BRAIN BEATS: TEMPO EXTRACTION FROM EEG DATA

Sebastian Stober[1] **Thomas Prätzlich**[2] **Meinard Müller**[2]

[1] Research Focus Cognititive Sciences, University of Potsdam, Germany

[2] International Audio Laboratories Erlangen, Germany

sstober@uni-potsdam.de, {thomas.praetzlich, meinard.mueller}@audiolabs-erlangen.de

ABSTRACT

This paper addresses the question how music information retrieval techniques originally developed to process audio recordings can be adapted for the analysis of corresponding brain activity data. In particular, we conducted a case study applying beat tracking techniques to extract the tempo from electroencephalography (EEG) recordings obtained from people listening to music stimuli. We point out similarities and differences in processing audio and EEG data and show to which extent the tempo can be successfully extracted from EEG signals. Furthermore, we demonstrate how the tempo extraction from EEG signals can be stabilized by applying different fusion approaches on the mid-level tempogram features.

1 Introduction

Recent findings in cognitive neuroscience suggest that it is possible to track a listener's attention to different speakers or music signals [1,24], or to identify beat-related or rhythmic features in electroencephalography (EEG) recordings [1] of brain activity during music perception. In particular, it has been shown that oscillatory neural activity is sensitive to accented tones in a rhythmic sequence [19]. Neural oscillations entrain (synchronize) to rhythmic sequences [2,14] and increase in anticipation of strong tones in a non-isochronous (not evenly spaced), rhythmic sequence [3,4,10]. When subjects hear rhythmic sequences, the magnitude of the oscillations changes for frequencies related to the metrical structure of the rhythm [16,17].

EEG studies [5] have further shown that perturbations of the rhythmic pattern lead to distinguishable electrophysiological responses—commonly referred to as event-related potentials (ERPs). This effect appears to be independent of the listener's level of musical proficiency. Furthermore, [26] showed that accented (louder) beats imagined by a listener on top of a steady metronome beat can be recognized

---

[1] Electroencephalography (EEG) is a non-invasive brain imaging technique that relies on electrodes placed on the scalp to measure the electrical activity of the brain. A recent review of neuroimaging methods for music information retrieval (MIR) that also includes a comparison of EEG with different approaches is given in [11].
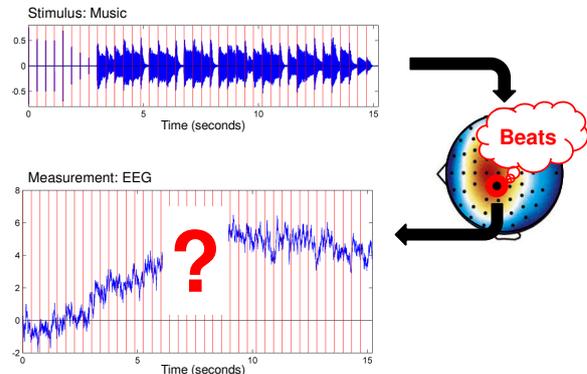
Figure 1. Question: Can we extract the tempo of a music recording from brain activity data (EEG) recorded during listening? The red vertical lines in the audio waveform (top) and the EEG signal (bottom) mark the beat positions.

from ERPs. EEG signals have also been used to distinguish perceived rhythmic stimuli [21] with convolutional neural networks. First preliminary results using autocorrelation for tempo estimation from the EEG signal during perception and imagination of music have been reported in [20].

This raises the question whether MIR techniques originally developed to detect beats and extract the tempo from music recordings could also be used for the analysis of corresponding EEG signals. One could argue that as the brain processes the perceived music, it generates a transformed representation which is captured by the EEG electrodes. Hence, the recorded EEG signal could in principle be seen as a mid-level representation of the original music piece that has been heavily distorted by two consecutive black-box filters—the brain and the EEG equipment.

This transformation involves and intermingles with several other brain processes unrelated to music perception and is limited by the capabilities of the recording equipment that can only measure cortical brain activity (close to the scalp). It further introduces artifacts caused by electrical noise or the participant's movements such as eye blinks. Figuratively speaking, this could be compared to a cocktail-party situation where the listener is not in the same room as the speakers but in the next room separated by a thick wall.

