2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
October 6-10, 2024. Sarawak, Malaysia

DECODING MUSICAL TIMBRE PERCEPTION FROM SINGLE-TRIAL EEG DATA

Praveena Satkunarajah, Sarah D. Power, Benjamin Rich Zendel

Memorial University of Newfoundland
Faculty of Medicine
300 Prince Philip Drive, St. John's, NL A1B 3V6, Canada

2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC) | 978-1-6654-1020-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/SMC54092.2024.10831757

ABSTRACT

Many users of hearing aids report challenges when listening to music. In the future, it may be possible to develop hearing aids that have electrodes which monitors brain activity in real-time and adapts the filters on the hearing aid to match the volitions of the user. In music, this could mean amplifying the sound of the instrument the listener wants to hear. One of the first steps in this research is to determine if a machine learning algorithm can identify to which instrument an individual is listening based only on a brief EEG signal. To test this possibility, participants were presented with a series of brief tones that varied in timbre (Trombone, Clarinet, Cello, Piano and Pure Tone) while their ongoing EEG was recorded from 73 electrodes. Linear Discriminant Analysis (LDA) was used. We investigated four different sets of features – Raw EEG, ERP-based features, Harmonics-based features and Regularity-based features. The Raw EEG based classifier performed significantly above chance (37%) when attempting to distinguish between responses to different musical instruments for 5-way classification. More advanced classification algorithms or different features may be able to better distinguish between tones with a musical timbre.

Index Terms— timbre, machine learning, EEG, hearing

1. INTRODUCTION

When listening to music, physical cues such as amplitude, frequency or timbre can be used to separate instruments that are playing concurrently. Both amplitude and frequency are unidimensional, can be measured using a single metric (i.e., decibels or Hertz), and have a direct, albeit non-linear relationship with a perceptual correlate (i.e., loudness and pitch). Timbre is multidimensional, and is defined as the quality of sound that differentiates two sounds with the same loudness, pitch location and duration [1]. It is what makes a guitar and piano playing the same note at the same amplitude sound different from each other. At the physical timbre is often associated with the relative amplitude of the harmonics, the amplitude dynamics of each harmonic, or the onset/offset slopes. Perceptually, a technique called multi-dimensional scaling (MDS) [2] has been used to determine how acoustic properties interact to engender the perception of timbre. MDS involves mapping dissimilarity in perceived timbre into a multidimensional space and attributing it to acoustical features of the sound.

Understanding how the brain represents musical timbres is a critical step in understanding how the brain uses timbral cues to perceptually segregate instruments during music listening. This type of research could benefit the development of Brain Computer Interface (BCI)-based hearing aids that could use an individual's ongoing electroencephalogram (EEG) signal to identify which incoming acoustic information could be selectively filtered, compressed, or amplified so that the hearing aid would adapt to both the acoustic environment that they are in, and the volitions of the listener.

While there have been previous studies that have focused on auditory attention to instruments, these studies do not fully explore how different characteristics of sound are represented in EEG data. Accordingly, the goal of this study is to determine if EEG data recorded while a person listens to different instruments can be classified on a trial-by-trial basis.

Previous studies on timbre discrimination show that perception of different timbres can indeed be encoded in EEG data. Auzou et al. (1995) reported that a timbral discrimination task led to changes in the EEG signal originating from the right hemisphere [3]. Meyer et al. (2006) reported that the amplitude of the N1 and P2 components of the auditory evoked response were larger when evoked by a musical instrument compared to a sine wave tone. However, both these studies made use of analysis based on event related potentials (ERPs) averaged over multiple trials. ERP based analyses normally require averaging multiple trials together in order to observe reliable and measureable peaks in the waveform. Accordingly, at the single trial level, it may be possible to measure peaks retrospectively. That is, a grand average of multiple trials can be used to identify when and where a peak occurs in the ERP waveform. Using this information, the amplitude of the waveform can be extracted at specific electrodes during a specific epoch that lines up with a grand average. This may not be practical in a real-world scenario without some prior knowledge of the expected dynamics of the evoked response.

