PAPER

Online EEG-Based Emotion Prediction and Music Generation for Inducing Affective States

Kana MIYAMOTO[†,††a)], ***Nonmember***, **Hiroki TANAKA**[†,††b)], ***and*** **Satoshi NAKAMURA**[†,††c)], ***Members***

SUMMARY Music is often used for emotion induction because it can change the emotions of people. However, since we subjectively feel different emotions when listening to music, we propose an emotion induction system that generates music that is adapted to each individual. Our system automatically generates suitable music for emotion induction based on the emotions predicted from an electroencephalogram (EEG). We examined three elements for constructing our system: 1) a music generator that creates music that induces emotions that resemble the inputs, 2) emotion prediction using EEG in real-time, and 3) the control of a music generator using the predicted emotions for making music that is suitable for inducing emotions. We constructed our proposed system using these elements and evaluated it. The results showed its effectiveness for inducing emotions and suggest that feedback loops that tailor stimuli to individuals can successfully induce emotions.

key words: *EEG, emotion induction, emotion prediction, music generation*

1. Introduction

Emotions have been studied extensively [1]–[3]. Negative emotions narrow thought-action repertoires, and positive emotions broaden the scope of attention [4]. Positive emotions are also effective for preventing and treating such mood disorders as depression [5]. Since emotions have various roles, they should be appropriately induced.

Music is a stimulus that induces emotions [6]. Its effects on emotions have also been investigated [7]–[9]. The emotions we perceive and feel when we listen to music are different [10], [11]. Felt emotions are often weaker or identical as perceived emotions. The effects of such musical parameters as tempo and rhythm on emotions have also been investigated. Emotions were expressed based on a circumplex model [12], which argues that they have a two-dimensional space with valence and arousal. Valence represents a range from pleasant to unpleasant. Arousal represents a range from activated to deactivated. Based on investigations of the relationship between musical parameters and induced emotions, scale and mode affect valence, and tempo, rhythm, and loudness affect arousal [13], [14]. In this way, music is related to emotional induction. However, not all people experience the same emotions even if they listen to the same music [15]. Music that effectively induces particular emotions varies from person to person. Therefore, it is necessary to play music according to the participant for more effective emotion induction.

Subjective evaluations are often used to judge emotions. The Self-Assessment Mannequin (SAM) is utilized for emotional evaluation experiments [16]. Many studies have used it to measure valence, arousal, and dominance. Unfortunately, since subjective evaluations often burden participants with too many tasks, interest is growing to use biological signals to recognize emotions. An electroencephalogram (EEG) records the electrical activity of the brain. Emotion prediction with it is being actively studied [17], [18], and it is expected to be used for a brain computer interface (BCI) and a human computer interaction (HCI). There has been a lot of research on EEG recording using video as a stimulus. One previous study predicted emotions from EEG while watching videos for video summarization [19]. AMIGOS [20] is a dataset containing biological signals while watching videos. There are also studies that use music as a stimulus. DEAP dataset [21] is a dataset containing biological signals while watching music videos. The use of the DEAP dataset to predict emotions with high accuracy has been actively researched. Although the cross-validation method differs depending on previous studies, a previous study reported that binary classification accuracy achieved over 70% [22].

We introduce three necessary methods for highly accurate predictions using EEG. The first is removing the noise, a step which is critical for highly accurate emotion predictions from EEG. To improve its signal-to-noise ratio, removing the noise with component analysis (ICA) [23], [24] and machine learning [25] has been proposed. The second is calculating the feature values, for example, with EEG waveforms [26], [27] and power spectral density (PSD) [28]. The third is determining the learning model. Models such as support vector machine (SVM) [29] and k-nearest neighbor (KNN) [30] have been used. Other research uses a convolutional neural network (CNN), especially for image recognition [31], [32]. In addition, transfer learning has been used because the amount of EEG data is limited [33]–[35]. Although we introduced the training of models using only EEG, models using EEG and other information have also been studied. Highly accurate predictions are expected by utilizing a plurality of information. In previous studies, EEG

Manuscript received August 15, 2021.
Manuscript revised January 8, 2022.
Manuscript publicized February 15, 2022.

