Tempo Estimation from the EEG signal during perception and imagination of music

Avital Sternin, Sebastian Stober, Adrian M. Owen, and Jessica A. Grahn *

Brain and Mind Institute, Department of Psychology
The University of Western Ontario
{asternin, sstober, adrian.owen, jgrahn}@uwo.ca

Abstract. Electroencephalography (EEG) recordings taken during the perception and the imagination of music contain information to estimate the tempo of a musical piece. Five participants listened to and imagined 12 short clips taken from familiar musical pieces – each 7s-16s long. Basic EEG preprocessing techniques were used to remove artifacts and a dynamic beat tracker was used to estimate average tempo. Autocorrelation curves were computed to investigate the periodicity seen in the average EEG waveforms, and the peaks from these curves were found to be proportional to stimulus measure length. As the tempo at which participants imagine may vary over time we used an aggregation technique that allowed us to estimate the tempo over the course of an entire trial. We propose future directions involving convolutional neural networks (CNNs) that will allow us to apply our results to build a brain-computer interface.

Keywords: Electroencephalography, Music Imagination, Music Information Retrieval, Brain-Computer Interface

1 Introduction

Everybody imagines music. Imagining music can be defined as a deliberate internal recreation of the perceptual experience of listening to music [1]. 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 and anyone who has ever had an "ear-worm" – a tune stuck in their head – has experienced imagining music.

Our goal is to use the brain signals that occur during the imagination of music to build a music-based brain computer interface (BCI). Recently studies have identified a close relationship between the brain areas that are active during the imagination and the perception of music [2,3,4,5]. We will capitalize on this close relationship to decode the brain activity that occurs during imagination using the activity that occurs during

* This work was supported by a fellowship within the Postdoc-Program of the German Academic Exchange Service (DAAD) (S.S.), the Canada Excellence Research Chairs (CERC) Program (A.M.O), an National Sciences and Engineering Research Council (NSERC) Discovery Grant and an Ontario Early Researcher Award (J.A.G.)

perception. Brain signals will be recorded during the perception of music in order to determine how the patient's brain responds to a stimulus. The patient will then imagine a piece of music and we will compare the signal collected during imagination to the signal collected during perception. The similarities will be used to determine what piece of music they imagined. Having the ability to decode music imagination and assign meaning to brain responses will lead to the building of a reliable BCI. This interface will allow for communication with patients who have no way to express themselves behaviourally due to various medical conditions (e.g. locked-in or minimally conscious patients). Similar communication devices have been developed that rely on the imagination of motor movements [6], but the proposed interface will be novel in its use of music imagination.

In developing this new BCI, we explore what characteristics of music create the most distinguishable changes in brain state. These characteristics include time signature, the presence/absence of lyrics, tempo, instrumentation, etc. By determining what makes a piece of music easy to classify we will ensure that our BCI will be robust and reliable. In this paper we present our findings regarding estimating music tempo from the EEG signal.

2 Related Work

The nature of our stimuli (pieces of music) need to be taken into consideration when choosing a suitable technique for our BCI. Previous decoding studies have used hemodynamic brain imaging techniques such as functional magnetic resonance imaging (fMRI) [2,3,6] to classify brain states. However, music unfolds temporally and hemodynamic techniques have poor temporal resolution. Another drawback to hemodynamic techniques is that they require expensive equipment and are not easily transported to the bedside of a patient. For a communication device to be widely used, a technique that is inexpensive and portable, like EEG, is ideal.

EEG has been used to investigate brain responses to auditory rhythms through frequency band analysis. Oscillatory neural activity in the gamma (20-60 Hz) band is sensitive to accented tones in a rhythmic sequence and anticipates isochronous tones [7]. Oscillations in the beta (20-30 Hz) band increase in anticipation of strong tones in isochronous and non-isochronous sequences [8,9,10]. Another approach to investigating music measures the magnitude of steady state evoked potentials (SSEPs) (reflecting neural oscillations entrained to the stimulus) while listening to rhythmic sequences [11].

EEG has also been used to investigate classification of perceived melodies [1]. Very short melodies ($\sim$3 secs long) were heard more than 140 times and the brain waves collected during each stimulus presentation were averaged together. These melodies produced unique waveforms, or event-related potentials (ERPs), in the EEG allowing the piece of music to be reliably identified with a single trial accuracy of 70%. However, perception of music is not useful for constructing a reliable BCI because perception is controlled by the experimenter and gives no indication of what a patient may want to communicate. Patients do have control over what they imagine, and a recent meta-analysis [12] has shown that EEG is capable of detecting brain activity during the *imagination* of music, supporting our decision to use EEG for a music-imagination-based BCI.

