Emotion Estimation from EEG Signals and Expected Subjective Evaluation

Kana Miyamoto, Hiroki Tanaka, Satoshi Nakamura
Division of Information Science
Nara Institute of Science and Technology
Ikoma-shi, Nara, Japan
{miyamoto.kana.mk4, hiroki-tan, s-nakamura}@is.naist.jp

***Abstract*—We estimated emotions from electroencephalogram (EEG) to construct an emotion induction system using music. Although a previous study proposed to construct a system from emotion estimation based on EEG only, the accuracy of emotion estimation just from EEG might be insufficient and the model was not tested in leave-one-music-out manner. In this study, we newly proposed emotion estimation using EEG and the emotions that are expected to be felt through music. First, we created a music generator that makes music that induces emotions that resemble the inputs of emotional values. Next, we recorded EEG and subjective evaluations of felt emotions while participants listened to the music created by the music generator. We estimated the emotions by the linear regression of the baseline model and the convolutional neural network (CNN) of the proposed model using recorded EEG and subjective evaluations and investigated whether using either all channels (29 EEG channels) or just 14 EEG channels was effective. Finally, we estimated emotions by neural network using the values estimated from the CNN and the music generator's inputs. As a result of adapting leave-one-music-out cross-validation, we obtained the lowest RMSE (valence: 0.151, arousal: 0.164) between the actual and estimated emotional values where the models combined the emotions estimated by CNN that were trained from 14 channels of EEG and the music generator's inputs. For the Wilcoxon signed-rank test result, we found a significant difference between our proposed model and the baseline model.**

***Index Terms*—electroencephalogram; emotion estimation; brain computer interface; human computer interaction**

I. Introduction

Music induces emotions [1], [2]. It has been used to improve the mental health such as senior citizens [3]. Researchers have studied systems that induce emotions using music that effectively changes emotions. Such systems suffer from two problems. The first concerns what kind of music to select. The emotions that are induced when listening to music are very subjective [4]. The same person may have different emotional levels depending on the situation. A system needs to select music that is appropriate for the individual and the situation. The second problem is a method that estimates the emotional state of participants. The system needs to change the music to adapt to the emotions of the person.

Ehrlich et al. proposed a system that automatically generates music that is designed to actively induce emotions [5]. They proposed emotion estimation using the electroencephalogram (EEG) of biological signals to construct their system. EEG records the brain's electrical activity and measures time-series data in multiple channels. Various researches use EEG, such as a combination of eye-tracking and BCI techniques for controlling 3D objects [6] and decoding robot errors using deep convolutional neural network [7]. EEG is also effective for emotion estimation [8], [9]. Emotions are expressed in two-dimensional space [10]. The horizontal dimension is the valence, representing a range from pleasant to unpleasant. The vertical dimension is arousal, representing a range from activation to deactivation. Ehrlich et al. proposed emotion estimation in the continuous values of valence and arousal using EEG. They generated music for participants to perceive the estimated emotions. The participants recognized their emotions from the music and induced the desired emotions by themselves. This study doesn't passively induce emotions.

We proposed a system that automatically generates music that induces emotions passively just by listening [11]. Our system estimated emotions as continuous values using valence and arousal by EEG and generated music to induce emotions. First, based on Ehrlich et al., we created a music generator that induced emotions. The music generator made music by inputting the valence and arousal of the emotions to be induced. We evaluated our music generator using crowdsourcing and concluded that it effectively induced emotions. Second, we recorded EEG and subjective evaluations of the felt emotions while participants listened to music created by the music generator and calculated the features for each 1 s of the recorded EEG for training emotion estimation models. We estimated the emotions by two models (linear regression and convolutional neural network (CNN)) with the features of 14 channels used by Ehrlich et al. The study obtained the RMSE between the actual and estimated emotional values by applying the holdout.

However, there are three problems with our previous study. The first is model selection. The study calculated the features of 14 selected channels instead of all the EEG channels. Emotion estimation using all the recorded channels must also be verified. In this paper, we compared linear regression and CNN using 14 channels that were used by Ehrlich et al. and linear regression and CNN using all EEG channels.