[†] The authors are with Nara Institute of Science and Technology, Ikoma-shi, 630–0192 Japan.
[††] The authors are with RIKEN, Center for Advanced Intelligence Project AIP, Ikoma-shi, 630–0192 Japan.
a) E-mail: miyamoto.kana.mk4@is.naist.jp (Corresponding author)
b) E-mail: hiroki-tan@is.naist.jp
c) E-mail: s-nakamura@is.naist.jp

DOI: 10.1587/transinf.2021EDP7171

and voice were used to predict speech quality [36]. EEG and face [37] and EEG and galvanic skin response [38] were used for emotion recognition. These three types of methods for highly accurate predictions using EEG must be chosen well depending on the task. For example, limited time can be spent on EEG processing in BCI and HCI.

Research that incorporates BCI and HCI into emotion induction is expected to contribute to the medical field [39], [40]. So far, proposals of emotion induction have played music based on emotions predicted from EEG. Two kinds of methods induce emotions: passive and active. The former is based on music therapy and requires no special effort other than participants who listen to music [41], [42]. In music therapy, a music therapist selects music after careful scrutiny of her patients. Such actions by a music therapist are replaced by emotion recognition from EEG. These recognized emotions are used to select appropriate music from a pre-prepared music database whose music is labeled as six different types of emotions. A previous study's system [42] selects music from a database that has the same label as the target emotion until the recognized emotion matches the target emotion. Although there are individual differences in felt emotions, generally this method can effectively induce emotions because its music is selected based on the states of the participants. However, note the following two concerns. First, the emotions that the method can recognize from EEG and induce are limited to six types. Even if emotions change while listening to music, they may not be reflected in the emotion recognition results because the emotions are treated as a group. Second, a music database is used. The amount of recorded music in it is limited, and sometimes its offerings might not be suitable for emotional induction.

In a method that actively induces emotions, the participants change their emotions to the target emotion on their own [43]. They perceive their own emotion after listening to music that expresses their emotion. They also strive to bring their perceived emotions closer to the target emotions. To generate music that reflects the emotions of participants, a logarithm of the EEG waveform variance is used as a feature value, and valence and arousal are predicted from linear discriminant analysis (LDA) and the sigmoid function as continuous values from 0 to 1. The proposed music generator in a previous study [43] automatically makes music that expresses the predicted valence and arousal. Since the continuous values of valence and arousal are used, music can be generated that seamlessly and continuously transitions between different emotional expression patterns. Using music that can effectively perceive one's emotions is fruitful for participants because they can know their own conditions. Unfortunately, the accuracy of emotion prediction using EEG might be insufficient because models of a previous study [43] are simple. Emotion prediction in an emotion induction system must be performed quickly, although prediction accuracy is also important in music generation. We expect that emotion prediction will be improved by devising a model.

Based on the above problems in previous studies on emotion induction, we propose a passive emotion induction system that automatically generates music based on the participants' emotions that are predicted with high accuracy from EEG. In this paper, we examine the following three elements for constructing our system:

1. Music generator: The amount of music in a music database is limited. We create a music generator that can make variety music. The music generator tries to induce emotions similar to its inputs. Since no music database is used, we are not limited by the available amount of music.
2. Emotion prediction using EEG: Classification models have difficulty detecting changes in emotions during emotion induction. We use regression models to predict detailed emotions. Simple regression models might not have sufficient performance in predicting emotions. We use CNN that takes into account the positional relationships of the electrodes to achieve highly accurate emotion prediction in a short time.
3. Control of music generator: The music generator creates music that induces emotions similar to the input emotions. However, there are individual differences in the emotions induced. We try to get target emotions by changing the inputs of the music generator based on the individual. We control the music generated by the music generator using emotions predicted by the EEG.

The remainder of this paper is outlined as follows. Section 2 describes our overall proposed system, and Sect. 3 describes the creation and evaluation of a music generator. Section 4 describes the training and validation of the emotion prediction model using EEG, and Sect. 5 describes the control of the music generator. In addition, we evaluate our proposed system that combines all three elements.

We developed an emotion induction system based on our previous works. This paper is an extended version of previously works [44], [45].