Before we can classify our stimuli we need to determine what characteristics in the data we are able to identify. Our analyses here will focus on estimating the tempo of the music stimulus from the EEG signal. It is possible to enhance the amplitude of SSEPs at frequencies related to specific tempos by imagining beats in groups of 2 or 3 [11], and we hope that the imagination of music will have a similar effect allowing us to pick up on the frequency of the beats in the music and therefore determine the tempo.

3 Methods

3.1 Participants

Five participants (1 male), aged 19-36, with normal hearing and no history of brain injury took part in this study. Four participants had formal musical training (1- 26 years), but none of the participants played instruments regularly at the time of data collection.

3.2 Stimuli

Stimuli were fragments of familiar musical pieces and were selected based on key signature (3/4 or 4/4 time) and the presence and absence of lyrics. The stimuli were kept as similar in length as possible with care taken to ensure that they all contained complete musical phrases. Stimulus details can be found in Table 1. Each musical fragment was preceded by approximately two seconds of clicks as a cue to the tempo and onset of the music. The beats began to fade out at the one second mark and stopped at the onset of the music.

3.3 Equipment and Procedure

We collected information about participants' previous music experience, their ability to imagine sounds, and musical sophistication using an adapted version of the Goldsmith's Musical Sophistican Index [13] combined with a clarity of auditory imagination scale. Participants also performed a beat tapping task to determine whether they were able to find and keep a beat. In order to participate in the EEG portion of the study participants needed to be familiar with 80% of the stimuli and to be able to tap along consistently to 90% of the stimuli. Participants were seated in an audiometric room (Eckel model CL-13) and the data were collected using a BioSemi Active-Two system with 64+2 EEG channels. Horizontal and vertical EOG channels were used to record eye movements. EEG was sampled at 512 Hz. A Cedrus StimTracker was used to ensure minimal delay ($<$0.05 ms) between the presentation of the stimulus to the participant and the marking of stimulus onset in the data. The experiment was programmed and presented using PsychToolbox run in MATLAB 2014a. A computer monitor displayed the instructions and fixation cross and speakers played the stimuli at a comfortable volume for each participant. The volume was kept constant across participants

In the first part of the task the perception of a music piece was paired directly with imagination. In the first condition, the song was presented preceded by 2 bars of clicks (cue). In the second condition, only the cue was played and participants were required to imagine the music on their own when the cue ended. In the third condition, no auditory cue was played and participants were asked to imagine the song at the presentation of a fixation cross. These three conditions appeared in the same order for every song. The order of the songs was randomized and songs were presented 5 times within this block.

Table 1. Information about the tempo, meter and length of the stimuli used in this experiment.

ID	Name	Meter	Length	Tempo	#Bars	Bar Length
1	Chim Chim Cheree (lyrics)	3/4	13.3s	212 BPM	16	0.85s
2	Take Me Out to the Ballgame (lyrics)	3/4	7.7s	189 BPM	8	0.95s
3	Jingle Bells (lyrics)	4/4	9.7s	200 BPM	8	1.20s
4	Mary Had a Little Lamb (lyrics)	4/4	11.6s	160 BPM	8	1.50s
11	Chim Chim Cheree	3/4	13.5s	212 BPM	16	0.85s
12	Take Me Out to the Ballgame	3/4	7.7s	189 BPM	8	0.95s
13	Jingle Bells	4/4	9.0s	200 BPM	8	1.20s
14	Mary Had a Little Lamb	4/4	12.2s	160 BPM	8	1.50s
21	Emperor Waltz	3/4	8.3s	178 BPM	8	1.01s
22	Hedwig's Theme (Harry Potter)	3/4	16.0s	166 BPM	15	1.08s
23	Imperial March (Star Wars Theme)	4/4	9.2s	104 BPM	4	2.30s
24	Eine Kleine Nachtmusik	4/4	6.9s	140 BPM	4	1.71s

In the second part of the task, only the third condition was presented to participants. There was no auditory cue. Again, the order of the stimuli was randomized and songs were presented 5 times within this block. After each song was imagined, participants reported whether or not they felt they had imagined the song correctly.

4 Analysis

We are interested in estimating the tempo of the music stimuli from the EEG signal. Our analyses here focus on the data gathered during the three conditions in the first block of the experiment. Participants reported that they were not confident in their performance on the trials collected from the second block of the experiment, therefore the data were excluded from analyses and will require further attention. Basic EEG preprocessing techniques were used to remove artifacts and a dynamic beat tracker was used to estimate average tempo. Autocorrelation curves were computed to investigate the periodicity seen in the average EEG waveforms and the peaks from these curves were found to be proportional to stimulus bar length. To deal with the tempo variance that inevitably occurred during the music imagination trials, we used a sliding window to aggregate over all channels and computed a more accurate tempo value.

