NMED-T: A TEMPO-FOCUSED DATASET OF CORTICAL AND BEHAVIORAL RESPONSES TO NATURALISTIC MUSIC

Steven Losorelli[1,2] Duc T. Nguyen[1,2] Jacek P. Dmochowski[3] Blair Kaneshiro[1,2]
[1]Center for the Study of Language and Information, Stanford University, USA
[2]Center for Computer Research in Music and Acoustics, Stanford University, USA
[3]Department of Biomedical Engineering, City College of New York, USA
{slosorelli, dtn006, blairbo}@stanford.edu jdmochowski@ccny.cuny.edu

ABSTRACT

Understanding human perception of music is foundational to many research topics in Music Information Retrieval (MIR). While the field of MIR has shown a rising interest in the study of brain responses, access to data remains an obstacle. Here we introduce the Naturalistic Music EEG Dataset—Tempo (NMED-T), an open dataset of electrophysiological and behavioral responses collected from 20 participants who heard a set of 10 commercially available musical works. Song stimuli span various genres and tempos, and all contain electronically produced beats in duple meter. Preprocessed and aggregated responses include dense-array EEG and sensorimotor synchronization (tapping) responses, behavioral ratings of the songs, and basic demographic information. These data, along with illustrative analysis code, are published in Matlab format. Raw EEG and tapping data are also made available. In this paper we describe the construction of the dataset, present results from illustrative analyses, and document the format and attributes of the published data. This dataset facilitates reproducible research in neuroscience and cognitive MIR, and points to several possible avenues for future studies on human processing of naturalistic music.

1. INTRODUCTION

Humans possess a unique ability to process music, and many topics in Music Information Retrieval (MIR) involve computational modeling of human perception. Tasks that humans often perform with ease—such as melody extraction, beat detection, and artist identification—remain open topics in MIR. At the same time, a full understanding of the cognitive and perceptual processes underlying human processing of music has yet to be reached.

Greater cross-disciplinary collaboration between MIR and neuroscience has been proposed [14], and a number of studies have incorporated approaches from both fields. For example, neural correlates of short- and long-term features introduced in MIR for genre classification [34] have been sought [1, 6, 10, 20], and brain responses have been used in MIR-related applications including tempo estimation [29, 30] and emotion recognition [5, 21]. Yet even as brain data become more prevalent in MIR research, experimental design, data collection, and data cleaning can present challenges [14]. Therefore, the research community can arguably benefit from curated, ready-to-use datasets of brain responses to real-world musical works.

Aiming to provide an open dataset with which several MIR and neuroscience topics can be explored, we introduce the Naturalistic Music EEG Dataset—Tempo (NMED-T), a dataset of EEG and behavioral responses to commercially available musical works. The dataset contains dense-array EEG responses from 20 adult participants who listened to 10 full-length songs, as well as tapped responses to the beat of shorter excerpts (collected in a separate listen). These responses have been cleaned and aggregated, and are ready to use in Matlab format along with ratings of familiarity and enjoyment, as well as basic demographic information about the participants.

NMED-T contributes to a growing body of publicly available music-related EEG repositories, including the DEAP [17], Music BCI [32], NMED-H [15], and OpenMIIR [31] datasets. It is well suited for MIR research in that the data are cleaned and ready to use but are also made available in raw form; stimuli are complete, naturalistic [1] musical works spanning a wide range of tempos; metadata links to stimulus audio are provided; and behavioral data are included. Moreover, as EEG was recorded while participants listened attentively but did not focus on any particular dimension of the songs, these data are suitable for studying many aspects of music processing.

The remainder of the paper is structured as follows. In § 2 we describe stimulus selection, study design, data collection, and data preprocessing. Illustrative analyses of the preprocessed data, which build upon past music perception and MIR approaches and reveal cortical and behavioral evidence of entrainment to musical beat, are presented in § 3. In § 4 we document the dataset itself. We conclude and discuss potential future uses of the data in § 5.

[1] Denoting real-world music—i.e., music that was created to be consumed in everyday life, as opposed to controlled stimuli created for experimental research.

2. METHODS

2.1 Stimuli

2.1.1 Stimulus Selection

As the present dataset is focused on naturalistic music and tempo, stimuli were ecologically valid, real-world musical works containing steady, electronically produced beats in duple meter at a variety of tempos. The 10 selected songs are all 4:30–5:00 in length, contain vocals (all but one in English), and are in the Western musical tradition. Song information is summarized in Table 1.