Treder et al., in a multi-streamed oddball experiment, used binary Linear Discriminant Analysis (LDA) classifiers

Authorized licensed use limited to: Memorial University. Downloaded on January 22,2025 at 12:46:04 UTC from IEEE Xplore. Restrictions apply.

to classify deviants into "attended" and "unattended" at a single-trial level. Both a general classifier and instrument-specific classifiers were explored. Results showed that the instrument-specific classifiers performed better than the general classifier, further supporting the idea that perceptions of different timbres were encoded differently in the EEG and therefore could possibly be detected by machine learning algorithms [4].

While stimulus reconstruction and ERP-based approaches have been explored in polyphonic music, our focus is on single-trial classification based on musical timbre. We explored whether perception of different instruments is distinctly encoded in EEG data such that it can be classified on a single-trial basis using machine learning. We also explored the effectiveness of different types of features.

2. MATERIALS AND METHODS

2.1. Participants

Fifteen participants were recruited for the study. However, due to technical difficulties with the EEG system data from only 10 participants (27 ± 11; 7 female, 3 male) were available to be used. Participants recruited were required to be right-handed and have normal hearing. Hearing thresholds were screened for using the Pure Tone Audiometry (PTA) assessment. Of the 10 participants included in the final data set, one was a formally trained musician, four were self-taught, and five were non-musicians. The amount of time participants spent actively listening to music - that is, in a focused, engaged manner - ranged from 0 to 27 hours a week.

2.2. Stimuli

The stimuli consisted of computer-generated tones with an F0 of 220 Hz (REAPER, Cockos, Inc., San Francisco, California, United States) in five different timbres, one in Pure Tone - a sine wave comprised of only a single frequency component - and the other in tones of four different instruments - the Clarinet, Piano, Trombone and Cello. Each Tone was 1s in length, and were normalised to -23 dB LUFS and presented at 70 dB SPL. Each of the 5 instrument tones were presented 200 times in sets of 5. Figure1 displays the sound envelopes of the tones used, and the amplitude spectrum of the tones.

2.3. Procedure

Participants first completed an orally administered questionnaire which included questions about their demographics and experience with music. Once this was done, participants were seated in a sound-attenuating booth and had their audiometric thresholds assessed using Pure Tone Audiometry (PTA). First, to ensure participants had normal hearing, pure-tone thresholds (PTTs) were measured by air conduction using ER-3A insert earphones evaluate hearing status. All participants had a PTT below 25 dB HL [5].

For the experimental task, participants were fitted with Etymotic E3A insert earphones, and were presented with the five instrument tones at 70 dB SPL. Before completing the task, participants were given time to familiarise themselves with the timbral differences of the 5 stimuli. Each participant was presented with a total of a thousand tones (200 per stimuli), which was broken down into 4 blocks of 250. Within each block, each instrument tone was presented in sets of 5. At the end of each set of 5 identical instrument tones, participants were asked to identify what tone they heard by pressing a button on a response box. The names of each of the five instrument tones was presented on a screen while they made their selection. Participants were easily able to classify the instruments, with average accuracy above 95%

2.4. EEG Recording and Preprocessing

EEG was recorded from 70 channels (64 electrodes and 6 face electrodes) at a sampling rate of 2048 Hz.

A highpass filter of 0.1 Hz was applied, and eye movements were removed using Independent Component Analysis (ICA). The processed data was then re-referenced to the mastoid electrodes, after which the data was segmented into epochs by stimulus type into 2s segments, including a 1-second pre-stimulus baseline.

2.5. Feature Extraction

We compared the classification performance on four different feature sets:

1. ERP-based Features: In this feature set, the amplitude, latency and mean voltages of the N1 and P2 components for each trial were used. The latency for each of the components were estimated based on the grand average ERPs across all participants.

2. Signal Regularity: Two measures of regularity were considered: the spectral entropy and the autocorrelation of the Fourier Transform between the instrument tone and the EEG data. Spectral entropy is a measure of the regularity of the power spectrum of a signal [6], and has been widely used in EEG problems [7, 8, 9, 10].