4.1 Preprocessing

The raw EEG and EOG data were processed using the python MNE [14] toolbox[1]. We applied a fft-bandpass filter keeping a frequency range between 0.5 and 30Hz. Afterwards, we down-sampled to a sampling rate of 64Hz. To remove artifacts caused by eye blinks, we computed independent components using the FastICA [15] algorithm and removed components that had a high correlation with the EOG channels. Finally, the 64 EEG channels were reconstructed from the remaining independent components without further dimensionality reduction.

[1] https://github.com/mne-tools/mne-python/

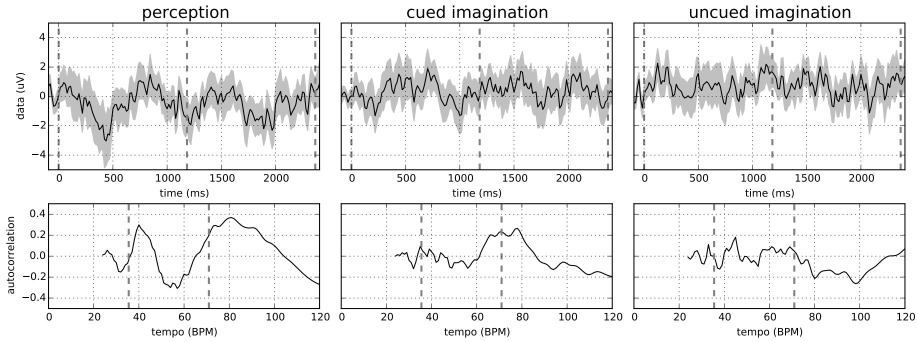

Fig. 1. Top: Mean and standard deviation over all 64 EEG channels of the bar-aligned ERPs (without epochs overlap) for "Chim Chim Cheree (lyrics)" in conditions 1–3 for participant P01. Each ERP was averaged over 25 epochs (5 from each trial). Bottom: Corresponding autocorrelation scores in the relevant tempo range. Dashed lines indicate downbeats (top) and the approximate bar tempo of the stimulus plus its lower tempo harmonic (bottom).

4.2 Analysis of Beat and Bar-Aligned ERPs

We analyzed the music stimuli with the dynamic beat tracker [16] provided by the *librosa* library.[2] This way, we obtained an estimation of the average tempo as well as annotations for all beats. The quality of these automatic annotations was verified through sonification.

Given the beat annotations of the stimuli and assuming that our participants would imagine the stimuli at a similar tempo, we computed bar-aligned ERPs using non-overlapping epochs from 100ms before to 2.4s after a downbeat annotation. This length was required to capture slightly more than a single bar for the slowest stimulus – number 23 with a bar length of more than 2.3s. As expected, the resulting averaged ERPs differ considerably between participants, stimuli, and conditions. However, we often observed a periodicity in the averaged signal proportional to the bar length. Figure 1 shows example ERPs for a specific participant and stimulus where this is clearly visible in all conditions.

In order to analyze this periodicity, we computed the autocorrelation curves by comparing each signal with itself at a range of time lags. To this end, we aggregated all 64 EEG channels into a mean signal. We further chose time lags corresponding to the bar tempo range of our stimuli. The lower end of 24 BPM was determined by our choice of the epoch length. Using longer epochs would allow us to extend the tempo range to slower tempi, but this would be at the expense of fewer epochs available for averaging.

In general, more distinct peaks in the autocorrelation were observed in the perception condition. For the two imagination conditions, peaks were more blurred as can also be seen in Figure 1. This is most likely caused by the lack of a time locking mechanism, which allows the tempo to vary. This hypothesis is also backed by the observation that artificially jittering the bar onsets results in a decrease in autocorrelation.