In this paper, we address the question whether well-established tempo and beat tracking methods, originally developed for MIR, can be used to recover tempo information from EEG data recorded from people listening to music, see Figure 1. In the remainder of this paper, we
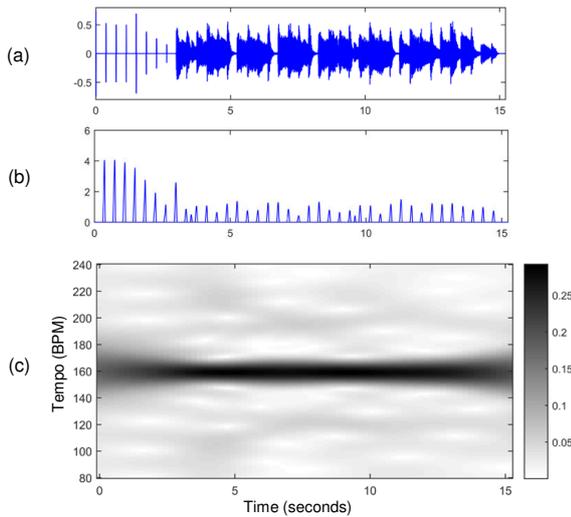
Figure 2. Tempogram computation for music signals. **(a)** Waveform signal. **(b)** Novelty curve. **(c)** Tempogram representation.

first briefly describe the EEG dataset (Section 2). As a first contribution, we explain how an MIR technique for tempo extraction can be applied on EEG signals (Section 3). Then, in Section 4, we evaluate the tempo extraction on the EEG signals by comparing it to the tempo extracted from the corresponding audio signals. As another contribution, we show that the tempo extraction on EEG signals can be stabilized by applying different fusion approaches. Finally, we conclude the paper with a summary and indication of possible research directions (Section 5).

2 Recording Setup and Dataset

In this study, we use a subset of the OpenMIIR dataset [22]—a public domain dataset of EEG recordings taken during music perception and imagination.[2] For our study, we use only the music perception EEG data from the five participants $p \in P := \{09, 11, 12, 13, 14\}$[3] who listened to twelve short music stimuli—each 7s to 16s long. These stimuli were selected from well-known pieces of different genres. They span several musical dimensions such as meter, tempo, instrumentation (ranging from piano to orchestra) and the presence of lyrics (singing or no singing present), see Table 1. All stimuli were normalized in volume and kept similar in length, while ensuring that they all contained complete musical phrases starting from the beginning of the piece. The EEG recording sessions consisted of five trials $t \in T := \{1, ..., 5\}$ in which all stimuli $s \in S := \{01, 02, 03, 04, 11, 12, 13, 14, 21, 22, 23, 24\}$ were presented in randomized order. This results in a total of $|S| \cdot |T| \cdot |P| = 12 \cdot 5 \cdot 5 = 300$ trials for the five

---

[2] The dataset is available at https://github.com/sstober/openmiir

[3] The remaining participants in the dataset had some of the stimuli presented at a slightly different tempo (c.f. [22]), which would not allow our fusion approaches discussed later in Section 4.
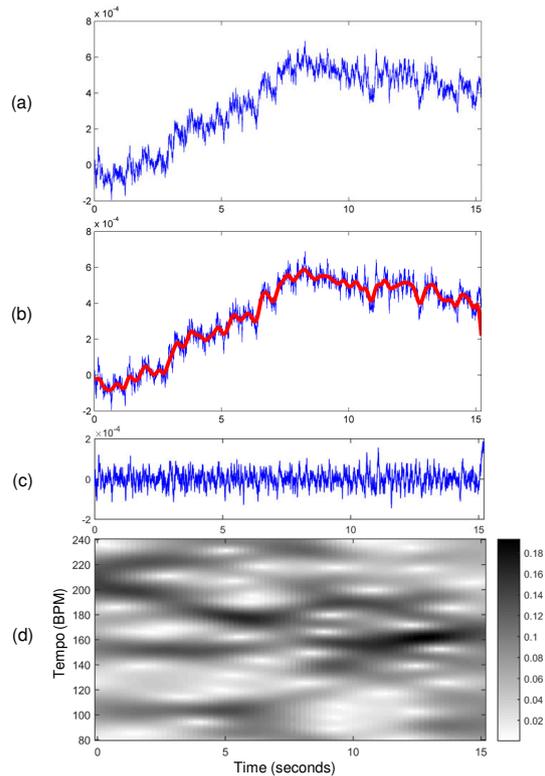
Figure 3. Tempogram computation for EEG signals. **(a)** EEG signal. **(b)** Local average curve. **(c)** Normalized EEG signal (used as novelty curve). **(d)** Tempogram representation.

participants, $|S| \cdot |T| = 12 \cdot 5 = 60$ trials per particpant, and $|P| \cdot |T| = 25$ trials per stimulus.

