Proceedings Paper:

Kumagai, Y., Arvaneh, M., Okawa, H. et al. (2 more authors) (2017) Classification of Familiarity Based on Cross-Correlation Features Between EEG and Music. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). The 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'17), 11-15 July 2017, Seogwipo, South Korea. IEEE , pp. 2879-2882.

https://doi.org/10.1109/EMBC.2017.8037458

Classification of Familiarity Based on Cross-Correlation Features Between EEG and Music

Yuiko Kumagai [1], Mahnaz Arvaneh [2], Haruki Okawa [1], Tomoya Wada [1], and Toshihisa Tanaka [1,3]

***Abstract*— An approach to recognize the familiarity of a listener with music using both the electroencephalogram (EEG) signals and the music signal is proposed in this paper. Eight participants listened to melodies produced by piano sounds as simple natural stimuli. We classified the familiarity of each participant using cross-correlation values between EEG and the envelope of the music signal as features of the support vector machine (SVM) or neural network used. Here, we report that the maximum classification accuracy was 100% obtained by the SVM. These results suggest that the familiarity of music can be classified by cross-correlation values. The proposed approach can be used to recognize high-level brain states such as familiarity, preference, and emotion.**

I. INTRODUCTION

Brain–computer interfaces (BCIs) are emerging interfaces that have a lot of potential as an application of signal processing and machine learning techniques. BCIs are not only used to control a computer or a device but also for music therapy and recommendations [1], [2], [3], [4]. For this purpose, some researchers have investigated neural responses to detect high-level brain states such as emotion and preference. Daly et al. [5] predicted music-induced emotion using electroencephalogram (EEG) as well as the properties of music. Hadjidimitriou et al. [6] classified music preferences under the parameter of familiarity using a time-frequency analysis of EEG. However, it has not investigated to recognize whether music is familiar or unfamiliar to a listener.

To investigate brain responses to sound, many researchers have measured event-related potentials (ERPs), such as mismatch negativity (MMN), in numerous contexts. MMN is a change-specific component of ERPs that has a peak at 150–250 ms after the onset of a deviant stimulus [7]. It has been shown that MMNs are elicited by deviant sounds in rhythmic sequences [8] and melodies [9]. Another approach is to measure an auditory steady-state response (ASSR), which can be elicited by periodically repeated sounds [10]. Recent studies have demonstrated that the cortical entrainment to periodic tones [11] and rhythms of music [12]. However, these MMN and ASSR approaches are not suitable to continuous stimuli such as natural music.

[1]Y. Kumagai, H. Okawa, T. Wada, and T. Tanaka are with Department of Electrical and Electronic Engineering at the Tokyo University of Agriculture and Technology, 2-24-16 Nakacho, Koganei-shi, Tokyo, Japan. [2]M. Arvaneh is with the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK. [3]T. Tanaka is also affiliated with RIKEN Brain Science Institute, Wako-shi, Saitama, Japan (e-mail: tanakat@cc.tuat.ac.jp).

*This work was supported by JSPS Grant 16K12456 and the Daiwa Anglo-Japanese Foundation.

Recent works on auditory perception have focused on cortical entrainment [13], [14], [15], [16], [17], [18], [19], [20], [21]. Cortical entrainment to the envelope of speech has been used in several studies [14], [15], [16]. It has been demonstrated that attended speech can be decoded by low-frequency neural oscillation and speech envelopes [17], [18]. Moreover, cortical entrainment to the envelope of speech can be seen even at the single-trial level [19], [20]. Cortical entrainment to the music has also been demonstrated in terms of music perception [21], [22].

In this paper, we present our findings on estimating the familiarity with music from EEG and music signals with a focus on entrainment. As features, we calculated cross-correlation function between the EEG while a participant is listening to music and the envelope of the music. The features were classified by a support vector machine (SVM) or deep neural network (DNN).

This paper is organized as follows: Section II describes the dataset we used, the preprocessing of the data, and feature extraction. In Section III, we present the results and discussion. In Section IV, conclusions are drawn.

II. MATERIALS AND METHODS

A. Dataset

The EEG data used in this study were collected in our previous work [22]. In order to classify familiarity, we used a subset of the data, that is, the trials labeled familiar and unfamiliar. A brief description of the dataset follows.

1) Participants: Eight males in their 20s participated in this experiment. All participants were healthy and had normal hearing. They each signed an informed consent form. The study was approved by the Human Research Ethics Committee of the Tokyo University of Agriculture and Technology.