To aid in song selection, we computed objective measures of tempo using publicly available Matlab code [8]. The computed tempos were then validated perceptually by four trained musicians. The final set of selected songs range in tempo from 56–150 BPM—a wide enough range to potentially explore octave errors [11, 35]. To facilitate further research on the audio as well as the responses, we purchased digital versions of all songs from Amazon, and include in Table 1 each song's Amazon Standard Identification Number (ASIN).

These real-world stimuli are complex and contain energy at various frequencies—not just those directly related to the beat. We followed the approach of Nozaradan et al. [27] and visualized low-frequency spectra of the stimuli. We extracted the amplitude envelope of each song using the MIR Toolbox, version 1.5 [18] at a sampling rate of 125 Hz (the sampling rate of the preprocessed EEG), and plotted magnitude spectra up to 15 Hz. As can be seen in Fig. 1, spectral peaks often occur at harmonics and subharmonics of the beat—implicating the hierarchical timescale of music—as well as at other frequencies.

2.1.2 Stimulus Preparation

To prepare the stimuli for the EEG experiment, full-length songs were first converted to mono using Audacity, version 2.1.2. [2] We then embedded the second audio channel with an intermittent click that was transmitted directly to the EEG amplifier (not played to participants) to ensure precise time stamping of the stimuli. For the behavioral experiment, we created 35-second excerpts of each song. Using Audacity, we selected the audio from 1:00–1:34 and applied a linear fade-in and fade-out to the first and last 2 seconds, respectively. We then appended 1 second of silence to make the conclusion of each excerpt more obvious to the participant.

2.2 Participants

Twenty right-handed participants, aged 18–29 years (mean age 23 years, 6 female) participated in the experiment. All reported normal hearing, fluency in English, and no cognitive or decisional impairments. We imposed no eligibility criteria related to formal musical training; 17 participants reported having received training (mean 8.4 years among those with training). Participants reported listening to music for 14.5 hours per week on average.

[2] `http://www.audacityteam.org`

2.3 Experimental Specifications & Data Collection

This study was approved by the Stanford University Institutional Review Board. All participants provided written informed consent before participating. Each participant filled out a general demographic and musical background questionnaire, after which the EEG and tapping blocks were completed, with the EEG block always occurring first.

2.3.1 EEG Experiment

First, each participant was informed that the general purpose of the experiment was to study human processing of music, and that he or she would be completing an EEG session and a behavioral test. As the EEG data were collected for the general study of music processing (not limited to beat perception), no explicit mention of beat or tempo was given at this stage of the experiment. Rather, participants were instructed simply to listen attentively to the songs as they played, and to avoid movement of any kind (including stretching, yawning, and tapping or moving to the beat) during the trials. Songs were presented in random order. Following each trial, participants delivered ratings of familiarity and enjoyment for the song just presented, on a scale of 1–9. The EEG experiment was split into two consecutive recording blocks in order to mitigate participant fatigue, limit data size of the EEG recordings, and allow for verification of electrode impedances between recordings. Therefore, a total of 40 EEG recordings were collected across the 20 participants.

The EEG experiment was programmed in Matlab version 2013b [3] with a custom template built on the Psychophysics Toolbox, version 3 [4]. Each participant sat comfortably in a chair at a desk for the duration of the experiment. Stimuli were presented through magnetically shielded Genelec 1030A speakers at a measured loudness level between 73–78 dB. During the trials, the participant viewed a fixation image presented on a computer monitor located 57 cm in front of him or her.

Dense-array EEG was recorded using the Electrical Geodesics, Inc. (EGI) GES300 system [33]. Data were recorded from 128 electrodes with vertex reference using an EGI Net Amps 300 amplifier and Net Station 4.5.7 acquisition software, sampled at 1 kHz with a range of 24 bits. Electrode impedances were verified to be no greater than 50 kΩ—an appropriate level for this system—at the start of each recording.

2.3.2 Behavioral Experiment

Following the EEG recordings, the electrode net was removed from the participant, and the behavioral test began. Here, each participant listened to the 35-second song excerpts, after receiving instructions to "tap to the steady beat of the song as you perceive it." If the participant had questions about tapping to multiple tempos for a given song, he or she was instructed to tap to the steady beat that best reflected his or her perception of it in the moment. Excerpts were presented in random order.