EEG was recorded with a BioSemi Active-Two system using 64+2 EEG channels at 512 Hz. Horizontal and vertical electrooculography (EOG) channels were used to record eye movements. As described in [22], EEG pre-processing comprised the removal and interpolation of bad channels as well as the reduction of eye blink artifacts by removing highly correlated components computed using extended Infomax independent component analysis (ICA) [12] with the MNE-python toolbox [6].

3 Computation of Tempo Information

In this section, we describe how tempo information can be extracted both from music and EEG signals. To this end, we transform a signal into a *tempogram* $\mathcal{T} : \mathbb{R} \times \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$ which is a time-tempo representation of a signal. A tempogram reveals periodicities in a given signal, similar to a spectrogram. The value $\mathcal{T}(t, \tau)$ indicates how predominant a tempo value $\tau \in \mathbb{R}_{>0}$ (measured in BPM) is at time position $t \in \mathbb{R}$ (measured in seconds) [15, Chapter 14].

In the following, we provide a basic description of the tempogram extraction for music recordings (Section 3.1) and EEG signals (Section 3.2). For algorithmic details,

we refer to the descriptions in [8, 15]. To compute the tempograms for the experiments in this work, we used the implementations from the Tempogram Toolbox.[4] Furthermore, we describe how the tempo information of a tempogram can be aggregated into a tempo histogram similar to [25] from which a global tempo value can be extracted (Section 3.3).

## 3.1 Tempogram for Music Audio Signals

To compute a tempogram, a given music audio signal is first transformed into a novelty curve capturing note onset information. In the following, we use a novelty curve computed as the positive part of a spectral flux, see [8]. Figure 2a shows the waveform of an audio stimulus, which begins with a set of cue clicks (in beats) followed by a short music excerpt of the same tempo. In Figure 2b, the novelty curve extracted from the waveform is shown. The onsets of the cue clicks are clearly reflected by peaks in the novelty curve. For the subsequent music excerpt, one can see that the peaks are similarly spaced as the cue clicks. However, there are some additional peaks in the music excerpt that correspond to additional notes or noise. Especially for music with soft onsets, the novelty curve may contain some noise in the peak structures. As for the tempo extraction, we further transform the novelty curve into an audio tempogram that reveals how dominant different tempi are at a given time point in the audio signal. In this study, we use a tempogram computed by short-term Fourier analysis of the novelty curve with a tempo window of 8 seconds, see [8] for details. The frequency axis (given in Hz) is transformed into a tempo axis (given in BPM). In Figure 2c, the audio tempogram of the example is shown, which reveals a predominant tempo of 160 BPM throughout the recording.

3.2 Tempogram for EEG Signals

In this section, we describe how we extract a tempogram from EEG signals that were measured when participants listened to a music stimulus. In principle, we use a similar approach for the tempo extraction from EEG signals as for the music recordings.

First, we aggregate the 64 EEG channels into one signal. Note that there is a lot of redundancy in these channels. This redundancy can be exploited to improve the signal-to-noise ratio. In the following, we use the channel aggregation filter shown in Figure 4. It was learned as part of a convolutional neural network (CNN) during a previous experiment attempting to recognize the stimuli from the EEG recordings [23]. In [23], a technique called "similarity-constraint encoding" (SCE) was applied that is motivated by earlier work on learning similarity measures from relative similarity constraints as introduced in [18]. The CNN
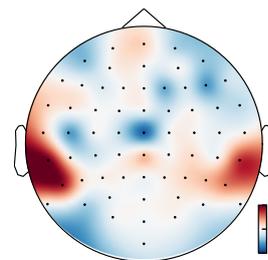
Figure 4. Topographic visualization of the SCE-trained channel aggregation filter used to compute a single signal from the 64 EEG channels (indicated by black dots). The filter consists of a weighted sum with the respective channel weights (shown in a color-coded fashion) and a subsequent application of the tanh which results in an output range of $[-1,1]$.

