trials and subjects. The corresponding distributions are reported at the margin of the scatter plot.*

5. Conclusion

MAD-EEG is a novel, open source dataset that enables studies on the problem of EEG-based decoding attention to a target instrument in realistic polyphonic music. The numerous variants in the stimuli and the behavioural data allow for investigating the impact that such factors may have on attention decoding.

We validated the usefulness of MAD-EEG for attention decoding studies, which distinguishes it from other music-related EEG datasets. We showed that even with a really simple linear regression model it is possible to reconstruct from the EEG response an audio representation that is more correlated with the attended instrument than with the unattended one, meaning that the EEG tracks relevant information of the attended source.

In future works, we will extend the dataset not only in terms of number of EEG recordings, but also in terms of variants present in the stimuli and behavioural data.

[2] https://zenodo.org/record/4537751.YWgAVRpBwuV

6. References

[1] B. N. Pasley, S. V. David, N. Mesgarani, A. Flinker, S. A. Shamma, N. E. Crone, R. T. Knight, and E. F. Chang, "Reconstructing speech from human auditory cortex," *PLoS biology*, vol. 10, no. 1, p. e1001251, 2012.

[2] N. Mesgarani and E. F. Chang, "Selective cortical representation of attended speaker in multi-talker speech perception," *Nature*, vol. 485, no. 7397, p. 233, 2012.

[3] 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*, vol. 25, no. 7, pp. 1697–1706, 2014.

[4] J. A. O'Sullivan, R. B. Reilly, and E. C. Lalor, "Improved decoding of attentional selection in a cocktail party environment with eeg via automatic selection of relevant independent components," in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 5740–5743.

[5] M. J. Crosse, G. M. Di Liberto, A. Bednar, and E. C. Lalor, "The multivariate temporal response function (mtrf) toolbox: a matlab toolbox for relating neural signals to continuous stimuli," *Frontiers in human neuroscience*, vol. 10, p. 604, 2016.

[6] S. A. Fuglsang, T. Dau, and J. Hjortkjær, "Noise-robust cortical tracking of attended speech in real-world acoustic scenes," *Neuroimage*, vol. 156, pp. 435–444, 2017.

[7] I. Sturm, M. Treder, D. Miklody, H. Purwins, S. Dähne, B. Blankertz, and G. Curio, "Extracting the neural representation of tone onsets for separate voices of ensemble music using multivariate eeg analysis." *Psychomusicology: Music, Mind, and Brain*, vol. 25, no. 4, p. 366, 2015.

[8] I. Sturm, S. Dähne, B. Blankertz, and G. Curio, "Multi-variate eeg analysis as a novel tool to examine brain responses to naturalistic music stimuli," *PloS one*, vol. 10, no. 10, p. e0141281, 2015.

[9] R. S. Schaefer, P. Desain, and J. Farquhar, "Shared processing of perception and imagery of music in decomposed eeg," *Neuroimage*, vol. 70, pp. 317–326, 2013.

[10] A. Ofner and S. Stober, "Shared generative representation of auditory concepts and eeg to reconstruct perceived and imagined music," *International Society for Music Information Retrieval Conference (ISMIR)*, 2018.

[11] F. Cong, A. H. Phan, Q. Zhao, A. K. Nandi, V. Alluri, P. Toiviainen, H. Poikonen, M. Huotilainen, A. Cichocki, and T. Ristaniemi, "Analysis of ongoing eeg elicited by natural music stimuli using nonnegative tensor factorization," in *Proc. of the 20th European Signal Processing Conference (EUSIPCO)*. IEEE, 2012, pp. 494–498.

[12] M. H. Thaut, "Rhythm, human temporality, and brain function," *Musical communication*, pp. 171–191, 2005.

[13] 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*, vol. 5, p. 742, 2014.

[14] S. Stober, T. Prätzlich, and M. Müller, "Brain beats: Tempo extraction from eeg data." in *International Society for Music Information Retrieval Conference (ISMIR)*, 2016, pp. 276–282.

[15] C. J. Plack, D. Barker, and D. A. Hall, "Pitch coding and pitch processing in the human brain," *Hearing Research*, vol. 307, pp. 53–64, 2014.

[16] A. Caclin, M.-H. Giard, B. K. Smith, and S. McAdams, "Interactive processing of timbre dimensions: A garner interference study," *Brain research*, vol. 1138, pp. 159–170, 2007.

[17] A. de Cheveigné, D. D. Wong, G. M. Di Liberto, J. Hjortkjær, M. Slaney, and E. Lalor, "Decoding the auditory brain with canonical component analysis," *NeuroImage*, vol. 172, pp. 206–216, 2018.

[18] M. S. Treder, H. Purwins, D. Miklody, I. Sturm, and B. Blankertz, "Decoding auditory attention to instruments in polyphonic music using single-trial eeg classification," *Journal of neural engineering*, vol. 11, no. 2, p. 026009, 2014.

[19] B. Kaneshiro, D. T. Nguyen, J. P. Dmochowski, A. M. Norcia, and J. Berger, "Naturalistic music eeg dataset - hindi (nmed-h)," https://purl.stanford.edu/sd922db3535, 2016.

[20] S. Losorelli, D. T. Nguyen, J. P. Dmochowski, and B. Kaneshiro, "Nmed-t: A tempo-focused dataset of cortical and behavioral responses to naturalistic music." in *ISMIR*, 2017, pp. 339–346.

[21] J. Appaji and B. Kaneshiro, "Neural tracking of simple and complex rhythms: Pilot study and dataset," *Late-Breaking Demos Session for ISMIR*, 2018.

[22] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[23] S. Stober, A. Sternin, A. M. Owen, and J. A. Grahn, "Towards music imagery information retrieval: Introducing the openmiir dataset of eeg rec.ings from music perception and imagination." in *International Society for Music Information Retrieval Conference (ISMIR)*, 2015, pp. 763–769.

[24] N. Mesgarani, S. V. David, J. B. Fritz, and S. A. Shamma, "Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex," *Journal of neurophysiology*, 2009.

[25] J. Vanthornhout, L. Decruy, J. Wouters, J. Z. Simon, and T. Francart, "Speech intelligibility predicted from neural entrainment of the speech envelope," *Journal of the Association for Research in Otolaryngology*, pp. 1–11, 2018.

[26] D. D. Wong, S. A. Fuglsang, J. Hjortkjær, E. Ceolini, M. Slaney, and A. de Cheveigné, "A comparison of temporal response function estimation methods for auditory attention decoding," *bioRxiv*, p. 281345, 2018.

[27] A. Yeh, "More accurate tests for the statistical significance of result differences," in *Proc. of the 18th conference on Computational linguistics-Volume 2*. Association for Computational Linguistics, 2000, pp. 947–953.

[28] E. W. Noreen, *Computer-intensive methods for testing hypotheses*. Wiley New York, 1989.