[3] `https://www.mathworks.com`

#	Song Title	Artist	ASIN	Tempo (BPM)	Tempo (Hz)	min:sec
1	"First Fires"	Bonobo	B00CJE73J6	55.97	0.9328	4:38
2	"Oino"	LA Priest	B00T4NHS2W	69.44	1.1574	4:31
3	"Tiptoes"	Daedelus	B011SAZRLC	74.26	1.2376	4:36
4	"Careless Love"	Croquet Club	B06X9736NJ	82.42	1.3736	4:54
5	"Lebanese Blonde"	Thievery Corporation	B000SF16MI	91.46	1.5244	4:49
6	"Canopée"	Polo & Pan	B01GOL4IB0	96.15	1.6026	4:36
7	"Doing Yoga"	Kazy Lambist	B01JDDVIQ4	108.70	1.8116	4:52
8	"Until the Sun Needs to Rise"	Rüfüs du Sol	B01APT6JKA	120.00	2.0000	4:52
9	"Silent Shout"	The Knife	B00IMN40O4	128.21	2.1368	4:54
10	"The Last Thing You Should Do"	David Bowie	B018GS2A46	150.00	2.5000	4:58

Table 1. Stimulus set. Songs were selected on the basis of vocals, electronically produced beats, genre, tempo, and length.

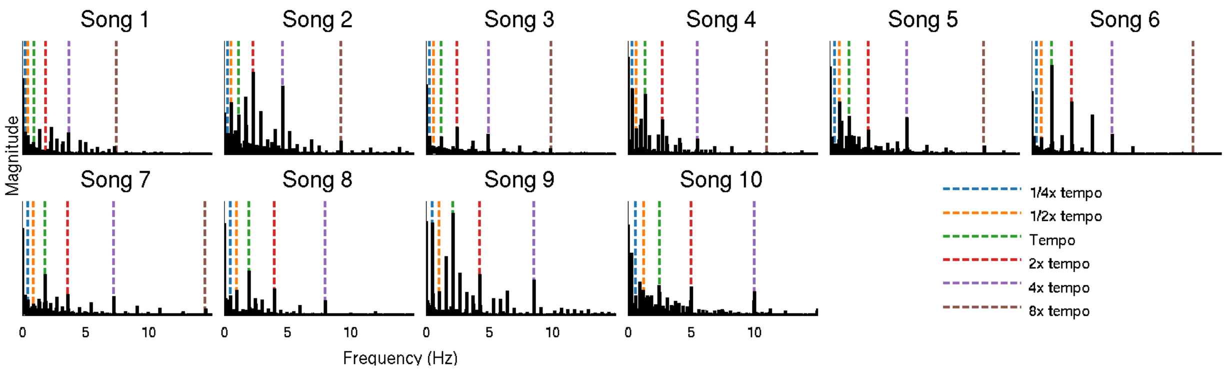

Figure 1. Low-frequency magnitude spectra of stimulus amplitude envelopes. Frequencies related to the musical beat hierarchy, from $1/4x$ the tempo (whole notes) to $8x$ the tempo (32nd notes) are denoted with vertical dashed lines.

Tapping responses were collected using Tap-It, an iOS application that plays audio while simultaneously recording responses tapped on the touchscreen [16]. We note a tap-to-timestamp latency of approximately 15 msec (st. dev. 5 msec) [16]. An Apple iPad 2 was used for this experiment, with stimuli delivered at a comfortable listening level using over-ear Sony MDR-V6 headphones.

2.4 Data Preprocessing

All data preprocessing and analysis was conducted using Matlab, versions 2013b and 2016b.

2.4.1 EEG Preprocessing

The following preprocessing steps were performed on individual EEG recordings that had been exported from Net Station to Matlab cell arrays. First, data from each electrode in the electrodes-by-time data matrix were zero-phase filtered using 8th-order Butterworth highpass (0.3 Hz) and notch (59–61 Hz) filters, and a 16th-order Chebyshev Type I lowpass (50 Hz) filter. Following this, the filtered data were temporally downsampled by a factor of 8 to a final sampling rate of 125 Hz.

We extracted trial labels, onsets, and behavioral ratings, and corrected the stimulus onset times using the click events sent directly from the audio to the EEG amplifier. The data for each trial were epoched, concatenated, and DC corrected (subtracting from each electrode its median value). Bad electrodes were removed from the data matrix, resulting in a reduction in the number of rows. We computed EOG components for tracking vertical and horizontal eye movements, and retained electrodes 1–124 for further analysis, excluding electrodes on the face. We applied a validated approach using Independent Components Analysis (ICA) to remove ocular and cardiac artifacts from the data [2, 13] using the `runica` function from the EEGLAB toolbox [7].