was trained using triplets of trials consisting of a reference trial, a paired trial from the same class (i.e., the same stimulus) and a third trial from a different class. For each triplet, the network had to predict which trial belongs to the same class as the reference trial. This way, it learned channel aggregation weights that produce signals that are most similar for trials belonging to the same class. In our earlier experiments, we found that the resulting aggregated EEG signals capture important characteristics of the music stimuli such as downbeats. We hypothesized that the learned filter from [23] could also be useful in our tempo extraction scenario, even though it is a very different task.[5]

Figure 3a shows an example of an aggregated EEG signal. From the aggregated EEG signal, we then compute a novelty curve. Here, opposed to the novelty computation for the audio signal, we assume that the beat periodicities we want to measure are already present in the time-domain EEG signal. We therefore interpret the EEG signal as a kind of novelty curve. As pre-processing, we normalize the signal by subtracting a moving average curve, see Figure 3b. This ensures that the signal is centered around zero and low frequent components of the signal are attenuated. The resulting signal (Figure 3c) is then used as a novelty curve to compute an EEG tempogram that reveals how dominant different tempi are at a given time point in the EEG signal (see Figure 3d). Note that, compared to the audio novelty curve, the EEG novelty curve is much nosier. As a result, there is more noise in the EEG tempogram compared to the audio tempogram, making it hard to determine a predominant global tempo.

3.3 Tempo Histograms

In this section, we explain how we extract a single tempo value from the audio and EEG tempograms. First, we aggregate the time-tempo information over the time by

---

[4] The Tempogram Toolbox contains MATLAB implementations for extracting various types of tempo and pulse related audio representations [9] A free implementation can be obtained at https://www.audiolabs-erlangen.de/resources/MIR/tempogramtoolbox

[5] We compared the tempo extraction on the SCE-trained channel aggregation with simply averaging the raw data across channels and found that the tempo extraction on the raw EEG data often performed roughly 10% points worse and was only on par with SCE in the best cases.
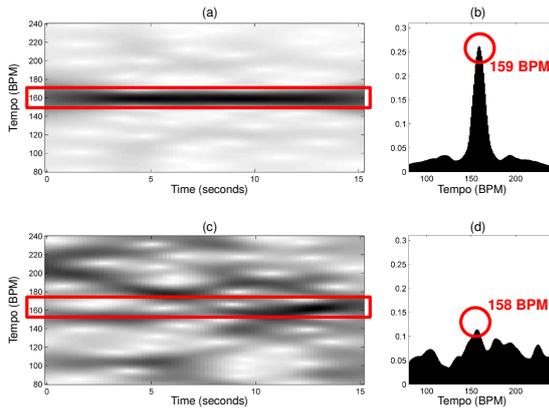
Figure 5. **(a)** Tempogram for the music signal from Figure 2 and **(b)** resulting tempo histogram. **(c)** Tempogram for EEG signal from Figure 3 and **(d)** resulting tempo histogram.

computing a tempo histogram $\mathcal{H} : \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$ from the tempogram (similar to [25]). A value $\mathcal{H}(\tau)$ in the tempo histogram indicates how present a certain tempo τ is within the entire signal. In Figure 5, a tempogram for a music recording and an EEG signal are shown along with their respective tempo histograms. In the audio tempo histogram, the highest peak at $\tau = 159$ BPM indicates the correct tempo of the music recording. The tempogram for the EEG data is much noisier, where it is hard to identify a predominant tempo from the tempogram. In the tempo histogram, however, the highest peak in the example corresponds to a tempo of 158 BPM, which is nearly the same as the main tempo obtained from the audio tempo histogram.