   In order to extract features of spectral entropy, we first used the multitaper method for spectral density estimation [11]. The spectrum was again broken down to delta, theta, alpha, beta and gamma bands. The spectral entropy was then computed within each of the bands for all channels.

   As another measure of regularity, we also extracted the auto-correlation of the Fourier transform for all channels.

Authorized licensed use limited to: Memorial University. Downloaded on January 22,2025 at 12:46:04 UTC from IEEE Xplore. Restrictions apply.

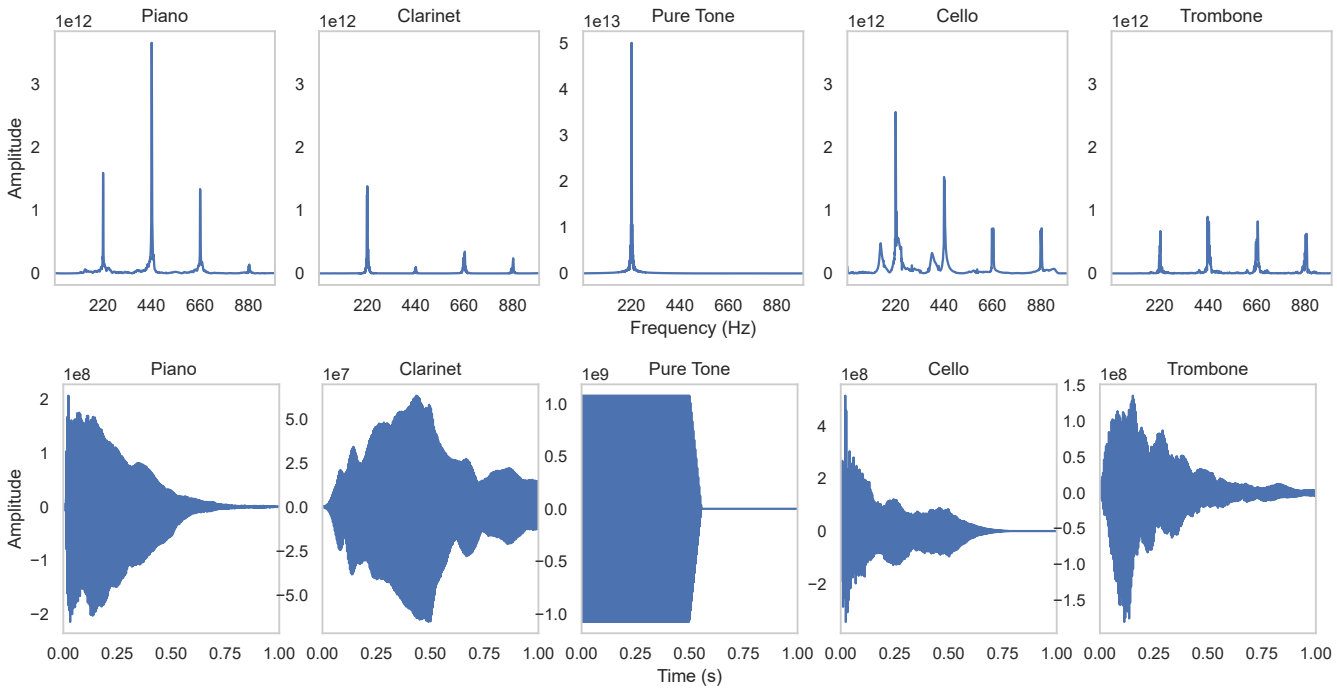

Fig. 1: Figure displaying Amplitude Spectrum (top) of tones and Sound envelope (bottom) of tones

3. Harmonics-based features: The power of the frequency spectrum around the harmonics and the subharmonics of the tone - meaning around frequency points of 27.5Hz, 55Hz, 110Hz, 220Hz, 440Hz and 880Hz – were calculated.

4. Raw EEG. The pre-processed EEG signals without any feature extraction (i.e., the EEG signal value at each time sample, for all 64 channels).