Most notably, we ensured that the bar-aligned epochs did not overlap by rejecting some of the epochs. If they overlap, a single data segment can contribute to multiple

[2] https://github.com/bmcfee/librosa

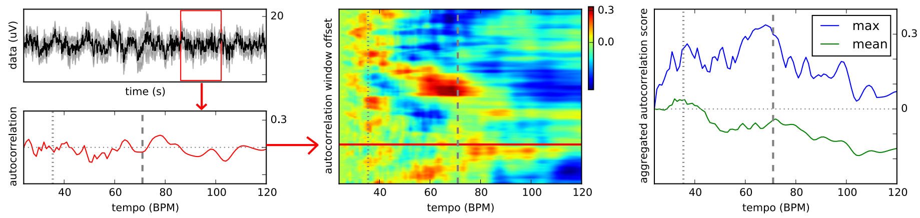

Fig. 2. Schema of the proposed tempo estimation technique. All plots refer to the first of the five trials contributing to the imagination ERP of "Chim Chim Cheree (lyrics)" in Figure 1, middle. Left top: EEG waveform (mean of all 64 channels) for the whole trial with the red box indicating the sliding window of 2.5 seconds. Left bottom: Autocorrelation curve for this specific segment of the trial. Middle: Vertically stacked autocorrelation curves for the whole trial with the red horizontal line indicating the position of the sliding window shown on the left. Right: Aggregated autocorrelation scores (mean and max) for the whole trial. Dashed vertical lines indicate the stimulus bar tempo. Dotted vertical lines refer to half the bar tempo.

epochs at different time points. This can induce misleading autocorrelation peaks that are not supported by the raw data.

4.3 Analyzing Single Trials

With the interesting observations reported in the preceding section, the question arises whether the tempo could similarly be estimated through autocorrelation from single trials. Single trial tempo estimation is important for our BCI in order to reduce patient fatigue during use. Here, we face several challenges. First, there are too few bar-aligned epochs in a single trial to use ERPs. Second, we neither know the tempo of the stimulus nor do we have beat annotations available. Therefore, there are no reference points for extracting bar-aligned epochs. Moreover, we wanted to address the problem of possible tempo variance in the imagination conditions.

Figure 2 illustrates our proposed solution to this problem. We move a 2.5-second sliding window over the mean EEG signal aggregated over all channels. At each position with a hop size of 5 samples at 64Hz, we compute the autocorrelation curve. The curves for the individual window segments are stacked into a 2-dimensional matrix with the first dimension corresponding to the window offset in the trial signal and the second dimension corresponding to the possible tempo values. Hence, each matrix value holds the score for a certain tempo at one specific point in the trial. The scores in the matrix are finally aggregated deriving an estimated tempo value for the trial. Figure 2 shows how the mean and maximum values are aggregated over all matrix rows.

Ideally, the score for the actual tempo should be stable throughout the whole trial. However, we observed substantial fluctuation within many trials. While the mean and maximum over all matrix rows often produce significant peaks in the aggregated autocorrelation curve, the following heuristic has led to slightly more stable results:

1. In each row, find the pair of tempo values with the maximal combined score.
2. Select the median of all selected pairs.

3. From this pair, return the tempo value with the higher mean value over all rows.

For the evaluation of our heuristic, we computed the mean absolute error of the estimated tempo and the actual tempo. We also considered the tempo harmonic below and above the correct value, i.e. half or twice the tempo, as a correct result. The prediction error, averaged across all stimuli, varied considerably between participants ranging from 7.07, 7.15, and 8.11 in the three conditions for participant P14 up to 9.81, 10.04, and 12.58 for P12. Figure 3 summarizes the errors for each stimulus over all trials of all participants. The figure clearly shows a trend that tempo is easier to predict for some stimuli, such as "Chim Chim Cheree" (ID 1 and 11) and "Mary had a little lamb" (ID 4 and 14), than for others. The slowest stimulus, the "Imperial March" (ID 23) has the highest variation of prediction accuracy. Although it is too early to draw definitive conclusions, the data also suggest that single-trial tempo estimation does not depend on the condition (perception or imagination) but rather relies on the tempo of a stimulus.

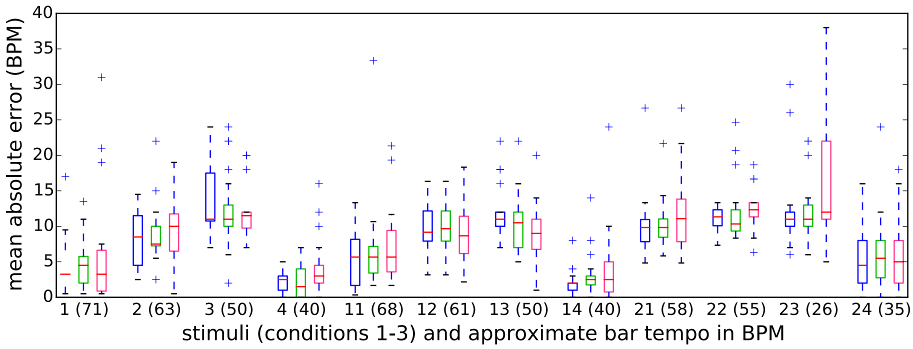

Fig. 3. Aggregated errors of single trial tempo prediction over all 5 participants. The true tempo is indicated in brackets with the mean absolute error for each stimulus indicated with a red bar. Each box contains 50% of the error values for perception (blue), cued imagination (green) and uncued imagination (magenta). The whiskers contain an additional 25% of the error values. Each cross indicates an outlier.