The second is the verification method. The study acquired multiple samples from one piece of music and verified them

978-1-7281-8485-2/21/$31.00 ©2021 IEEE

using the holdout without considering the type of music. EEG while listening to the same music was be included in the test and training data in the holdout. When listening to the same music, EEG has similar features. This result might lower the RMSE more than the actual environment used. In this study, we calculated the RMSE between the actual and estimated emotional values by applying the leave-one-music-out cross-validation instead of the holdout.

The third is a further performance improvement. The accuracy of the estimated emotions from EEG is important in the emotion induction system. However, the accuracy of emotion estimation just using EEG might be insufficient. If the estimated emotions differ significantly from the emotions felt by the participants, the system might generate music that struggles to induce the desired emotions. Therefore we chose to use EEG and other information. Previous studies used EEG and galvanic skin response to estimate emotions [12] and EEG and audio information to predict speech quality [13]. Since the emotion induction system needs very quick emotion estimation, using minimal biological signals and emotion estimation models is required. We propose an emotion estimation method that combines EEG for biological signals and the inputs of a music generator that induces emotions that resemble the emotions of the inputs. The music generator's inputs can be regarded as expected subjective evaluations. We expect the expected subjective evaluations to complement the emotion estimation by EEG. From the estimated values using EEG and the values of expected subjective evaluations, we estimated emotions by the neural network.

In summary, this paper investigates the following three aspect:

1) Selection of EEG channels and models.
2) Adaptation of leave-one-music-out cross-validation.
3) Emotion estimation using EEG and expected subjective evaluations.

II. EEG Data Collection

We recorded the EEG and the subjective evaluations of the felt emotions while participants listened to music produced by the music generator. In this section, we introduce our stimuli of generated music and experimental design for recording the EEG and the subjective evaluations.

A. Stimuli of Generated Music

Ehrlich et al. created a music generator, which helps participants perceive the intended emotions of music [5]. Based on previous studies, the felt and perceived emotions are different, but these emotions are the same or the felt emotion often appears lower than the perceived emotion [14], [15]. Therefore we created a music generator to induce emotions based on Ehrlich et al. 's system. A music generator composes music by calculating five musical parameters: tempo, rhythm, loudness, pitch, and mode from the inputs of valence and arousal [16]. We calculated these musical parameters from the inputs of valence and arousal between 0 and 1 and generated musical instrument digital interface (MIDI) signals, which were sent to DAW software that generated music by piano, violin, and cello. After evaluating the music generator using crowdsourcing, we concluded that it effectively induced emotions. In this study, we used the same stimuli as our previous study [11].

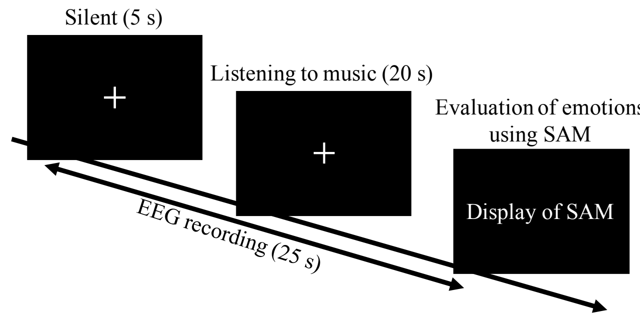


Fig. 1. Overview of experimental design: This figure outlines the experiment for one piece of music. This outline was repeated for 41 pieces of music.

B. Participants

Twenty healthy undergraduate and graduate students (10 males, 10 females) participated in this experiment, which was approved by the ethics committee of Nara Institute of Science and Technology.

C. Design

The experimental procedure is shown in Fig. 1. The participants silently gazed at a cross in the center of the monitor for 5 s. Then they listened to music for 20 s and gazed at the cross. The screen changed after they listened to the music for the evaluation of their felt emotions. They evaluated the valence and arousal of the felt emotions in nine steps between 0 and 1 with a self-assessment mannequin (SAM) [17].

The participants sat in front of a desk on which a monitor was placed. Before putting on the electroencephalograph, they wore earphones and listened to five pieces of music sample (15 s) {val, aro}={0,0}; {0,1}; {0.5,0.5}; {1,0}; {1,1}. They practiced the experiment with two pieces of music (20 s) {val, aro}={0.125,0.25}; {0.875,0.75}. After the practice, we put a CGX Quick-30 electroencephalograph on them and recorded their EEG and their subjective evaluations of the 41 pieces of music (20 s). The 41 pieces of music were chosen uniformly from the coordinates of valence and arousal.