As final preprocessing steps, transients exceeding 4 standard deviations of each electrode's mean power were identified in an iterative fashion and replaced with NaNs. We then reconstituted missing rows corresponding to previously identified bad electrodes with rows of NaNs, ensuring that each data matrix contained the same number of rows. We appended a row of zeros—representing the vertex reference—and converted the data frame to average reference (subtracting from each electrode the mean of all electrodes). All missing values (NaNs) were imputed with the spatial average of data from neighboring electrodes, and a final DC correction was performed. Finally, the epochs were separated once again into single trials. Therefore, after preprocessing, each recording produced a cell array of EEG data, each element of which contained an electrodes-by-time matrix of size $125 \times T$, where T varied according to the length of the stimulus.

After preprocessing all recordings, we aggregated the data on a per-song basis. The data frame for each song is thus a 3D electrodes-by-time-by-participant matrix of size $125 \times T \times 20$.

2.4.2 Preprocessing of Tapping Responses

The Tap-It application stores the timestamps of taps, in seconds, measured from the device touchscreen on a per-trial basis, as well as each participant's randomized stimulus ordering array [16]. We aggregated the tapping responses in a cell array and the ordering arrays in a matrix.

3. ILLUSTRATIVE ANALYSES

The following analyses are presented to illustrate basic properties of the dataset.

3.1 EEG Responses

One approach to studying beat processing using EEG involves low-frequency ($\leq$ 20 Hz) steady-state evoked potentials (SS-EPs). In an SS-EP paradigm, stimuli presented (e.g., flashed or sounded) at a particular frequency elicit brain responses at that same frequency. While SS-EPs are more often used to study vision processing [25], the approach has in recent years been used to study responses to auditory rhythms. Here, SS-EPs have shown evidence of entrainment to musical beat, peaking at beat- and meter-related frequencies even when metrical accents are imagined [26] or when beat frequencies do not dominate low-frequency stimulus spectra [27]. To our knowledge, music SS-EP studies have to date used simple, synthesized rhythmic patterns as stimuli. Our first illustrative analysis extends this approach to complex, naturalistic music.

Spatial filtering is a technique for EEG analysis whereby a weighted sum of electrodes is computed subject to some criterion [3]. Advantages of concentrating activity of interest from many electrodes to a few spatially filtered components include dimensionality reduction, improved SNR, and a reduction in multiple comparisons. For the present analysis we consider two simple spatial filters. The first is simply the mean across all electrodes (ME), which can be thought of as a constant weight applied to each electrode. For the second, we perform Principal Components Analysis (PCA), and analyze the first PC of data.

We first averaged each song's 3D electrodes-by-time-by-participant matrix across participants, producing an electrodes-by-time matrix for each song. Then, so that we analyzed the same amount of data for each song and to account for the time course of listener entrainment to the beat [9], we retained 4 minutes of data from each song, starting 15 seconds into the song.

To compute the spatial filters, we concatenated the participant-averaged data frames across all songs, producing an electrodes-by-aggregated-time matrix. Then, for the ME spatial filter, we computed the mean across electrodes, while for the PCA filter we computed electrode weightings for PC1 using Singular Value Decomposition (SVD). Finally, we reshaped each resulting song-concatenated component vector into a songs-by-time matrix. As our current interest is on SS-EPs, we present the magnitude spectrum of each component on a per-song basis.

The SS-EPs are shown in Fig. 2; y-axis scaling is consistent within each spatial filtering technique. By inspection of the plots, low frequencies (<15 Hz) of ME spectra occasionally contain peaks at frequencies in the musical beat hierarchy (e.g., Song 5). PC1 performs better, eliciting more robust spectral peaks at beat-related frequencies. Moreover, EEG PC1 appears to peak at frequencies directly related to musical beat, while suppressing many of the other spectral peaks that were observed in the magnitude spectra of stimulus amplitude envelopes (Fig. 1).

Spatial filters can be visualized by projecting the filter weights on a 2D scalp topography. While it is common to convert the spatial filter weights to a so-called "forward model," which captures the projection of filtered activity on the scalp, for PCA the spatial filter is equivalent to the forward model [28]. The ME filter, applying a constant weight to all electrodes, would reveal no spatial variation. However, the PC1 filter topography (Fig. 2, bottom right) applies a range of positive and negative weights to the electrodes, which may help to explain why this filter produces more prominent spectral peaks at beat frequencies.

3.2 Behavioral Ratings

Participant ratings of familiarity and enjoyment are shown in Fig. 3. Familiarity with the songs was low overall; ratings of enjoyment tended to be higher, and also varied more across participants.