5 Discussion

In the current experiment we have very little information about when a participant starts or finishes imagining a piece of music. Although we have given participants an auditory cue that signals when to begin we don't know that they are complying with the instructions. It is also inevitable that participants will imagine music at different rates resulting in different end times. To control for these differences in the future we will ask participants to signal the start and finish of their music imagination with a button press. This will make the estimation of tempo for the EEG signal easier to accomplish.

We see a high potential of improving the aggregation methods used to combine information from multiple channels in a trial and for the score matrix (Figure 2, middle). In a pilot test, we have obtained encouraging first results using convolutional neural networks (CNNs), a popular technique from the field of deep learning, to predict the tempo based on the score matrix. We further plan to investigate along this path, possibly also integrating the computation of the score matrix and the preceding channel aggregation step into the neural network for a holistic modeling and training approach.

To promote further research and improvements in this direction the data collected in the experiment and the related Matlab and Python code is available under the Open Data Commons Public Domain Dedication and License (PDDL): http://dx.doi.org/10.6084/m9.figshare.1404206.

References

1. Schaefer, R.S., Farquhar, J., Blokland, Y., Sadakata, M., Desain, P.: Name that tune: decoding music from the listening brain. NeuroImage **56** (2011) 843–9
2. Halpern, A.R., Zatorre, R.J., Bouffard, M., Johnson, J.A.: Behavioral and neural correlates of perceived and imagined musical timbre. Neuropsychologia **42** (2004) 1281–92
3. Kraemer, D.J.M., Macrae, C.N., Green, A.E., Kelley, W.M.: Musical imagery: sound of silence activates auditory cortex. Nature **434** (2005) 158
4. Herholz, S., Lappe, C., Knief, A., Pantev, C.: Neural basis of music imagery and the effect of musical expertise. The European journal of neuroscience **28** (2008) 2352–60
5. Herholz, S., Halpern, A., Zatorre, R.: Neuronal correlates of perception, imagery, and memory for familiar tunes. Journal of cognitive neuroscience **24** (2012) 1382–97
6. Owen, A.M., Coleman, M.R., Boly, M., Davis, M.H., Laureys, S., Pickard, J.D.: Detecting awareness in the vegetative state. Science (New York, N.Y.) **313** (2006) 1402
7. Snyder, J.S., Large, E.W.: Gamma-band activity reflects the metric structure of rhythmic tone sequences. Cognitive Brain Research **24** (2005) 117–126
8. Iversen, J.R., Repp, B.H., Patel, A.D.: Top-down control of rhythm perception modulates early auditory responses. Annals of the New York Academy of Sciences **1169** (2009) 58–73
9. Fujioka, T., Trainor, L.J., Large, E.W., Ross, B.: Beta and gamma rhythms in human auditory cortex during musical beat processing. Annals of the New York Academy of Sciences **1169** (2009) 89–92
10. Fujioka, T., Trainor, L.J., Large, E.W., Ross, B.: Internalized Timing of Isochronous Sounds Is Represented in Neuromagnetic Beta Oscillations. Journal of Neuroscience **32** (2012) 1791–1802
11. Nozaradan, S., Peretz, I., Missal, M., Mouraux, A.: Tagging the neuronal entrainment to beat and meter. The Journal of neuroscience : the official journal of the Society for Neuroscience **31** (2011) 10234–10240
12. Schaefer, R.S., Desain, P., Farquhar, J.: Shared processing of perception and imagery of music in decomposed EEG. NeuroImage **70** (2013) 317–26
13. Müllensiefen, D., Gingras, B., Musil, J., Stewart, L.: The musicality of non-musicians: An index for assessing musical sophistication in the general population. PLoS ONE **9** (2014) e89642
14. Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hämäläinen, M.S.: {MNE} software for processing{MEG} and{EEG} data. NeuroImage **86** (2014) 446 – 460
15. Hyvärinen, a., Hyvärinen, a., Oja, E., Oja, E.: Independent component analysis: algorithms and applications. Neural networks : the official journal of the International Neural Network Society **13** (2000) 411–30
16. Ellis, D.P.W.: Beat tracking by dynamic programming. Journal of New Music Research **36** (2007) 51–60