D. Preprocessing

We used MATLAB (R2019a) and EEGLAB [18] for preprocessing, which consisted of the following seven steps: 1) We removed the data that caused problems such as the music wasn't played; 2) The EEG signals were downsampled from 1000Hz to 200Hz; 3) The silent state of 2 to 5 s was divided into three epochs of 1 s data; 4) The music-listening state of 0 to 20 s was divided into 20 epochs of 1 s data; 5) We designed 2nd order zero phase Chebyshev IIR bandpass filters that pass theta (4-7 Hz), alpha (8-13 Hz), low beta (14-21 Hz), high beta (22-29 Hz), and gamma (30-45 Hz); 6) EEG signals were divided into five frequency bands by the designed

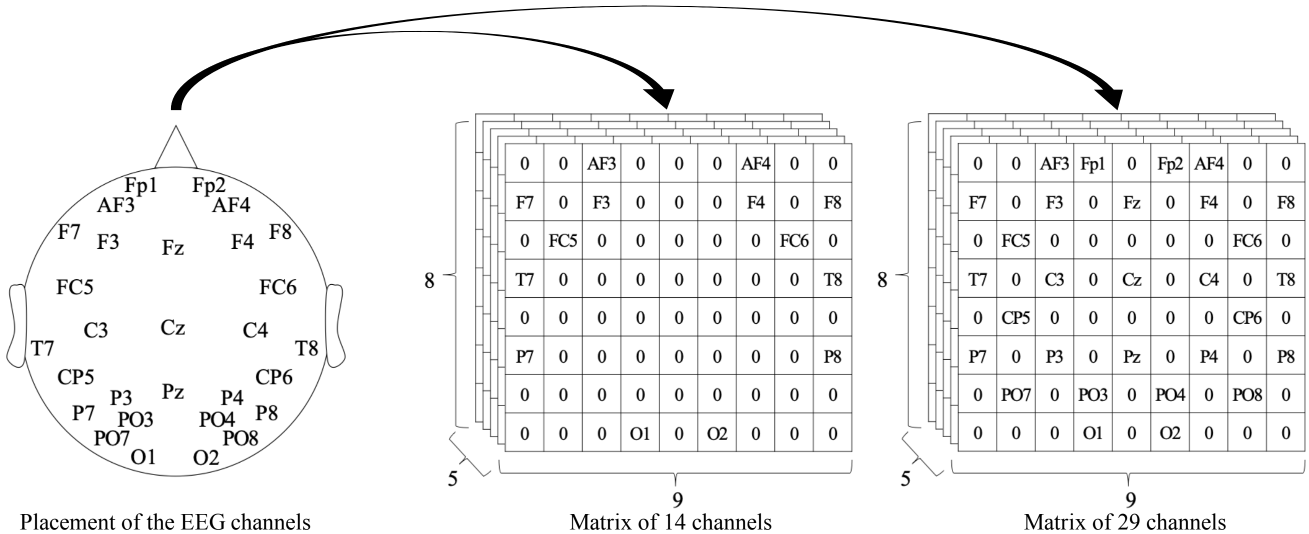


Fig. 2. Matrix based on placement of EEG channels: We used two matrices to estimate valence and arousal in CNN.

filters to calculate the $f = log(var(EEGdata))$, which is the logarithm of the waveform variance for each bit of data; 7) The average of the logarithm of the silent state waveform variance was subtracted from the logarithmic waveform variance during the music-listening state for each type of music.

III. Emotion Estimation

We verified the three proposed emotion estimations using the preprocessing results. We used MATLAB (R2019b) for training the emotion estimation models.

A. Selection of EEG channels and models

We verified the models for estimating emotions only using EEG. We prepared two types of channels: just 14 channels and all 29 channels. We prepared two types of models: linear regression and CNN. By combining the channels and models, we examined four models: linear regression using 14 channels, CNN using 14 channels, linear regression using 29 channels, and CNN using 29 channels. We evaluated the models using the leave-one-music-out cross-validation.

1) Emotion estimation using 14 channels: Ehrlich et al. [5] acquired 14 channels using Emotiv EPOC: AF3, AF4, F3, F4, F7, F8, FC5, FC6, T7, T8, P7, P8, O1, and O2. We used CGX Quick-30, acquired 29 channels, and selected 14 from the 29 channels we recorded.