2. Design of Emotion Induction System

We construct an emotion induction system by combining a music generator, predicting emotions using EEG, and controlling the music generator. An outline of our proposed system is shown in Fig. 1.

We express valence and arousal as continuous values from 0 to 1. The system induces target emotions by the following procedure: 1) Input the target valence and arousal to the music generator, which produces music that induces emotions close to the input values; 2) Emotions are predicted from EEG while participants listen to generated music; 3) From the difference between the predicted and target emotions, we calculate the valence and arousal to be input to the music generator in a feedback loop; 4) Repeat steps 2) and 3). Even if emotions cannot be induced to the target values in 1), we approach the target emotion by changing the inputs to the music generator based on the predicted emotions. Our proposed system is expected to provide emotional

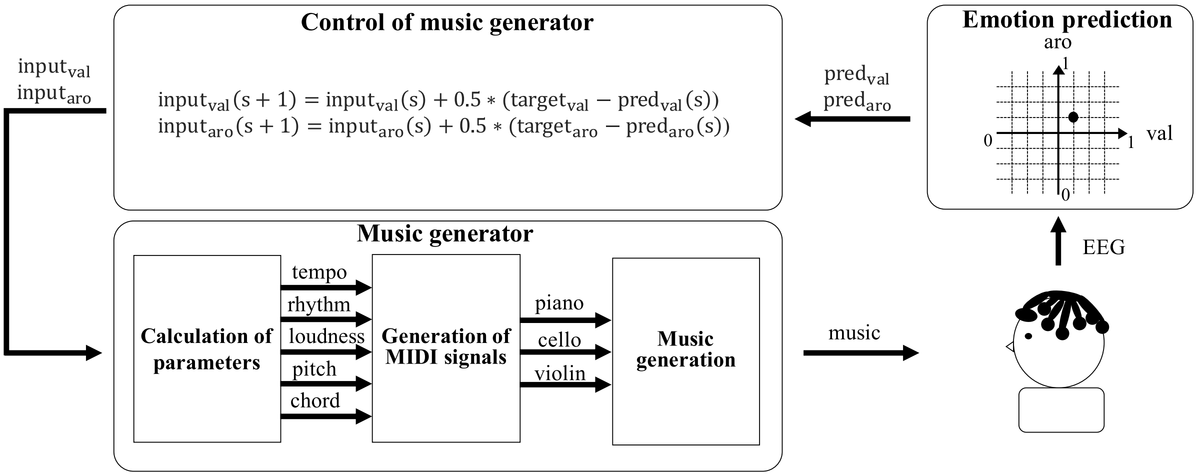

Control of music generator
input_val
input_aro
input_val (s + 1) = input_val (s) + 0.5 ∗ (target_val − pred_val (s))
input_aro (s + 1) = input_aro (s) + 0.5 ∗ (target_aro − pred_aro (s))
pred_val
pred_aro
Emotion prediction
aro
1
0
1
val
0
EEG
Music generator
Calculation of parameters
tempo
rhythm
loudness
pitch
chord
Generation of MIDI signals
piano
cello
violin
Music generation
music

Fig. 1 Emotion induction system: Input = input to music generator, target = emotion of target in induction, and pred = emotion predicted from EEG while listening to music.

induction that takes into account individual differences in emotions when listening to music.

3. Music Generator

We describe the creation and evaluation of a music generator for inducing emotions. Since perceived and felt emotions are similar when listening to music [10], [11], we create music based on a music generator for perceiving the intended emotions proposed in a previous study [43]. We reproduced the music generator of the previous study and improved it for inducing emotions.