2.6. Feature Reduction and Training

To reduce the number of features used for classification, we performed Principal Component Analysis (PCA). Components accounting for 90% of the data variance were retained and used as features in the classification. Figure 2 displays the number of features which were retained after feature reduction. To enable real-time online processing and classification in the long term, we chose Linear Discriminant Analysis (LDA) for its low computational cost. Grid Search cross validation was used to find the best solver for the model. First, within-participant classification was performed. The model was trained using 10 runs of 5-fold cross validation. Within-participant classification was done using all the feature sets from Section E. Next, across-subject classification was performed. That is, Leave-One-Subject-Out Cross Validation (LOSOCV) was performed, where in each iteration a model was trained on 9 participants, and used to test one unseen participant. This model was trained using Raw EEG as the input only.

We measured the performance of the classifiers using accuracy - the proportion of correct predictions out of all predictions made.

3. RESULTS

Figure 3 displays the within-participant classification results for the four feature sets. To compare these classification accuracies against chance, the class labels were randomized and the within-subject classification analysis (10 runs of 5-fold cross-validation) was repeated. The participant accuracies with the true and random labels were compared via a one-sided paired sample t-test. The results indicated that the classifier trained on Raw EEG (M = 0.381, SD = 0.16) and ERP-based (M = 0.32, SD = 0.16) features on average performed significantly above chance (p = .026 and p = .006 respectively), but those trained on harmonics (M = 0.21, SD = 0.02) and regularity-based features (M = 0.21, SD = 0.03) did not (p = .1 and p = .83 respectively).

A further independent samples t-test against chance was performed within each participant for classifiers trained on Raw EEG and ERP-based features. Accuracies achieved for 8 out of 10 participants were significantly above chance ($p <$ .001) for both Raw EEG and ERP-based features.

Classifier accuracy for each individual participant by fea-

Authorized licensed use limited to: Memorial University. Downloaded on January 22,2025 at 12:46:04 UTC from IEEE Xplore. Restrictions apply.

ture is shown in Table 1.

Table 1: Table of LDA classifier Accuracy by Participant

Participant	Raw EEG	ERP-based	Regularity	Harmonics
P2	0.22	0.22	0.27	0.20
P4	0.22	0.20	0.20	0.21
P5	0.19	0.19	0.19	0.19
P6	0.47	0.32	0.22	0.22
P7	0.52	0.31	0.24	0.18
P9	0.26	0.24	0.19	0.19
P10	0.45	0.56	0.21	0.25
P12	0.42	0.28	0.15	0.20
P14	0.70	0.65	0.21	0.20
P15	0.36	0.25	0.20	0.21

The LOSOCV model yielded a mean accuracy of 28%, (SD = 0.048) across the 10 participants, performing significantly above chance of 22.1% (p = .003). Chance level was computed using inverse binomial cumulative distribution function binoinv (1 - α, n, 1/c) × 100/n with parameters α = .001, n = 1000 for 1000 samples, and c = 0.2 for 5-way classification - a method used by Combrisson and Jerbi [12]. Figure 4 displays the difference in performance between the two models.

4. CONCLUSION

From the results above, we see that classifiers trained on Raw EEG and ERP-based features performed significantly above chance, while those trained on harmonics and regularity based features did not perform above chance levels. Previous work has shown that ERP-based features (i.e., N1 and P2) are modulated by timbre, which may explain why using ERP-based features allowed classifiers to perform significantly above chance [13, 14, 15].

The fact that the models performed better when Raw EEG and ERP-based features were used could be an indication that the classifier relied on temporal features for the classification. In the ERP based analysis, specific epochs were used for the analysis, while using Raw EEG preserved the temporal features of the raw data. Moreover, the finding that Raw EEG allowed for better classification accuracy compared to other features, suggests that feature extraction might not be a necessary step, thus streamlining the classification process. This potentially makes it more ideal for real-time processing.