4   Evaluation

In this section, we report on our experiments to show to which extent the tempo extraction for the audio signals and the EEG signals are related. In the following, $\mathcal{H}_{s,p,t}$ denotes to the tempo histogram stemming from the audio stimulus $s \in S$, participant $p \in P$, and trial $t \in T$ (see Section 2). An overview of the stimuli is given in Table 1. For all experiments, we used a tempo window of 8 seconds, see [7]. Furthermore, we applied a moving average filter on the EEG data of 0.5 seconds. In Section 4.1, we introduce our evaluation measures and discuss quantitative results for different tempo extraction strategies. Then, in Section 4.2, to better understand the benefits and limitations of our approach, we look at some representative examples for tempograms and tempo histograms across the dataset.

4.1   Quantitative Results

To determine the tempo a of a given audio stimulus, we consider the highest peak in the respective audio tempo histogram $\mathcal{H}^{\mathrm{audio}}$, see Table 1. [6] The EEG tempo

---

[6] The OpenMIIR dataset also provides ground-truth tempi in the meta-data. Except for stimulus 21 with a difference of 4 BPM, our computed

Table 1. Information about the tempo, meter and length of the stimuli (with cue clicks) used in this study. Note that stimuli 1–4 and 11–14 are different versions of the same song with and without lyrics.

ID	Name	Meter	Length with cue	Tempo [BPM]
1	Chim Chim Cheree (lyrics)	3/4	14.9s	213
2	Take Me Out to the Ballgame (lyrics)	3/4	9.5s	188
3	Jingle Bells (lyrics)	4/4	12.0s	199
4	Mary Had a Little Lamb (lyrics)	4/4	14.6s	159
11	Chim Chim Cheree	3/4	15.1s	213
12	Take Me Out to the Ballgame	3/4	9.6s	188
13	Jingle Bells	4/4	11.3s	201
14	Mary Had a Little Lamb	4/4	15.2s	159
21	Emperor Waltz	3/4	10.3s	174
22	Hedwig's Theme (Harry Potter)	3/4	18.2s	165
23	Imperial March (Star Wars Theme)	4/4	11.5s	104
24	Eine Kleine Nachtmusik	4/4	10.2s	140
	mean		12.7s	175

histogram $\mathcal{H}^{\mathrm{EEG}}$ is much noisier. To obtain some insights on the tempo information contained in $\mathcal{H}^{\mathrm{EEG}}$, we look at the tempi corresponding to the highest peak as well as subsequent peaks. To this end, after selecting the tempo corresponding to the highest peak, we set the values within ± 10 BPM in the neighborhood of the peak in the tempo histogram to zero. This procedure is repeated until the top n peaks are selected. In the following, we consider the first three tempi b_1, b_2, b_3 obtained from a given tempo histogram and build the sets of tempo estimates $\mathcal{B}_1 := \{b_1\}$ (top 1 peak), $\mathcal{B}_2 := \{b_1, b_2\}$ (top 2 peaks), and $\mathcal{B}_3 := \{b_1, b_2, b_3\}$ (top 3 peaks). To determine the error of the tempo estimates $\mathcal{B}_n$ with $n \in \{1, 2, 3\}$, we compute the *minimum absolute BPM deviation* compared to the audio tempo: $\varepsilon(\mathcal{B}_n, a) := \min_{b \in \mathcal{B}_n} |b - a|$. Furthermore, as small errors are less severe as large errors, we quantify different error classes with an error tolerance $\delta \geq 0$. To this end, we compute the *BPM error rate* $E_\delta(\mathcal{B}_n)$ which is defined as the percentage of absolute BPM deviations with $\varepsilon(\mathcal{B}_n, a) > \delta$. In our experiments, we use different $\delta \in \{0, 3, 5, 7\}$ (given in BPM).

We performed the tempo extraction from the EEG tempo histograms with three different strategies:

(S1) *Single-trial tempo extraction*: For each trial, the tempo is extracted individually. This results in extracting the tempi from $|S| \cdot |P| \cdot |T| = 12 \cdot 5 \cdot 5 = 300$ tempo histograms (see Section 4).