3.1 Structure of Music Generator

The music generator proposed in a previous study makes music from the valence and arousal of an emotion to be perceived [43]. The music generator's valence and arousal inputs are represented as continuous values between 0 and 1. A valence closer to 1 indicates a pleasant state; an arousal closer to 1 indicates an activated state. We calculated five music parameters based on the formulas proposed by a previous study using the following formulas. We limited each bar to about 1 sec.

$$\text{tempo}: \quad note_{dur} = 0.3 - \text{aro} * 0.15 \subset \mathbb{R} \tag{1}$$

$$\text{rhythm}: \quad p(note = 1) = \text{aro} \tag{2}$$

$$\text{loudness}: \quad note_{vel} = unif\{50, 30 * \text{aro} + 60\} \subset \mathbb{N} \tag{3}$$

pitch :

$$note_{reg} = \begin{cases} p(\text{C3}) = 2 * (0.5 - \text{val}) & if \ \text{val} < 0.5 \\ p(\text{C5}) = 2 * (\text{val} - 0.5) & if \ \text{val} \geq 0.5 \\ \text{C4} & \text{otherwise} \end{cases} \tag{4}$$

$$\text{chord}: \quad 7 - (6 * \text{val}) \in 1, \ldots, 7 \subset \mathbb{N} \tag{5}$$

The tempo is the length of the note in seconds. A note is shorter if the arousal is high. The rhythm is the probability that the note will appear. A note's probability is higher if the arousal is high. However, in this formula, no music is generated when the arousal is 0. We set the arousal to 0.03 and calculated the rhythm when the arousal is less than 0.03. Loudness, which indicates the note's volume, increases when the arousal is high. The pitch indicates which scale is used. If the valence is high, a higher scale is more likely to be used. The chord determines the types of chords that are used in the music generation. The probability of using a higher scale increases if the valence is high. The chord determines which will be used in the music generation. A previous study used such modes as Ionian, Dorian, Phrygian, Lydian, Mixolydian, Aeolian, and Locrian [46]. In this experiment, the model was limited to Ionian. The chords used for each measure are shown in Table 1.

Table 1 Chords used for music generation

Chord	1 bar	2 bar	3 bar	4 bar
1	F_{maj}	B_{dim}	C_{maj}	F_{maj}
2	C_{maj}	F_{maj}	G_{maj}	C_{maj}
3	G_{maj}	C_{maj}	D_{min}	G_{maj}
4	D_{min}	G_{maj}	A_{min}	D_{min}
5	A_{min}	D_{min}	E_{min}	A_{min}
6	E_{min}	A_{min}	B_{dim}	E_{min}
7	B_{dim}	E_{min}	F_{maj}	B_{dim}

These five musical parameters were calculated by MATLAB (R2019a). The MIDI signals, which were also generated by MATLAB (R2019a), were sent to the DAW software Cakewalk using virtual MIDI cable software called LoopBe1. Music was generated by three virtual instruments: piano, violin, and cello. A sample can be heard here: https://sites.google.com/view/music-generator.

3.2 Assessment Methods

We evaluated our music generator with crowdworkers from CrowdWorks, Inc. The crowdworkers first listened to five samples (Each sample was 15 sec). The following are the inputs of the samples to the music generator: {val, aro}={0,0}; {0,1}; {0.5,0.5}; {1,0}; {1,1}. The crowdworkers then