2) Emotion estimation using 29 channels: We acquired 29 channels using CGX Quick-30: Fp1, Fp2, AF3, AF4, F7, F8, F3, Fz, F4, FC5, FC6, T7, T8, C3, Cz, C4, CP5, CP6, P7, P8, P3, Pz, P4, PO7, PO8, PO3, PO4, O1, and O2. We used all 29 channels to train the emotion estimation models.

3) Linear regression: Ehrlich et al. estimated emotions based on linear discriminant analysis (LDA) and the sigmoid function using the logarithm of the EEG variances. Linear regression is a similar model of Ehrlich et al. We used linear regression as the baseline model. Our model trained with 14 channels has 70 features (14 channels$\times$5 frequency bands). Our model trained with 29 channels has 145 features (29 channels$\times$5 frequency bands). The features were input in vector format and the data were normalized using z-score normalization. In linear regression, the models that estimate valence and arousal are different.

4) CNN: CNN has feature extraction capabilities. CNN is also used in studies on EEG [19], [20]. We used CNN for training that takes into account the positional relationships of the EEG channels. We mapped the matrix for each of the five frequency bands (Fig. 2). The placement was based on the method of previous studies [21], [22]. We used zero to fill the element where there were no EEG channels. We prepared $8\times9\times5$ matrix. As shown in Fig. 4, the CNN consists of a convolution layer (2×2 size, 1 stride), a batch normalization layer, a ReLU layer, a convolution layer (2×2 size, 1 stride), a batch normalization layer, a ReLU layer, a dropout layer (dropout rate of 0.5), a fully connected layer (output dimensionality of 2), and a regression output layer. We used the Adam optimizer. The learning rate was 0.001. The batch size was 64. The learning epoch was 100. In CNN, valence and arousal were estimated from one model.

B. Adaptation of leave-one-music-out cross-validation

From the preprocessing, we acquired 20 samples from a piece of music (Fig. 3). We assume that adjacent samples have similar features. If the holdout is applied without considering the type of music, adjacent samples are included in the test and training data. The emotion induction system needs to estimate emotions from the EEG when listening to unknown music. Since the holdout might show higher accuracy than emotion estimation in actual environments, we apply the leave-one-music-out cross-validation to test one piece of music that is excluded from the training data.

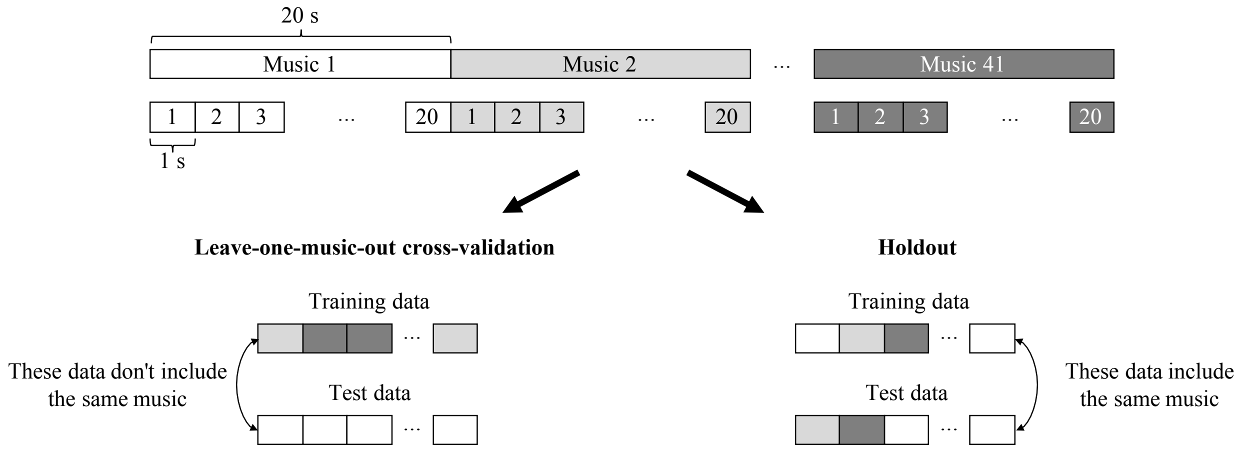


Fig. 3. Overview of the leave-one-music-out cross-validation and the holdout.