2) EEG data acquisition: We used the Ag/AgCl active electrodes from Guger Technologies (g.tec) named g.LADYbird, g.LADYbirdGND (for GND), and g.GAMMAearclip (for reference) for the EEG. These were driven by a power supply unit named g.GAMMAbox (g.tec). Thirty-two electrodes were placed over the scalp in accordance with the international 10–10 system, as shown in Fig. 1. The electrodes for GND and the reference were placed at AFz and A1, respectively. The EEG signals were amplified using MEG-6116 (Nihon Kohden), which applied low and high-pass analog filters for each channel. The cut-off frequencies of the low and the high-pass filters were set to 100 and 0.08 Hz, respectively. The EEG signals were sampled by A/D converter (AIO-163202FX–USB, Contec) with a sampling rate of 1,024 Hz. The signals were

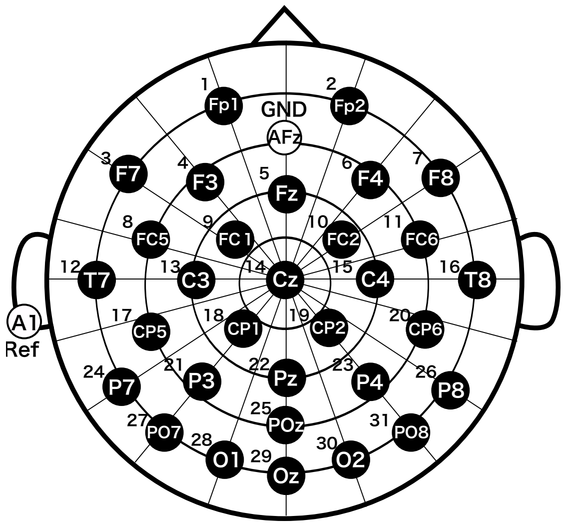

Fig. 1: Electrode positions. The numbers indicate indices.

TABLE I: Pieces of music used in the experiment. For each music in this table, we created a piece consisting of melodies produced by piano sounds.

Composer	Title
Popular English lullaby	Twinkle Twinkle Little Star
A. L. Vivaldi	The Four Seasons, Spring
P. I. Tchaikovsky	The Nutcracker, March
P. I. Tchaikovsky	Swan Lake, Scene
A. Dvorak	Symphony No. 9 "From The New World"
T. Hakase	Jounetsu Tairiku
A. Khachaturyan	Masquerade
J. Pachelbel	Canon
L. v. Beethoven	Ode to Joy
W. A. Mozart	Eine Kleine Nachtmusik
I. Albeniz	Piano Sonate Op.82
F. Kuhla	Sonatine Op.55–1
A. Diabelli	Sonatine Op.151–2
A. Diabelli	Sonatine Op.168–2
P. I. Tchaikovsky	Six Pieces Op.51–1
G. Faure	Dolly Suite, Kitty–Valse
L. v. Beethoven	Piano Sonate Op.14–1
L. v. Beethoven	The Creatures of Prometheus, Introduction
F. Mendelssohn	Lieder Ohne Worte Op.19–1
W. A. Mozart	Piano Sonate KV309

recorded with the Data Acquisition Toolbox by MATLAB (MathWorks).

3) Musical pieces: We extracted 20 musical pieces, using the music computation and notation software Sibelius (Avid Technology, USA). Each piece consisted of melodies produced by piano sounds, as shown in Table I. The length of each musical piece was 32 seconds with the tempo set to 150 beats per minute (bpm) (i.e., the frequency of a quarter of a note was 2.5 Hz). The sampling frequency was set to 44,100 Hz and resampled to 32,768 Hz for analysis.

4) Tasks: In the experiment, the participants listened to the musical pieces while visually fixating on a stationary position. The experiment consisted of two sessions where each session included 20 musical pieces. After listening to each musical pieces, the participants were asked whether they were familiar with the presented piece. The EEG recordings with respect to all the pieces were assessed to detect certain qualities, such as large-amplitude spikes. Each EEG recording was visually inspected during the experiment. If the EEG was contaminated with a large amount of artifacts, the same musical piece was replayed to record the EEG.

After listening to each musical piece, we labeled the piece as familiar or unfamiliar according to the answers of the participants. If participants' answer are not consistent across the sessions, they were excluded from the datasets. The number of data for each participant is summarized in Table II.

B. Feature Extraction

Our previous work [22] showed that the cross-correlation function averaged across trials, channels, and participants, showed a pronounced peak. At the peak, the cross-correlation magnitudes while participants listening to unfamiliar and scrambled music were significantly larger than they were while listening to familiar music [22]. In this paper, we considered the cross-correlation values between EEG and music envelope for each channel as potential features for classification.