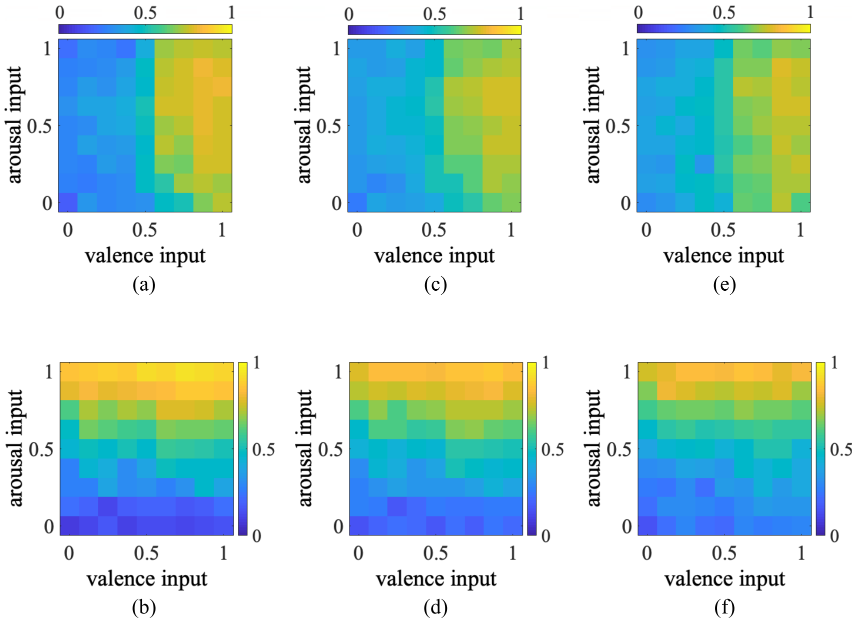


Fig. 2 Color map of perceived and felt emotions: (a) perceived valence evaluation; (b) perceived arousal evaluation; (c) felt valence evaluation; (d) felt arousal evaluation; (e) felt valence evaluation using recreated music generator; (f) felt arousal evaluation using recreated music generator.

evaluated the music by listening to 81 pieces (Each sample was 30 sec) that had the following combined total values: val = 0, 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875, 1, and aro = 0, 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875, 1. We evaluated the music with a Self-Assessment Mannequin (SAM), which evaluates the valence, arousal, and dominance of emotions using illustrations. In this study, the valence and the arousal of the emotions were evaluated on a 9-point scale from 0 to 1.

3.3 Evaluation of Music Generator

Our music generator is designed to help participants perceive the emotion intended by a music sample. The perceived emotions as intended by the music and those that are actually felt are the same, or the felt emotions are weaker than the perceived emotions [10], [11]. We investigated how effectively the music generator induced and perceived emotions.

3.3.1 Evaluation of Perceived Emotion

One hundred and one crowdworkers evaluated the emotions perceived by music. The evaluated valence and arousal were normalized to a minimum value of 0 and a maximum value of 1 for each crowdworker. Using the normalized valence and arousal, we used two methods to investigate whether the crowdworkers perceived the intended emotion of the music.

The first method calculated Pearson's linear correlation coefficient between the inputs of the music generator and the evaluations by each crowdworker. If the correlation coefficient is high, then the crowdworkers perceived the emotion intended by the music. The median of Pearson's linear correlation coefficient was r = 0.76 for valence and r = 0.86 for arousal. In a previous study that proposed the music generator [43], the median of the correlation coefficient was r = 0.52 for valence and r = 0.74 for arousal.

The second method represents the colored maps of the averages of the music generator's inputs and the crowdworker's evaluations. The color maps obtained from the experiments are shown in Fig. 2. The horizontal axis is the valence input to the music generator, and the vertical axis is the arousal input to the music generator. When the valence evaluated by the crowdworkers matches the valence input to the music generator, the colors are aligned vertically. When the arousal evaluated by the crowdworkers matches the arousal input to the music generator, the colors are aligned horizontally. The color map of valence in Fig. 2 (a) shows that a high valence is not easily perceived when the arousal input to the music generator is low. The color map of arousal in Fig. 2 (b) shows that the colors are aligned horizontally. Similar color maps were obtained in a previous study that proposed the music generator [43].

Since the above two methods showed similar results to those of a previous study that proposed the music generator [43], our music generator reproduced the music generator of the previous study.

3.3.2 Evaluation of Felt Emotion

One hundred and eight crowdworkers evaluated the emotions felt by our music. We used the same two methods that evaluated the emotions perceived by the music to investigate whether the crowdworkers felt emotions due to the music. The median of Pearson's linear correlation coefficient was r = 0.59 for valence and r = 0.83 for arousal. The color map of valence in Fig. 2 (c) shows that a high valence is not easily felt when the arousal input to the music generator is either low or high. The color map of arousal in Fig. 2 (d) shows that the colors are aligned horizontally.

The above two methods suggest that the music generator effectively induced arousal emotions. However, inducing valence emotions is affected by the arousal input to the music generator.

3.4 Improvement of the Music Generator Using Support Vector Regression

In our evaluation of the motions of crowdworkers felt by the music, our music generator struggled to induce high valence when the arousal input was low or high. Therefore, we trained a support vector regression model to regulate the arousal input to the music generator. This model predicted the arousal input to the music generator from the valence and arousal evaluated by the crowdworkers in 3.3.2. The support vector regression model was trained with 3-fold cross-validation using the default values in MATLAB and a Gaussian kernel. The result was RMSE = 0.0240. Since we thought that this support vector regression model could adjust the arousal input to the music generator, we trained it using all the data and connected it to the front of the music generator.