C. Emotion estimation using EEG and expected subjective evaluations

The accuracy of the emotion estimation from EEG might only be insufficient. One solution to this problem is estimating emotions by adding information other than EEG. We used EEG and the music generator's inputs. After evaluating the music made by our music generator, we found that it induced emotions similar to the inputs of valence and arousal. Accordingly, we regarded the inputs of the music generator as expected subjective evaluations and assumed that the inputs can be used for emotion estimation. We estimated emotions using emotions estimated from EEG and the inputs to the music generator. Using the music generator's inputs, we expect that the influence on the final estimated emotions will be reduced even if the estimated values by EEG are incorrect.

We prepared four inputs of emotions estimated from EEG and those of the music generator. As shown in Fig. 4, the neural network consists of a fully connected layer (output dimensionality of 8), a ReLU layer, a dropout layer (dropout rate of 0.2), a fully connected layer (output dimensionality of 2), and a regression output layer. We used the Adam optimizer. The learning rate was 0.001. The batch size was 64. The learning epoch was 100. We evaluated the models using leave-one-music-out cross-validation.

IV. Result

A. Selection of EEG channels and models

We estimated the emotions only using EEG from four models: linear regression using 14 channels, CNN using 14 channels, linear regression using 29 channels, and CNN using 29 channels. Table I shows the results of the comparison of the felt and estimated emotions. We obtained the lowest RMSE (valence: 0.214, arousal: 0.239) by CNN using 14 channels. For the Wilcoxon signed-rank test result, we found a significant difference between CNN using 14 channels and the linear regression using 14 channels of the baseline model for both valence and arousal ($p<0.05$).

TABLE I
RMSE of felt and estimated emotions using four models. Bold indicates lowest RMSE.

Par.	Linear regression				CNN			
	14 channels		29 channels		14 channels		29 channels	
	val	aro	val	aro	val	aro	val	aro
1	0.285	0.330	0.290	0.341	0.261	0.282	0.261	0.283
2	0.307	0.234	0.328	0.246	0.287	0.217	0.292	0.227
3	0.276	0.286	0.290	0.279	0.247	0.229	0.253	0.246
4	0.118	0.164	0.131	0.191	0.104	0.140	0.116	0.153
5	0.334	0.283	0.350	0.316	0.312	0.250	0.324	0.252
6	0.244	0.285	0.259	0.313	0.222	0.266	0.236	0.276
7	0.403	0.410	0.423	0.424	0.353	0.372	0.362	0.379
8	0.259	0.308	0.288	0.323	0.217	0.277	0.228	0.291
9	0.213	0.290	0.224	0.302	0.192	0.254	0.200	0.265
10	0.074	0.166	0.083	0.193	0.073	0.152	0.076	0.153
11	0.165	0.233	0.171	0.243	0.151	0.217	0.160	0.220
12	0.216	0.291	0.211	0.307	0.195	0.263	0.199	0.260
13	0.218	0.318	0.224	0.333	0.208	0.302	0.216	0.307
14	0.193	0.227	0.195	0.249	0.174	0.193	0.175	0.202
15	0.252	0.239	0.253	0.233	0.228	0.210	0.232	0.210
16	0.081	0.165	0.090	0.182	0.067	0.148	0.085	0.160
17	0.464	0.379	0.496	0.399	0.411	0.369	0.417	0.372
18	0.178	0.189	0.177	0.192	0.152	0.152	0.155	0.159
19	0.077	0.222	0.082	0.231	0.067	0.190	0.072	0.197
20	0.384	0.367	0.421	0.386	0.365	0.299	0.363	0.329
mean	0.237	0.269	0.249	0.284	**0.214**	**0.239**	0.221	0.247
std	0.108	0.071	0.116	0.072	0.099	0.068	0.098	0.068

B. Adaptation of leave-one-music-out cross-validation

We found that CNN using 14 channels of EEG had the lowest RMSE of the four models that only used EEG. We used CNN with 14 channels and compared emotion estimation by considering the type of music using leave-one-music-out cross-validation and emotion estimation without considering the type of music using k-fold cross-validation. In this paper, we applied k-fold cross-validation instead of the holdout to improve the generalizability. The amount of test data for the leave-one-music-out cross-validation and the k-fold cross-validation was identical. Tables II shows the results of the comparison of the felt and estimated emotions. We obtained the lower RMSE (valence: 0.202, arousal: 0.226) with the k-fold cross-validation than RMSE (valence: 0.214, arousal:

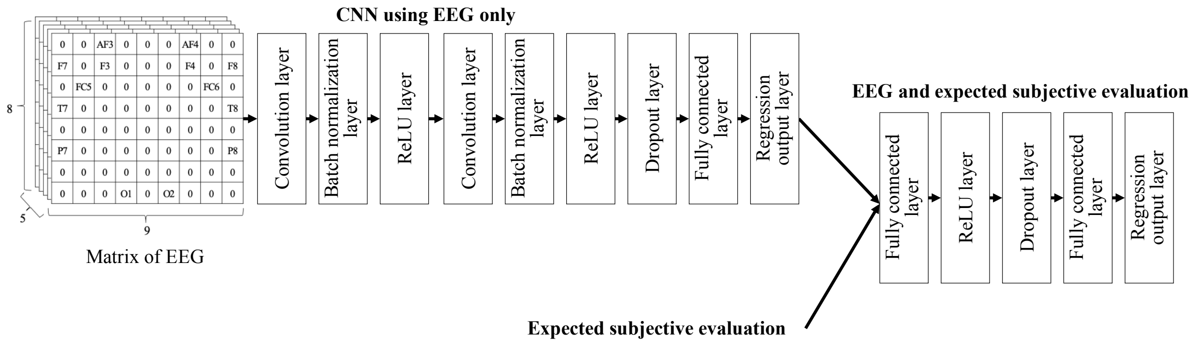


Fig. 4. Neural network using EEG and expected subjective evaluations: This neural network estimates valence and arousal from four inputs. Four inputs are valence and arousal only estimated from EEG and valence and arousal of the music generator's inputs.

TABLE II
RMSE OF FELT AND ESTIMATED EMOTIONS APPLYING LEAVE-ONE-MUSIC-OUT CROSS-VALIDATION AND K-FOLD CROSS-VALIDATION. BOLD FONTS INDICATE THE LOWEST RMSE.

Par.	leave-one-music-out		k-fold	
	val	aro	val	aro
1	0.261	0.282	0.240	0.268
2	0.287	0.217	0.255	0.210
3	0.247	0.229	0.247	0.234
4	0.104	0.140	0.105	0.140
5	0.312	0.250	0.285	0.239
6	0.222	0.266	0.211	0.251
7	0.353	0.372	0.331	0.340
8	0.217	0.277	0.209	0.252
9	0.192	0.254	0.188	0.237
10	0.073	0.152	0.068	0.150
11	0.151	0.217	0.141	0.201
12	0.195	0.263	0.188	0.240
13	0.208	0.302	0.189	0.270
14	0.174	0.193	0.167	0.193
15	0.228	0.210	0.209	0.209
16	0.067	0.148	0.087	0.154
17	0.411	0.369	0.374	0.320
18	0.152	0.152	0.145	0.144
19	0.067	0.190	0.072	0.182
20	0.365	0.299	0.328	0.293
mean	0.214	0.239	**0.202**	**0.226**
std	0.099	0.068	0.086	0.057

TABLE III
RMSE OF FELT AND ESTIMATED EMOTIONS USING EEG AND EXPECTED SUBJECTIVE EVALUATIONS. BOLD FONTS INDICATE THE LOWEST RMSE.

Par.	CNN		Music gen.		Comb.	
	val	aro	val	aro	val	aro
1	0.261	0.282	0.140	0.125	0.124	0.150
2	0.287	0.217	0.174	0.216	0.180	0.144
3	0.247	0.229	0.140	0.140	0.137	0.130
4	0.104	0.140	0.266	0.301	0.097	0.128
5	0.312	0.250	0.192	0.152	0.205	0.143
6	0.222	0.266	0.293	0.122	0.223	0.169
7	0.353	0.372	0.122	0.149	0.143	0.188
8	0.217	0.277	0.402	0.244	0.228	0.197
9	0.192	0.254	0.238	0.128	0.157	0.128
10	0.073	0.152	0.256	0.360	0.053	0.122
11	0.151	0.217	0.232	0.186	0.115	0.157
12	0.195	0.263	0.305	0.253	0.166	0.220
13	0.208	0.302	0.177	0.201	0.135	0.222
14	0.174	0.193	0.277	0.244	0.172	0.134
15	0.228	0.210	0.175	0.306	0.143	0.177
16	0.067	0.148	0.293	0.244	0.068	0.137
17	0.411	0.369	0.146	0.137	0.155	0.169
18	0.152	0.152	0.220	0.250	0.102	0.139
19	0.067	0.190	0.253	0.213	0.057	0.134
20	0.365	0.299	0.347	0.284	0.367	0.282
mean	0.214	0.239	0.232	0.213	**0.151**	**0.164**
std	0.099	0.068	0.075	0.070	0.071	0.041