1) Preprocessing: Thirty-second epochs of the EEG and music recordings were used, excluding the first second after the onset of the musical pieces and the last second before the end of them to remove filtering edge effects. For the recorded music, a zero-phase digital high-pass Butterworth filter (1 Hz) was applied to the recorded musical pieces. Then, the envelope of the filtered musical pieces was calculated using the Hilbert transform. Moreover, a zero-phase digital bandpass Butterworth filter between 1 and 40 Hz was applied to the recorded EEG and the envelopes. The filtered EEG and envelopes were down sampled to 256 Hz. Finally, the z-scores were calculated.

2) Cross-correlation function: First, the cross-correlations between the envelope of the music signals and the EEG signals were computed for time lags between -0.6 and 0.6 seconds, as follows:

$$C(n,\tau) = \sum_t x(t)y(n,t+\tau), \tag{1}$$

where $x(t)$ and $y(t)$ denote the filtered standardized (z-scored) envelope of a sound stimulus and the corresponding filtered standardized (z-scored) EEG response at time t and channel n, respectively. In addition, τ denotes the time lag between the envelope and the EEG signal. Second, we calculated the standard deviation of the cross-correlation function across the electrodes. Then, the peak was defined as the maximum value of the standard deviation in the time lag with the largest peak. This peak was used as a feature.

C. Classification

To determine the familiarity with music, feature classification was achieved using three different methods, Radial Basis Function based SVM (RBF–SVM), linear SVM (LSVM), and DNN. Since the number of electrodes was 31, the feature vector consisted of 31 cross-correlation features. The

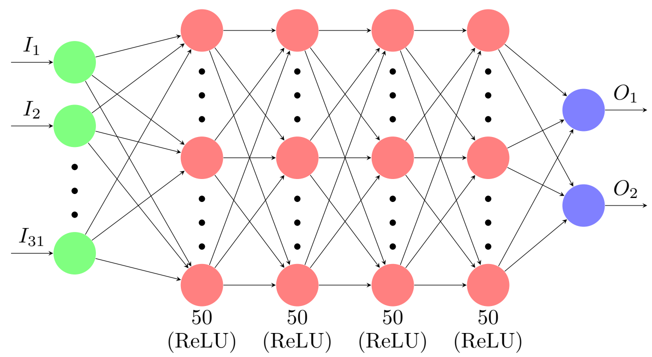

Fig. 2: Architecture of the neural network. ReLU [24] was used as the activation function, and the final layer was a softmax classifier. The size of the hidden layer was 4, and the number of neuron elements was 50.

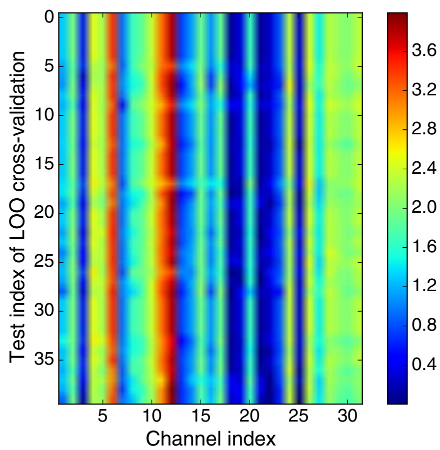

Fig. 3: Weight of the linear support vector machine for Participant 1 (s1m). The horizontal axis indicates the channel index corresponding to Fig. 1.

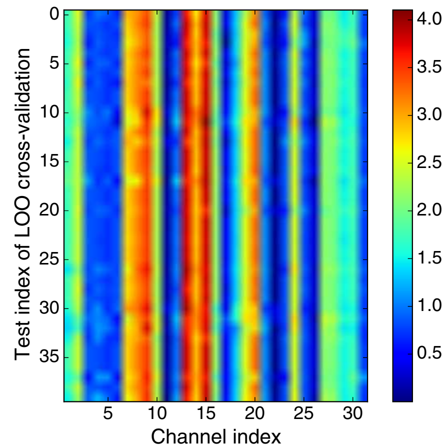

Fig. 4: Weight of the linear support vector machine for Participant 2 (s2m). The horizontal axis indicates the channel index corresponding to Fig. 1.