3.3 Tapped Responses

For each trial of tapping data, we first converted each inter-tap interval to an instantaneous measure of tempo in Hz, mapped it to the midpoint of the interval, and then linearly interpolated the result to a consistent timing grid with a temporal resolution of 2 Hz. We analyze and plot data from a 17-second interval starting 15 seconds into the excerpt (i.e., starting at time 1:15 in the complete song).

The aggregate tapping responses are shown in Fig. 4. We present two visualizations of these results. First, the top figure for each song shows instantaneous tempo over the time of the excerpt for individual participants (gray curves), with the median across participants plotted in black. In bottom figures, we compute the median tempo across time for each individual participant, and summarize with histograms. Beat-related frequencies are shown in the orange (1/2x tempo frequency), green (tempo frequency), and red (2x tempo frequency) lines. To a large extent, participants tended to tap at what we had previously determined to be the tempo frequency. However, there are cases of lower agreement, particularly for the slowest songs (Song 1 and Song 2). Here, the histograms suggest a nearly bimodal distribution of tapped tempos, split between the computational measure and twice that, with the higher measure lying closer to what is considered the preferred tempo region for humans [23].

4. PUBLISHED DATASET

We publish the cleaned EEG data, aggregated behavioral ratings, aggregated tapped responses, and basic demographic data about the participants in Matlab `.mat` format. Example code and helper functions for the illustrative analyses are provided, also in Matlab format. Finally, we publish raw EEG recordings (for researchers who wish to apply their own preprocessing pipelines) as well as individual files of the tapped responses. The dataset is available for download from the Stanford Digital Repository [22], [4] published under a Creative Commons CC-BY license.

[4] `https://purl.stanford.edu/jn859kj8079`

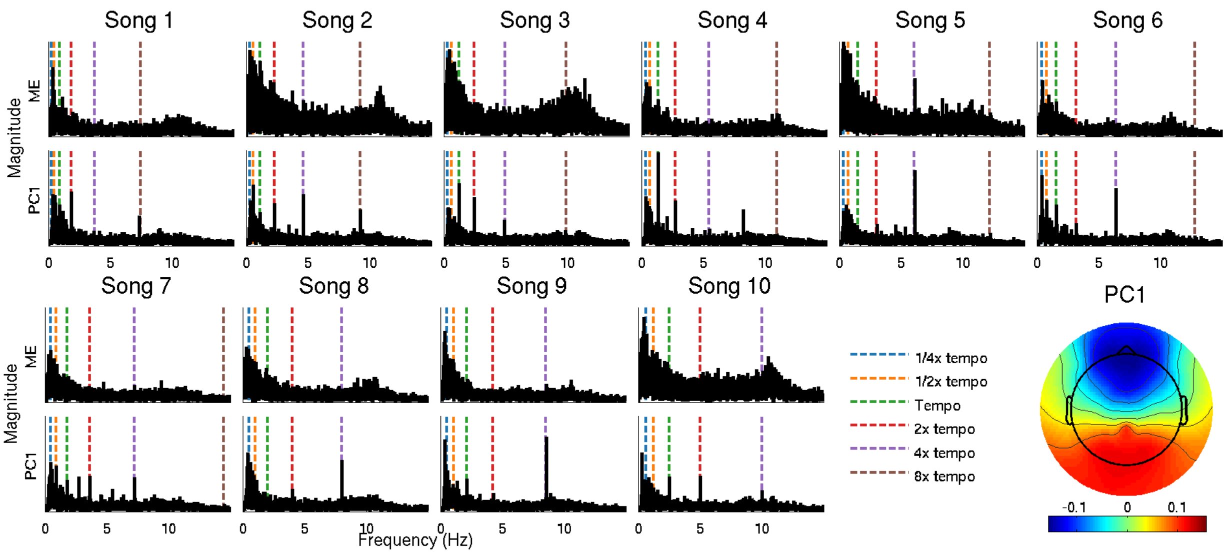

Figure 2. Low-frequency EEG spectra using a mean-electrode spatial filter (top) and PC1 spatial filter (bottom) for each song. Beat-related frequencies are shown with dashed vertical lines. Bottom right: PC1 spatial filter weights.

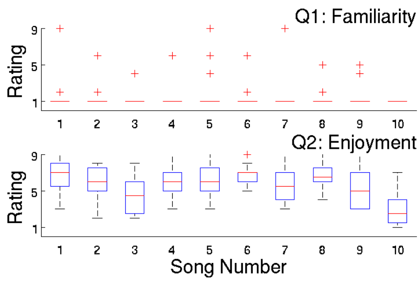

Figure 3. Participant ratings of familiarity and enjoyment.