0.239) with the leave-one-music-out. For the Wilcoxon signed-rank test result, we found a significant difference between and the k-fold cross-validation and the leave-one-music-out cross-validation for both the valence and arousal ($p<0.05$).

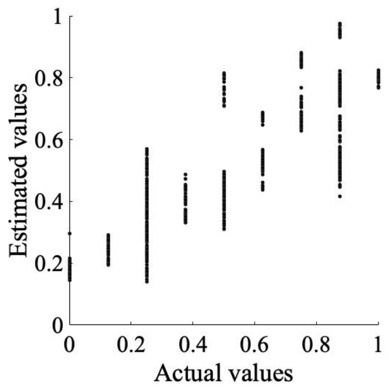
C. Emotion estimation using EEG and expected subjective evaluations

We estimated the emotions from CNN using 14 channels and the music generator's inputs. Table III shows the results of the comparison of the felt and estimated emotions. CNN is CNN using 14 channels of EEG only, Music gen. is the inputs of the music generator, and Comb. is the neural network that combines CNN and Music gen. We obtained the lowest RMSE (valence: 0.151, arousal: 0.164) by the EEG and expected subjective evaluations. For the Wilcoxon signed-rank test result, we found a significant difference between our proposed method and the linear regression using 14 channels of the baseline model for both the valence and arousal ($p<0.05$).

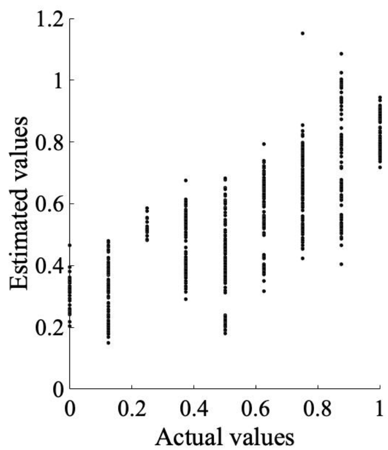
We show the plots of the actual and estimated emotions of the valence and arousal of participant 1 in Fig. 5. There was a positive correlation between the actual emotions and those estimated by EEG and the expected subjective evaluations.

V. CONCLUSION

We proposed an emotion estimation method that combines EEG and inputs of a music generator. We first estimated the emotions using EEG only and obtained the lowest RMSE by CNN using 14 channels. We assume the CNN that takes into account the positional relationship of channels is more effective than linear regression. Our result also indicates the effectiveness of selecting just some channels instead of all of them. We also estimated emotions by applying two cross-validation methods. The RMSE was higher when the leave-



Actual and estimated emotions of valence
Pearson's correlation coefficient ($r=0.841$, $p<0.05$)



Actual and estimated emotions of arousal
Pearson's correlation coefficient ($r=0.809$, $p<0.05$)

Fig. 5. Plots of the actual and estimated emotions of participant 1: The emotions were estimated by EEG and expected subjective evaluations.

one-music-out cross-validation was applied than when the k-fold cross-validation was applied, although we assume that the leave-one-music-out cross-validation is similar to the actual environment. We then estimated the emotions by a model combining values estimated from EEG and inputs of the music generator. The RMSE by the model was the lowest of all our experiments. We consider that the inputs of the music generator compensated for the results of emotion estimation from EEG.

In the future, we will construct an emotion induction system that combines the music generator and our proposed method of emotion estimation and compare the estimated emotions when using the system with desired emotions.

ACKNOWLEDGMENT

This work was supported by JST CREST Grant Number JPMJCR19A5, Japan.