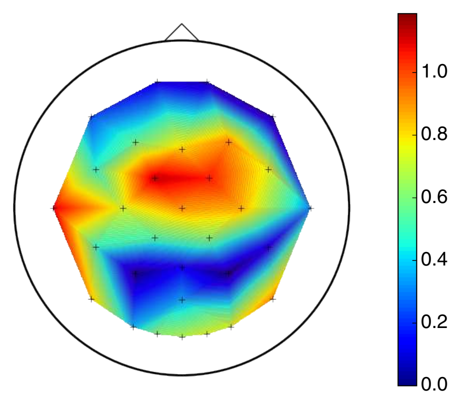

Fig. 5: The weight of each electrode averaged across participants.

classification analysis was performed individually for each participant. We used leave–one–out (LOO) cross-validation.

1) SVM: As for RBF–SVM, parameter C was set to 1, and parameter γ was set to 100. The LSVM's parameter C was set to 10. RBF–SVM and LSVM were utilized as implemented in the open-source machine learning library scikit-learn [23].

2) Deep neural network: The neural network architecture that we used is shown in Fig. 2. The DNN was trained using batch training with momentum. We used ReLU as the activation function [24], and the final layer was a softmax classifier [25], [26]. The number of hidden layers was 4, and the number of neuron elements was 50. The learning rate was 0.01.

III. RESULTS AND DISCUSSION

The classification accuracy for each participant and each classification method are shown in Table II. RBF–SVM showed the best performance with an accuracy of 100%. All classifiers' accuracies were over 50% (i.e., above the chance level).

In order to discuss which channel of the EEG is important for classification, we plotted the magnitude of the weight of the LSVM by each LOO cross-validation. Figs. 3 and 4 show the weights of participants 1 (s1m) and 2 (s2m). The horizontal axis indicates the channel index corresponding to Fig. 1, and the vertical axis indicates the test index of LOO cross-validation. As shown in Figs. 3 and 4, the weights were consistent across the LOO tests. Fig. 5 shows the weight of the LSVM averaged across LOO tests and participants as a topographical map.

According to Figs. 3–5, the weight coefficient of the classifier is larger in the frontal area than in the occipital area. These results suggest that the frontal area is important to classify brain states. Meltzer et al. [12] showed that the EEG responses to the beat of the music were larger in the

TABLE II: Classification accuracy for each participant-specific model. We used RBF–SVM, LSVM, and DNN. The best performance was 100% with RBF–SVM. The bracket below the participant's index indicates the number of data for each participant.

	Accuracy [%]								
Model	s1m (40)	s2m (40)	s3m (38)	s4m (34)	s5m (36)	s6m (40)	s7m (36)	s8m (36)	Mean ± Standard deviation
Radial Basis Function based support vector machine	93	100	63	68	69	63	81	58	74 ± 15
Linear support vector machine	95	98	53	53	54	58	81	56	68 ± 19
Deep neural network	93	98	61	62	64	65	81	53	72 ± 16

frontal, fronto-parietal, and central areas. Kong et al. [27] reported that the central area was activated strongly while listening to speech. The above-mentioned result is supportive of the findings of previous studies [12], [27].

IV. CONCLUSION

We proposed an approach to evaluate the familiarity with music using EEG. The cross-correlation values between the EEG and the envelope of the music signals were used in the RBF–SVM, LSVM, and DNN. The best participant achieved a classification rate of 100% using RBF–SVM. These results suggest that cross-correlation values can be used to recognize high-level brain states such as familiarity, preference, and emotion. Further, the proposed feature can be adapted for BCIs using natural music.

REFERENCES

[1] 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." *J. Neural. Eng.*, vol. 11, no. 2, p. 026009, 2014.

[2] I. Daly, D. Williams, A. Kirke, J. Weaver, A. Malik, F. Hwang, E. Miranda, and S. J. Nasuto, "Affective brain-computer music interfacing," *J. Neural. Eng.*, vol. 13, no. 4, p. 046022, 2016.

[3] Y. Pan, C. Guan, J. Yu, K. K. Ang, and T. E. Chan, "Common frequency pattern for music preference identification using frontal EEG," in *Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on*. IEEE, 2013, pp. 505–508.

[4] K. C. Tseng, B.-S. Lin, A. M.-K. Wong, and B.-S. Lin, "Design of a mobile brain computer interface-based smart multimedia controller," *Sensors*, vol. 15, no. 3, pp. 5518–5530, 2015.

[5] I. Daly, D. Williams, J. Hallowell, F. Hwang, A. Kirke, A. Malik, J. Weaver, E. Miranda, and S. J. Nasuto, "Music-induced emotions can be predicted from a combination of brain activity and acoustic features," *Brain Cogn.*, vol. 101, pp. 1–11, 2015.