(S2) *Fusion I*: Fixing a stimulus $s \in S$ and a participant $p \in P$, we average over the tempo histograms of the trials $t \in T$:
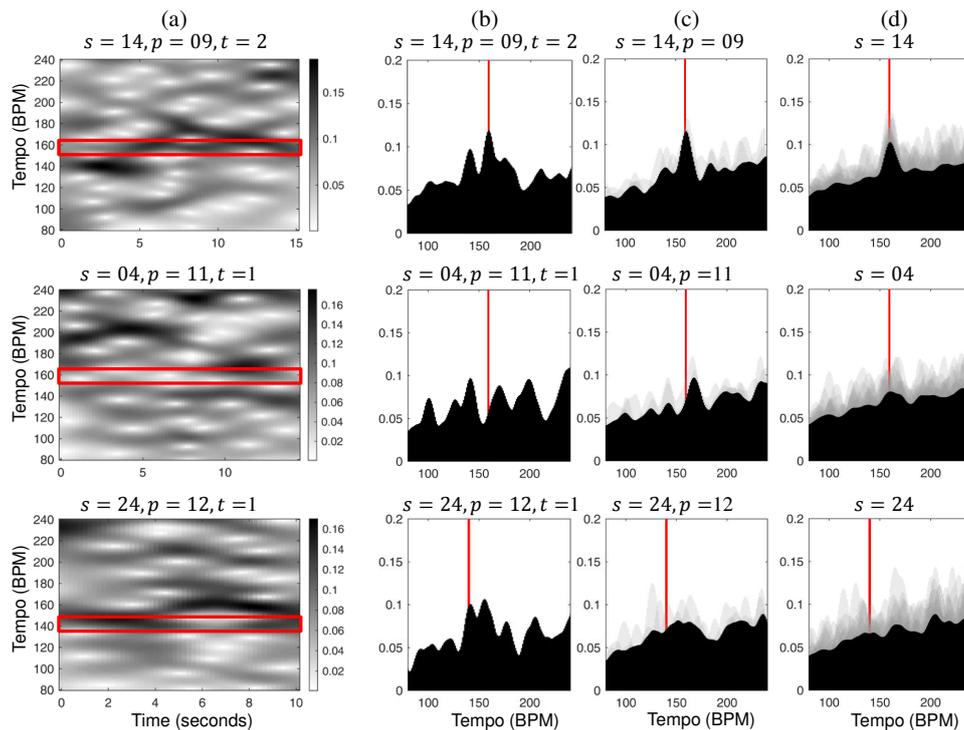
$$\mathcal{H}_{s,p}(\tau) := \frac{1}{|T|} \sum_{t \in T} \mathcal{H}_{s,p,t}(\tau).$$

This results in extracting the tempi from $|S| \cdot |P| = 60$ tempo histograms.

(S3) *Fusion II*: Fixing a stimulus $s \in S$, we average the tempo histograms over the participants $p \in P$ and the

---

tempi differed at most 1 BPM from the OpenMIIR ground-truth.
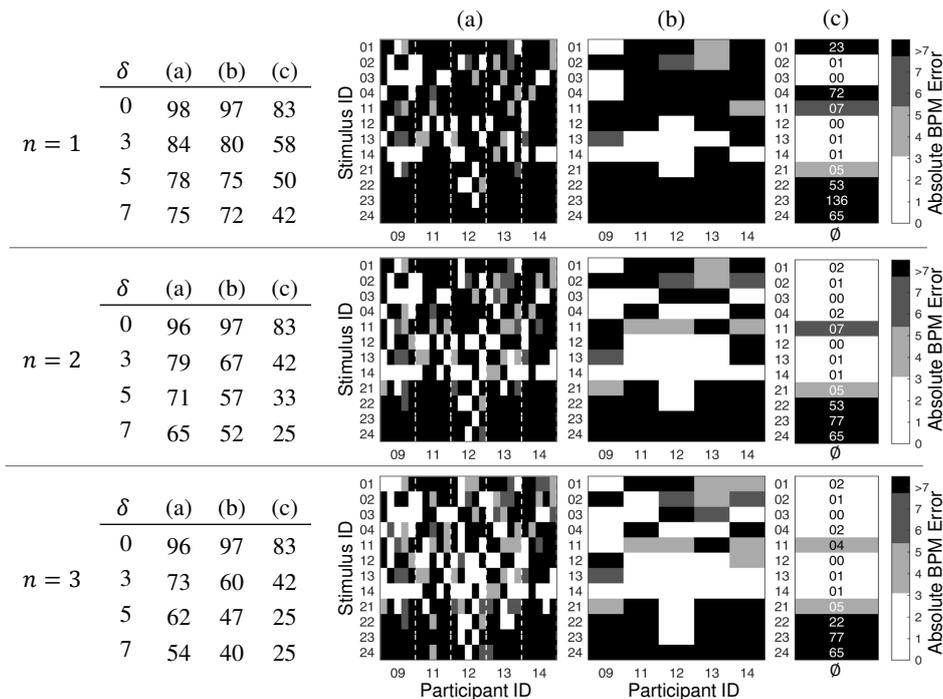
Figure 6. Tables with BPM error rates in percent (left) and absolute BPM error (right) for the set of tempo estimates $\mathcal{B}_n$. **(a)** strategy S1. Note that for each participant, there are five columns in the matrix that correspond to the different trials. **(b)** strategy S2. **(c)** strategy S3.