3.4.1 Evaluation of Improved Music Generator

One hundred and four crowdworkers evaluated the emotions felt by the music. The same two methods that evaluated the emotions perceived by the music were used to investigate whether the crowdworkers felt emotions due to the music. The median of the Pearson's linear correlation coefficient was r = 0.60 for valence and r = 0.76 for arousal. The color map of valence in Fig. 2 (e) does not show a tendency that high valence is hard to induce when the arousal input to the music generator is low or high. The color map of arousal in Fig. 2 (f) shows that the colors are aligned horizontally.

From the above two methods, the median of Pearson's linear correlation coefficient of the arousal of the music generator using support vector regression was lower than the original music generator. However, the tendency to induce high valence when the arousal input to the music generator was low or high was removed from the color map. This suggests that our proposed music generator using support vector regression effectively induced emotion. In the following experiments, music was created using the music generator based on support vector regression.

4. Emotion Prediction Using EEG

In this section, we describe emotion prediction using EEG. We investigated models that can achieve sufficiently accurate emotion prediction that is required for our proposed emotion induction system. First, we introduce EEG recordings and subjective evaluation of emotions felt while our participants listened to music. Next we describe the validation method of the emotion prediction models and show the results of comparing the models.

4.1 Data Collection

We recorded the EEG data and the subjective evaluations of the emotions the participants felt while listening to the music. The collected EEG data were preprocessed.

4.1.1 Participants

Twenty healthy participants (age: 24.3 years; 10 males, 10 females) joined this experiment, which was approved by the ethics committee of the Nara Institute of Science and Technology.

4.1.2 Stimuli

The music generator using support vector regression generated 41 pieces of music, each of which were 20 sec. The inputs to the music generator can be found shown in our previous work [44].

4.1.3 Experimental Protocol

At the beginning of the experiment, the participants sat in front of a desk with a monitor and wore earphones. They listened to five samples (Each sample was 15 sec) with the following input values to the music generator: {val,aro}={0,0};{0,1};{0.5,0.5};{1,0};{1,1}.

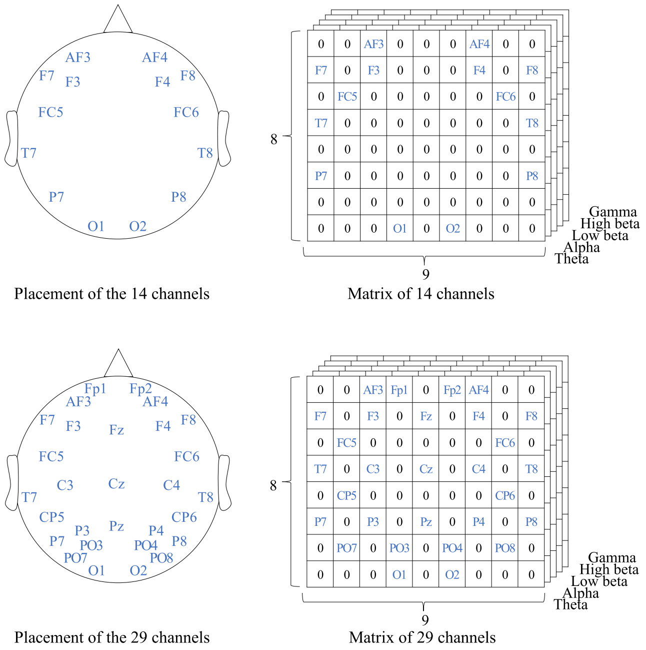
Then we conducted a practice session. The participants silently gazed at a cross mark in the center of the monitor for 5 sec and then listened to each 20-sec music sample while continuing to gaze at the cross mark. After listening to the music, they evaluated the valence and arousal of their felt emotions using SAM on a 9-point scale from 0 to 1. They practiced the experiment with two pieces of music: {val, aro}={0.125,0.25}; {0.875,0.75}.

After the practice, they put on an electroencephalograph, CGX Quick-30. We repeated the same procedure from the practice session to record the EEG data and the subjective evaluations of the felt emotions while they listened to the 41 pieces of music.