4.1 Cleaned EEG Data

The `.mat` file `songSS_Imputed.mat` contains the cleaned EEG records, aggregated across participants, for song `SS` (§ 2.4.1). There are 10 such files, one per song. Each `.mat` file contains the following variables:

- *dataSS*: 3D electrodes-by-time-by-participant data frame. The size is $125 \times T \times 20$, with T varying according to the song.
- *subsSS*: Cell array of participant ids. Contents are the same for all songs, but are included in order to link these data to raw EEG files, raw tapping responses, and participant demographics.
- *fs*: Sampling rate, in Hz (always 125).

4.2 Raw EEG Data

We provide the raw EEG records in their exported state before preprocessing. No filtering, epoching, or cleaning has been performed. As each participant underwent two recordings, there are a total of 40 raw EEG files. The file `PP_R_raw.mat` refers to recording `R`$\in$ 1, 2 from participant `PP`. Each file contains the following variables:

- *X*: Raw data frame. Size is electrodes-by-time, $129 \times T$, where T is the total length of the recording, including time periods not related to the experimental trials. The vertex reference electrode is row 129.
- *DIN_1*: Cell array containing all event labels (triggers) and times. We provide the helper function `parseDIN.m` to extract the labels and onsets into numeric vectors. Full specification on labels is provided in the `README` file accompanying the dataset.
- *fs*: Sampling rate, in Hz (always 1000).

4.3 Behavioral Ratings

Participants delivered ratings of familiarity (Q1) and enjoyment (Q2) of each song during the EEG session. The file `behavioralRatings.mat` contains a single variable *behavioralRatings*, which is a 3D participant-by-song-by-question ($20 \times 10 \times 2$) matrix.

4.4 Tapping Responses

Aggregated and raw tapping responses are stored in the file `TapIt.zip`. This archive contains the file `TapIt.mat`, which comprises the following variables:

- *allTappedResponses*: Aggregated tapped response times across all participants and songs. This is a participants-by-song (20×10) cell array. Each entry is a column vector of tap times in seconds, recorded from the device touchscreen.
- *allSongOrders*: Song-order vectors, aggregated across all participants. This is a participants-by-trial (20×10) matrix, where each row contains the stimulus presentation order for the respective participant. Numbering starts at 1.

Individual response files are also included in the `.zip` file:

- *PPP_SS.txt*: Single trial of tapped responses, in seconds, for participant `PPP` and song `SS`.
- *PPP_play_order.txt*: Stimulus presentation ordering for participant `PPP`. Numbering starts at 0.

4.5 Participant Demographics

The file `participantInfo.mat` contains a struct array *participantInfo* with participant demographics. Fields

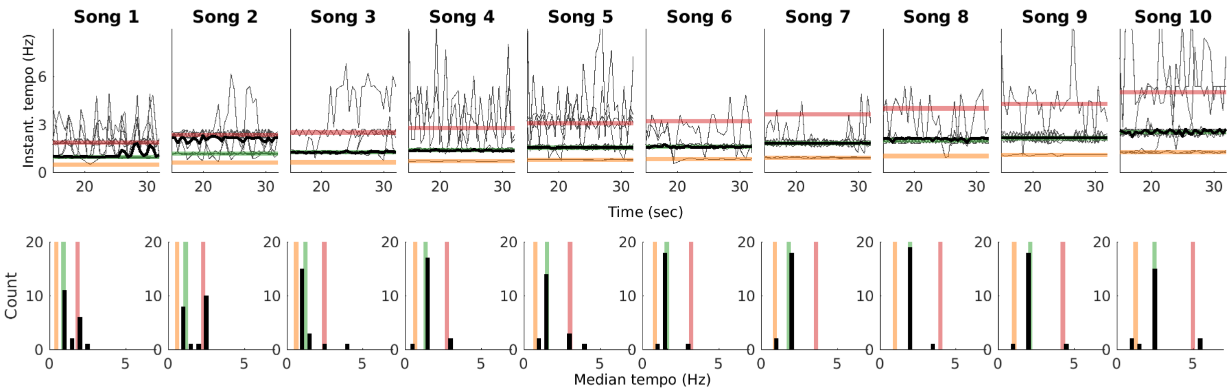

Figure 4. Tapping responses. Top: Instantaneous tempo over time for individual participants (gray), with median across participants in black. Bottom: Histograms of median tempo, over time, for individual participants. Ground-truth tempos are shown with orange ($1/2x$ tempo frequency), green (tempo frequency), and red ($2x$ tempo frequency) lines.

include *age*, *nYearsTraining*, *weeklyListening* (hours), and *id* (participant identifier link to raw filenames).