Figure 7. Tempograms and tempo histograms for stimuli 14, 04, and 24 (top to bottom). The red boxes and lines mark the audio tempo. The gray histograms in the background were averaged in the fusion strategies. **(a)** Tempogram for S1. **(b)** Tempo histogram for S1, derived from (a). **(c)** Tempo histogram for S2. $\mathcal{H}_{s,p}(\tau)$ was computed from five tempo histograms (5 trials). **(d)** Tempo histogram for S3. $\mathcal{H}_s(\tau)$ was computed from the 25 tempo histograms (5 participants with 5 trials each).

trials $t \in T$:

$$\mathcal{H}_s(\tau) := \frac{1}{|P| \cdot |T|} \sum_{p \in P} \sum_{t \in T} \mathcal{H}_{s,p,t}(\tau).$$

This results in extracting the tempi from $|S| = 12$ tempo histograms.

Note that it is a common approach in EEG signal processing to average the EEG signals over different trials as described in [13]. This usually reduces the noise in the signals. In this study, instead of averaging over the EEG signals, we averaged over the tempo histograms, which is a kind of mid-level feature representation.

Figure 6 shows the BPM error rates (left) as well as the absolute BPM error (right). Each row in the figure corresponds to the results for a different set of tempo estimates $\mathcal{B}_n$. For $n = 1$, a strict error tolerance of $\delta = 0$, and strategy S1, the tempo extraction basically fails, having a BPM error rate of 98%. This is not surprising, as no deviation from the audio tempo is allowed. When allowing a deviation of five BPM ($\delta = 5$), the tempo extraction using only the top peak ($n = 1$) fails in 78% of the cases. By applying the fusion strategy S2 for the tempo extraction, the BPM error rate significantly drops to 75%, which is an improvement of 3% points. The BPM error rate goes down to 50% for the fusion strategy S3 which averages over all trials for a given stimulus. This shows that averaging stabilizes the results. When looking at the results by considering the set of tempo estimates $\mathcal{B}_2$ ($n = 2$) and $\mathcal{B}_3$ ($n = 3$), we can see that the second and third peak often correspond to the expected tempo. For example, for $\delta = 5$ and strategy S3, the BPM error rate goes down from 50% (for $n = 1$), to 33% (for $n = 2$), and 25% (for $n = 3$).

Furthermore, Figure 6 shows that the results strongly depend on the music stimulus used. The extraction for stimulus $s = 14$, for example, works well for nearly all participants. This is a piece performed on a piano which has clear percussive onsets. Also, for the first eight stimuli ($01 - 04$ and $11 - 14$) the tempo extraction seems to work better than for the last four stimuli ($21 - 24$). This may have different reasons. For instance, $s = 21$, $s = 23$ and $s = 24$ are amongst the shortest stimuli in the dataset and $s = 22$ has very soft onsets. Furthermore, the stimuli 21–24 are purely instrumental (soundtracks and classical music) without lyrics.

4.2 Qualitative examples

Figure 7 shows the tempograms and tempo histograms for some representative examples. We subsequently discuss the top, middle, and bottom row of the figure corresponding to stimulus 14, 04, and 24, respectively.

The EEG tempogram shown in Figure 7a (top row) clearly reflects the correct tempo of the music stimulus. In the corresponding tempo histogram (b), a clear peak can be seen at the correct tempo. In the tempo histograms (c) and (d), corresponding to strategies S2 and S3, one can clearly see the stabilizing and noise reducing effect of the two fusion strategies, resulting in a very clear tempo peak.