Furthermore, results suggest that BCI-based hearing aids may need to be customized to each user. Given that the average performance of the LOSOCV (i.e. between-participant) models were lower compared to the within-participant models, it is likely that each individual encodes timbre in a unique way, possibly linked to the high variability observed in the

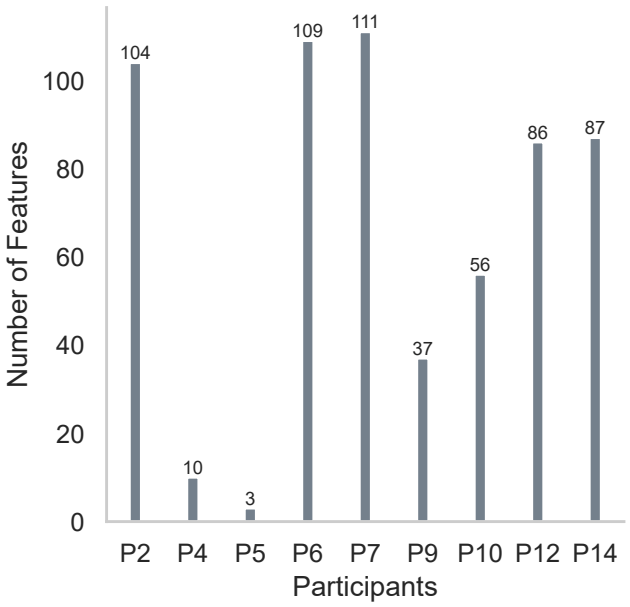

Fig. 2: Average number of Features over 50 folds after Feature Reduction for Raw EEG by Participant

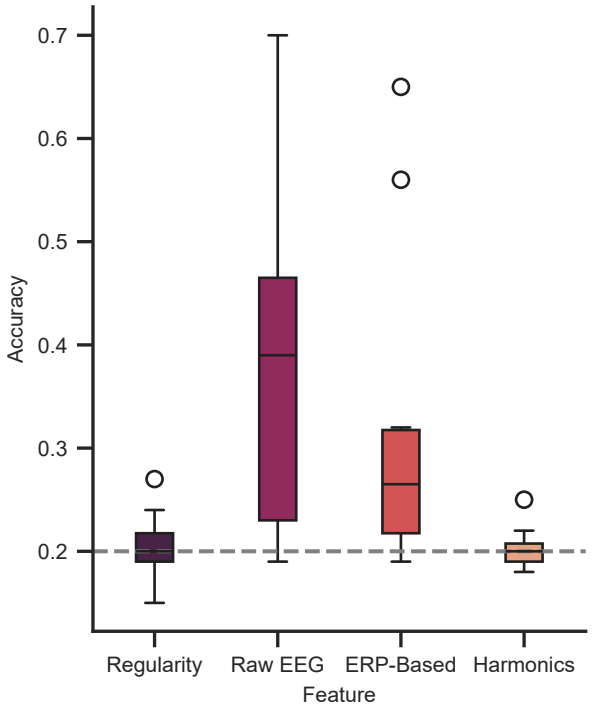

Fig. 3: Classification results for within-participant model for all 5 features

Authorized licensed use limited to: Memorial University. Downloaded on January 22,2025 at 12:46:04 UTC from IEEE Xplore. Restrictions apply.

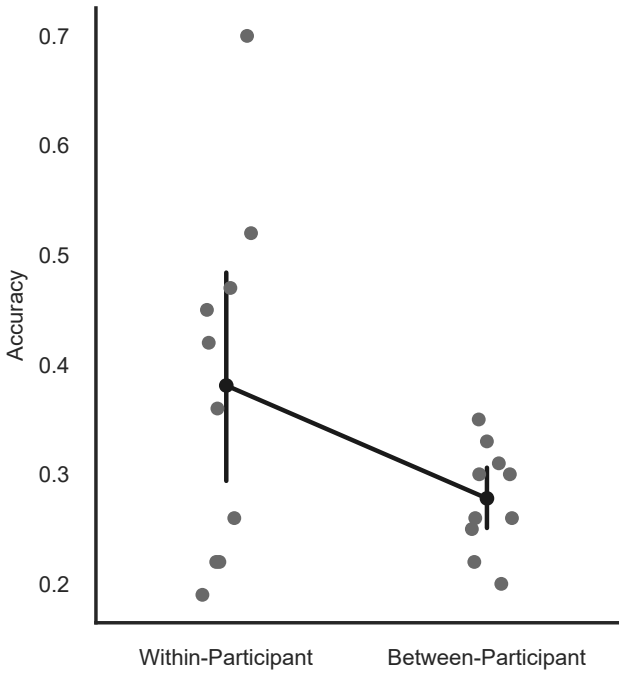

Fig. 4: Comparison between within and between participant model performance

within-participant classifiers.