4.6 Code

The file `Code.zip` contains the Matlab scripts for the analyses performed in § 3. A variety of helper functions and files (e.g., electrode location map, script to parse the *DIN_1* variable in raw EEG files) are also provided here.

5. DISCUSSION

This paper introduces NMED-T, an open dataset of electrophysiological and behavioral responses collected from 20 participants listening to real-world musical excerpts. The published data include both raw and preprocessed dense-array EEG and tapping responses, behavioral ratings of the songs, and basic demographic information.

Our illustrative analyses validate the frequency-tagging, SS-EP approach [26, 27] with responses to complex, naturalistic music (Fig. 2). Even a simple PCA filter computed from trial-averaged responses highlights beat-related frequencies in the EEG spectra. Many PC1 spectra show prominent peaks between 5–10 Hz, regardless of tempo; future research could use this dataset to investigate further the stimulus and response attributes contributing to this phenomenon. The variability in tapping responses (Fig. 4) highlights the challenge of defining a 'ground truth' for tempo and beat identification, particularly for complex music [24]. Here we see various, sometimes conflicting results across and within participants' tapped responses. Past research has suggested that humans inherently prefer certain frequencies related to natural movement [23, 35]. This may help to explain why some participants tapped at twice the tempo for the slowest songs, tending toward the postulated 2-Hz natural resonant frequency.

We faced several trade-offs when designing the study. Collection of EEG data, while relatively inexpensive [14], still incurs costs of equipment and time. Participant fatigue must also be taken into account when planning the overall duration of an experiment. As we wished to collect EEG responses to a set of full-length songs from every participant, we were limited in the number of songs we could use, and relegated the secondary tapping task to shorter excerpts. Stimulus selection, too, is often a compromise of breadth and depth. For example, the OpenMIIR dataset [31] uses shorter stimuli from a variety of genres, but at the expense of depth within any one genre; while the NMED-H [15] includes various stimulus manipulations of complete songs, but only four songs from a single genre. Our focus on full-length songs with a steady beat and a variety of tempos limited the range of genres somewhat. We also deliberately avoided massively popular songs in order to minimize possible effects, on the brain responses, of varying familiarity, established personal preferences, and autobiographical associations with the songs [12].

There are shortfalls to the dataset. One potential confound is that the EEG session always preceded the behavioral task; thus, participants were more familiar with the music during the tapping task. As a result, the tapping data may not be suitable for studying the time course of beat entrainment. However, we chose this arrangement so that participants would not be focused specifically on beat while EEG responses were recorded. Second, the tapping data show variations in tapped tempo across participants and within-participant over time. Whether this reflects our participant pool (not all trained musicians), inadequate instruction for the task, or is merely characteristic of this response is not addressed in the present illustrative analyses. Finally, listeners are known to exhibit variations in tempo octave during tapping while largely agreeing on whether a song is fast or slow [19], but we unfortunately did not collect data here to explore this distinction.

Generally speaking, this dataset facilitates research on encoding and decoding of naturalistic music. While the study design and initial analyses focused primarily on beat and tempo, the EEG responses can be analyzed in conjunction with various other stimulus features as well. Investigation of individual differences is also possible (e.g., predicting a particular participant's tapping tempo or preference rating from his or her own EEG). Other researchers might consider augmenting the dataset with complementary responses to the same songs. Ideally, the dataset will find applications in MIR and neuroscience research beyond those envisioned by the authors of this study.

6. ACKNOWLEDGMENTS

This research was supported by the Patrick Suppes Gift Fund and the Wallenberg Network Initiative: Culture, Brain, Learning. The authors thank Anthony Norcia, Dan Ellis, and the anonymous ISMIR reviewers for helpful feedback on the manuscript.

7. REFERENCES

[1] V. Alluri, P. Toiviainen, I. P. Jääskeläinen, E. Glerean, M. Sams, and E. Brattico. Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *NeuroImage*, 59(4):3677–3689, 2012.

[2] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

[3] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Muller. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Processing Magazine*, 25(1):41–56, 2008.

[4] D. H. Brainard. The psychophysics toolbox. *Spatial Vision*, 10(4):433–436, 1997.

[5] R. Cabredo, R. S. Legaspi, P. S. Inventado, and M. Numao. An emotion model for music using brain waves. In *ISMIR*, pages 265–270, 2012.