REFERENCES

[1] S. Koelsch, "A neuroscientific perspective on music therapy," *Ann. NY Acad. Sci.*, vol. 1169, pp. 374–384, 2009.
[2] I. Wallis, T. Ingalls, and E. Campana, "Computer-generating emotional music: The design of an affective music algorithm," *DAFx-08, Espoo, Finland*, vol. 712, pp. 7–12, 2008.
[3] R. Ramirez, M. Palencia-Lefler, S. Giraldo, and Z. Vamvakousis, "Musical neurofeedback for treating depression in elderly people," *Frontiers in neuroscience*, vol. 9, p. 354, 2015.
[4] P. J. Rentfrow and S. D. Gosling, "The do re mi's of everyday life: the structure and personality correlates of music preferences." *Journal of personality and social psychology*, vol. 84, no. 6, p. 1236, 2003.
[5] S. K. Ehrlich, K. R. Agres, C. Guan, and G. Cheng, "A closed-loop, music-based brain-computer interface for emotion mediation," *PloS one*, vol. 14, no. 3, pp. 1–24, 2019.
[6] J. Chun, B. Bae, and S. Jo, "Bci based hybrid interface for 3d object control in virtual reality," in *2016 4th International Winter Conference on Brain-Computer Interface (BCI)*. IEEE, 2016, pp. 1–4.
[7] J. Behncke, R. T. Schirrmeister, W. Burgard, and T. Ball, "The signature of robot action success in eeg signals of a human observer: Decoding and visualization using deep convolutional neural networks," in *2018 6th International Conference on Brain-Computer Interface (BCI)*. IEEE, 2018, pp. 1–6.
[8] M. Z. Soroush, K. Maghooli, S. K. Setarehdan, and A. M. Nasrabadi, "A review on eeg signals based emotion recognition," *International Clinical Neuroscience Journal*, vol. 4, no. 4, p. 118, 2017.
[9] H. Dabas, C. Sethi, C. Dua, M. Dalawat, and D. Sethia, "Emotion classification using eeg signals," in *Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence*, 2018, pp. 380–384.
[10] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.
[11] K. Miyamoto, H. Tanaka, and S. Nakamura, "Music generation and emotion estimation from eeg signals for inducing affective states," in *Companion Publication of the 2020 International Conference on Multimodal Interaction*, 2020, pp. 487–491.
[12] Y.-H. Kwon, S.-B. Shin, and S.-D. Kim, "Electroencephalography based fusion two-dimensional (2d)-convolution neural networks (cnn) model for emotion recognition system," *Sensors*, vol. 18, no. 5, p. 1383, 2018.
[13] I. H. Parmonangan, H. Tanaka, S. Sakti, and S. Nakamura, "Combining audio and brain activity for predicting speech quality," *Proc. Interspeech 2020*, pp. 2762–2766, 2020.
[14] E. Schubert, "Emotion felt by the listener and expressed by the music: literature review and theoretical perspectives," *Frontiers in psychology*, vol. 4, p. 837, 2013.
[15] A. Gabrielsson, "Emotion perceived and emotion felt: Same or different?" *Musicae scientiae*, vol. 5, no. 1_suppl, pp. 123–147, 2001.
[16] M. A. Schmuckler, "Expectation in music: Investigation of melodic and harmonic processes," *Music Perception: An Interdisciplinary Journal*, vol. 7, no. 2, pp. 109–149, 1989.
[17] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.
[18] 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*, vol. 134, no. 1, pp. 9–21, 2004.
[19] G. Xu, X. Shen, S. Chen, Y. Zong, C. Zhang, H. Yue, M. Liu, F. Chen, and W. Che, "A deep transfer convolutional neural network framework for eeg signal classification," *IEEE Access*, vol. 7, pp. 112 767–112 776, 2019.
[20] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from eeg with deep recurrent-convolutional neural networks," *arXiv preprint arXiv:1511.06448*, 2015.
[21] J. Li, Z. Zhang, and H. He, "Hierarchical convolutional neural networks for eeg-based emotion recognition," *Cognitive Computation*, vol. 10, no. 2, pp. 368–380, 2018.
[22] Y. Yang, Q. Wu, Y. Fu, and X. Chen, "Continuous convolutional neural network with 3d input for eeg-based emotion recognition," in *International Conference on Neural Information Processing*. Springer, 2018, pp. 433–443.