Future work will involve exploring additional features and classification algorithms that could lead to increased performance.

5. REFERENCES

[1] Acoustics Accredited Standards Committee S1, "American National Standard: Acoustical Terminology. ANSI/ASA S1.1-2013," *Ansi S1.1-1994*, vol. 2013, 2014.

[2] Stephen McAdams, "Perspectives on the contribution of timbre to musical structure," *Computer music journal*, vol. 23, no. 3, pp. 85–102, 1999.

[3] Pascal Auzou, F Eustache, P Etevenon, H Platel, P Rioux, J Lambert, B Lechevalier, E Zarifian, and JC Baron, "Topographic eeg activations during timbre and pitch discrimination tasks using musical sounds," *Neuropsychologia*, vol. 33, no. 1, pp. 25–37, 1995.

[4] Matthias S Treder, Hendrik Purwins, Daniel Miklody, Irene Sturm, and Benjamin Blankertz, "Decoding auditory attention to instruments in polyphonic music using single-trial eeg classification," *Journal of neural engineering*, vol. 11, no. 2, pp. 026009, 2014.

[5] Larry E Humes, "The world health organization's hearing-impairment grading system: an evaluation for unaided communication in age-related hearing loss," *International journal of audiology*, vol. 58, no. 1, pp. 12–20, 2019.

[6] Jürgen Fell, Joachim Röschke, Klaus Mann, and Cornelius Schäffner, "Discrimination of sleep stages: a comparison between spectral and nonlinear eeg measures," *Electroencephalography and clinical Neurophysiology*, vol. 98, no. 5, pp. 401–410, 1996.

[7] N Sriraam et al., "Eeg based detection of alcoholics using spectral entropy with neural network classifiers," in *2012 International Conference on Biomedical Engineering (ICoBE)*. IEEE, 2012, pp. 89–93.

[8] Aihua Zhang, Bin Yang, and Ling Huang, "Feature extraction of eeg signals using power spectral entropy," in *2008 international conference on BioMedical engineering and informatics*. IEEE, 2008, vol. 2, pp. 435–439.

[9] Rui Zhang, Peng Xu, Rui Chen, Fali Li, Lanjin Guo, Peiyang Li, Tao Zhang, and Dezhong Yao, "Predicting inter-session performance of smr-based brain–computer interface using the spectral entropy of resting-state eeg," *Brain topography*, vol. 28, no. 5, pp. 680–690, 2015.

[10] Mitul Kumar Ahirwal and Narendra Londhe, "Power spectrum analysis of eeg signals for estimating visual attention," *International Journal of computer applications*, vol. 42, no. 15, pp. 22–25, 2012.

[11] David J Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.

[12] Etienne Combrisson and Karim Jerbi, "Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy," *Journal of Neuroscience Methods*, vol. 250, pp. 126–136, 2015, Cutting-edge EEG Methods.

[13] Martin Meyer, Simon Baumann, and Lutz Jancke, "Electrical brain imaging reveals spatio-temporal dynamics of timbre perception in humans," *Neuroimage*, vol. 32, no. 4, pp. 1510–1523, 2006.

[14] Przemysław Tużnik, Paweł Augustynowicz, and Piotr Francuz, "Electrophysiological correlates of timbre imagery and perception," *International journal of psychophysiology*, vol. 129, pp. 9–17, 2018.

[15] Antoine Shahin, Larry E Roberts, Christo Pantev, Laurel J Trainor, and Bernhard Ross, "Modulation of p2 auditory-evoked responses by the spectral complexity of musical sounds," *Neuroreport*, vol. 16, no. 16, pp. 1781–1785, 2005.

Authorized licensed use limited to: Memorial University. Downloaded on January 22,2025 at 12:46:04 UTC from IEEE Xplore. Restrictions apply.