[6] F. Cong, V. Alluri, A. K. Nandi, P. Toiviainen, R. Fa, B. Abu-Jamous, L. Gong, B. G. W. Craenen, H. Poikonen, M. Huotilainen, and T. Ristaniemi. Linking brain responses to naturalistic music through analysis of ongoing EEG and stimulus features. *IEEE Trans. Multimedia*, 15(5):1060–1069, 2013.

[7] A. Delorme and S. Makeig. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, 2004.

[8] D. P. W. Ellis. Beat tracking by dynamic programming. *Journal of New Music Research*, 36(1):51–60, 2007.

[9] P. Fraisse and B. H. Repp. Anticipation of rhythmic stimuli: Speed of establishment and precision of synchronization (1966). *Psychomusicology: Music, Mind, and Brain*, 22(1):84, 2012.

[10] N. Gang, B. Kaneshiro, J. Berger, and J. P. Dmochowski. Decoding neurally relevant musical features using Canonical Correlation Analysis. In *ISMIR*, 2017.

[11] J. Hockman and I. Fujinaga. Fast vs slow: Learning tempo octaves from user data. In *ISMIR*, pages 231–236, 2010.

[12] P. Janata. The neural architecture of music-evoked autobiographical memories. *Cerebral Cortex*, 19(11):2579–2594, 2009.

[13] T.-P. Jung, C. Humphries, T.-W. Lee, S. Makeig, M. J. McKeown, V. Iragui, and T. J. Sejnowski. Extended ICA removes artifacts from electroencephalographic recordings. In *NIPS*, pages 894–900, 1998.

[14] B. Kaneshiro and J. P. Dmochowski. Neuroimaging methods for music information retrieval: Current findings and future prospects. In *ISMIR*, pages 538–544, 2015.

[15] B. Kaneshiro, D. T. Nguyen, J. P. Dmochowski, A. M. Norcia, and J. Berger. Naturalistic music EEG dataset—Hindi (NMED-H). In *Stanford Digital Repository*, 2016.

[16] H.-S. Kim, B. Kaneshiro, and J. Berger. Tap-It: An iOS app for sensori-motor synchronization experiments. In *ICMPC12*, 2012.

[17] 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 Trans. Affective Computing*, 3(1):18–31, 2012.

[18] O. Lartillot and P. Toiviainen. A Matlab toolbox for musical feature extraction from audio. In *DAFx*, pages 237–244, 2007.

[19] M. Levy. Improving perceptual tempo estimation with crowd-sourced annotations. In *ISMIR*, pages 317–322, 2011.

[20] Y. P. Lin, J. R. Duann, W. Feng, J. H. Chen, and T. P. Jung. Revealing spatio-spectral electroencephalographic dynamics of musical mode and tempo perception by independent component analysis. *Journal of NeuroEngineering and Rehabilitation*, 11(1), 2014.

[21] Y. P. Lin, C. H. Wang, T. P. Jung, T. L. Wu, S. K. Jeng, J. R. Duann, and J. H. Chen. EEG-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering*, 57(7):1798–1806, 2010.

[22] S. Losorelli, D. T. Nguyen, J. P. Dmochowski, and B. Kaneshiro. Naturalistic music EEG dataset—Tempo (NMED-T). In *Stanford Digital Repository*, 2017.

[23] D. Moelants. Preferred tempo reconsidered. In *ICMPC7*, pages 1–4, 2002.

[24] D. Moelants and M. F. McKinney. Tempo perception and musical content: What makes a piece fast, slow, or temporally ambiguous? In *ICMPC8*, pages 558–562, 2004.

[25] A. M. Norcia, L. G. Appelbaum, J. M. Ales, B. R. Cottereau, and B. Rossion. The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, 15(6):4, 2015.

[26] 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.

[27] 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.

[28] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda. Recipes for the linear analysis of EEG. *NeuroImage*, 28(2):326–341, 2005.

[29] A. Sternin, S. Stober, J. A. Grahn, and A. M. Owen. Tempo estimation from the EEG signal during perception and imagination of music. In *BCMI/CMMR*, 2015.

[30] S. Stober, T. Prätzlich, and M. Meinard. Brain beats: Tempo extraction from EEG data. In *ISMIR*, pages 276–282, 2016.

[31] 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 and imagination. In *ISMIR*, 2015.

[32] 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, 2014.

[33] D. M. Tucker. Spatial sampling of head electrical fields: The geodesic sensor net. *Electroencephalography and Clinical Neurophysiol.*, 87(3):154–163, 1993.

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

[35] L. van Noorden and D. Moelants. Resonance in the perception of musical pulse. *Journal of New Music Research*, 28(1):43–66, 1999.