In Figure 7b (middle row), the tempo histogram does not reveal the expected tempo. As also indicated by the tempogram in Figure 7a, the listener does not seem to follow the beat of the music stimulus. However, when averaging over the trials of participant $p = 11$, the tempo peak near 160 BPM becomes more dominant (see tempo histogram (c)). When averaging over all trials and all participants for the stimulus $s = 04$, the tempo peak becomes more blurry, but appears at the expected position, (see tempo histogram (d)).

For the third example in Figure 7 (bottom row), the tempogram (a) shows predominant values near the correct tempo. In the corresponding tempo histogram (b), the correct tempo is revealed by the second peak. However, the histograms for strategy S2 (c) and S3 (d) lead to very blurry peaks where the correct tempo peak is not among the top three peaks. These examples illustrate that the fusion strategies often stabilize the tempo extraction. When the data is too noisy, however, these strategies may sometimes degrade the results.

5 Conclusions

In this paper, we presented a case study where we applied an MIR tempo extraction technique, originally developed for audio recordings, to EEG signals. In experiments, we showed that it is possible to extract the tempo from EEG signals using a similar technique as for audio signals. We could see that the averaging over trials and participants typically stabilized the tempo estimation. Furthermore, we noticed that the quality of the tempo estimation was highly dependent on the music stimulus used. Exploring this effect is beyond the scope of this small study. To properly understand the reasons for this effect, a large-scale music perception experiment using stimuli with systematically adapted tempi would be needed. Possible reasons might be the complexity of the music stimuli, the presence of lyrics, the participants, or the applied methodology and techniques. Investigating these issues could be a starting point for interdisciplinary research between MIR and music perception. Supplementary material and code is available at https://dx.doi.org/10.6084/m9.figshare.3398545.

6 References

[1] A. Aroudi, B. Mirkovic, M. De Vos, and S. Doclo. Auditory attention decoding with EEG recordings using noisy acoustic reference signals. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 694–698, 2016.

[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] 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.

[4] 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.

[5] 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, January 2009.

[6] 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, December 2013.

[7] P. Grosche and M. Müller. A mid-level representation for capturing dominant tempo and pulse information in music recordings. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 189–194, 2009.

[8] P. Grosche and M. Müller. Extracting predominant local pulse information from music recordings. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(6):1688–1701, 2011.

[9] P. Grosche and M. Müller. Tempogram toolbox: Matlab implementations for tempo and pulse analysis of music recordings. In *Late-Breaking News of the International Society for Music Information Retrieval Conference (ISMIR)*, 2011.

[10] 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.

[11] B. Kaneshiro and J.P. Dmochowski. Neuroimaging methods for music information retrieval: Current findings and future prospects. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 538–544, 2015.

[12] 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.

[13] S.J. Luck. *An introduction to the event-related potential technique*. MIT press, 2014.

[14] H. Merchant, J.A. Grahn, L.J. Trainor, M. Rohrmeier, and W.T. Fitch. Finding the beat: a neural perspective across humans and non-human primates. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 370(1664), 2015.

[15] M. Müller. *Fundamentals of Music Processing*. Springer Verlag, 2015.

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

[17] 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, December 2012.

[18] M. Schultz and T. Joachims. Learning a distance metric from relative comparisons. *Advances in neural information processing systems (NIPS)*, pages 41–48, 2004.

[19] 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.

[20] A. Sternin, S. Stober, J.A. Grahn, and A.M. Owen. Tempo estimation from the EEG signal during perception and imagination of music. In *International Workshop on Brain-Computer Music Interfacing / International Symposium on Computer Music Multidisciplinary Research (BCMI/CMMR)*, 2015.

[21] 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 (NIPS)*, pages 1449–1457, 2014.

[22] S. Stober, A. Sternin, A.M. Owen, and J.A. Grahn. Towards music imagery information retrieval: Introducing the OpenMIIR dataset of EEG recordings from music perception an imagination. In *Proc. of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 763–769, 2015.

[23] S. Stober, A. Sternin, A.M. Owen, and J.A. Grahn. Deep feature learning for EEG recordings. *arXiv preprint arXiv:1511.04306*, 2015.

[24] 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*, 11(2):026009, April 2014.

[25] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.

[26] 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, August 2011.