[6] S. K. Hadjidimitriou and L. J. Hadjileontiadis, "EEG-based classification of music appraisal responses using time-frequency analysis and familiarity ratings," *IEEE Trans. Affect. Comput.*, vol. 4, no. 2, pp. 161–172, 2013.

[7] R. Näätänen, A. W. Gaillard, and S. Mäntysalo, "Early selective-attention effect on evoked potential reinterpreted," *Acta. Psychol.*, vol. 42, no. 4, pp. 313–329, 1978.

[8] C. Lappe, O. Steinsträter, and C. Pantev, "Rhythmic and melodic deviations in musical sequences recruit different cortical areas for mismatch detection." *Front. Hum. Neurosci.*, vol. 7, no. June, p. 260, 2013.

[9] P. Virtala, M. Huotilainen, E. Partanen, and M. Tervaniemi, "Musicianship facilitates the processing of western music chords–an ERP and behavioral study," *Neuropsychologia*, vol. 61, no. 1, pp. 247–258, 2014.

[10] O. G. Lins and T. W. Picton, "Auditory steady-state responses to multiple simultaneous stimuli," *Electroencephalogr. Clin. Neurophysiol./Evoked Potentials Section*, vol. 96, no. 5, pp. 420–432, 1995.

[11] S. Nozaradan, "Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 369, no. 1658, p. 20130393, 2014.

[12] B. Meltzer, C. S. Reichenbach, C. Braiman, N. D. Schiff, A. Hudspeth, and T. Reichenbach, "The steady-state response of the cerebral cortex to the beat of music reflects both the comprehension of music and attention," *Front. Hum. Neurosci.*, vol. 9, p. 11, 2015.

[13] N. Ding and J. Z. Simon, "Cortical entrainment to continuous speech: Functional roles and interpretations," *Front. Hum. Neurosci.*, vol. 8, p. 311, 2014.

[14] E. Ahissar, S. Nagarajan, M. Ahissar, A. Protopapas, H. Mahncke, and M. M. Merzenich, "Speech comprehension is correlated with temporal response patterns recorded from auditory cortex," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 98, no. 23, pp. 13 367–13 372, 2001.

[15] S. J. Aiken and T. W. Picton, "Envelope and spectral frequency-following responses to vowel sounds," *Hear. Res.*, vol. 245, no. 1-2, pp. 35–47, 2008.

[16] K. V. Nourski, R. A. Reale, H. Oya, H. Kawasaki, C. K. Kovach, H. Chen, M. A. Howard, and J. F. Brugge, "Temporal envelope of time-compressed speech represented in the human auditory cortex," *J. Neurosci.*, vol. 29, no. 49, pp. 15 564–15 574, 2009.

[17] E. C. Lalor, A. J. Power, R. B. Reilly, and J. J. Foxe, "Resolving precise temporal processing properties of the auditory system using continuous stimuli," *J. Neurophysiol.*, vol. 102, no. 1, pp. 349–359, 2009.

[18] J. R. Kerlin, A. J. Shahin, and L. M. Miller, "Attentional control of ongoing cortical speech representations in a "cocktail party"," *J. Neurosci.*, vol. 30, no. 2, pp. 620–628, 2010.

[19] N. Ding and J. Z. Simon, "Emergence of neural encoding of auditory objects while listening to competing speakers," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 109, no. 29, pp. 11 854–11 859, 2012.

[20] J. A. O'Sullivan, A. J. Power, N. Mesgarani, S. Rajaram, J. J. Foxe, B. G. Shinn-Cunningham, M. Slaney, S. A. Shamma, and E. C. Lalor, "Attentional selection in a cocktail party environment can be decoded from single-trial EEG," *Cereb. Cortex*, vol. 25, no. 7, pp. 1697–1706, 2015.

[21] K. B. Doelling and D. Poeppel, "Cortical entrainment to music and its modulation by expertise," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 112, no. 45, pp. E6233–E6242, 2015.

[22] Y. Kumagai, M. Arvaneh, and T. Tanaka, "Familiarity affects entrainment of EEG in music listening," *Submitted*.

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[24] A. Mnih and G. E. Hinton, "A scalable hierarchical distributed language model," in *Adv. Neural. Inf. Process. Syst.*, 2009, pp. 1081–1088.

[25] F. Seide, G. Li, and D. Yu, "Conversational speech transcription using context-dependent deep neural networks," in *Interspeech*, 2011, pp. 437–440.

[26] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, 2012.

[27] Y.-Y. Kong, A. Mullangi, and N. Ding, "Differential modulation of auditory responses to attended and unattended speech in different listening conditions," *Hear. Res.*, vol. 316, pp. 73–81, 2014.