4.1.4 Preprocessing

We preprocessed each participant based on the following procedure using MATLAB (R2019a) and EEGLAB [47].

1) We removed the data that caused problems, such as no music was played; 2) The EEG signals were downsampled from 1000 to 200 Hz; 3) The silent states of 2 to 5 sec were divided into three epochs of 1-sec data; 4) The music-listening states of 0 to 20 sec were divided into 20 epochs of 1-sec data; 5) We designed 2nd order zero phase



Placement of the 14 channels

Matrix of 14 channels

0	0	AF3	0	0	0	AF4	0	0
F7	0	F3	0	0	0	F4	0	F8
0	FC5	0	0	0	0	0	FC6	0
T7	0	0	0	0	0	0	0	T8
0	0	0	0	0	0	0	0	0
P7	0	0	0	0	0	0	0	P8
0	0	0	0	0	0	0	0	0
0	0	0	O1	0	O2	0	0	0

Placement of the 29 channels

Matrix of 29 channels

0	0	AF3	Fp1	0	Fp2	AF4	0	0
F7	0	F3	0	Fz	0	F4	0	F8
0	FC5	0	0	0	0	0	FC6	0
T7	0	C3	0	Cz	0	C4	0	T8
0	CP5	0	0	0	0	0	CP6	0
P7	0	P3	0	Pz	0	P4	0	P8
0	PO7	0	PO3	0	PO4	0	PO8	0
0	0	0	O1	0	O2	0	0	0

Fig. 3 Matrix for CNN input

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 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 listening state for each type of music.

From the preprocessing, we obtained 145 dimensional features of all 29 channels of the electroencephalograph × 5 frequency bands for 20 samples per piece of music.

4.2 Models

We evaluated four different regression models for predicting the felt emotions using EEG while listening to music. The baseline model is linear regression, which is a similar model used in a previous study of emotion induction using EEG [43]. We proposed CNN that can take into account the information of the channel positions in the EEG. Three types of models were proposed and compared with the baseline model: CNN, CNN with transfer learning, and neural network with emotions predicted from CNN and the inputs of the music generator.

4.2.1 Linear Regression

A previous study of active emotion induction showed the prediction of emotions based on LDA and the sigmoid function [43]. In this experiment, we selected linear regression as a similar model to the previous study's model [43]. The training data were normalized by Z-scores.

4.2.2 CNN

CNN, which is effective for feature extraction, is widely used in such fields as image recognition as well as in the study of EEG [33], [48]. By using matrices that take into account the positions of the channels, we expect that training will consider the relationships between channels. In this experiment, we created the matrices shown in Fig. 3 based on the channel locations and five frequency bands from previous studies [49], [50]. The size of the matrix is 8 × 9 × 5. We used zero to fill the elements without any EEG channels.

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 with 100 learning epochs.

4.2.3 CNN with Transfer Learning

The transfer learning method updates a model learned in a domain using data from another domain [51]. Since it is possible to update a model with a small amount of data, it is being applied to research related to EEG. However, image recognition models that do not use EEG data for training are often used [33]. In this paper, we train a model of a different domain using different EEG datasets from the present experiment and re-train it using the EEG data recorded in our experiment. To train the model in a different domain, we use the DEAP dataset [21], which is comprised of EEG data and subjective evaluations of valence and arousal gathered while watching music videos. We used the DEAP dataset for pretraining because both it and our experiment include music, although the stimuli in our experiment do not include videos. The DEAP dataset has 32 participants and 1-min music videos. The sampling frequency of the EEG data is 128 Hz, and we applied a band-pass filter from 4 Hz to 45 Hz. We preprocessed this DEAP dataset as follows.

1) The silent state of 0 to 3 sec was divided into three epochs of 1-sec data; 2) The music-listening state of 0 to 60 sec was divided into 60 epochs of 1-sec data; 3) 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); 4) EEG signals were divided into five frequency bands by the designed filters to calculate the $f = log(var(EEGdata))$, which is the logarithm of the waveform variance for each bit of data; 5) The average of the logarithm of the silent state waveform variance was subtracted from the logarithmic waveform variance during the listening state for each type of music.

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

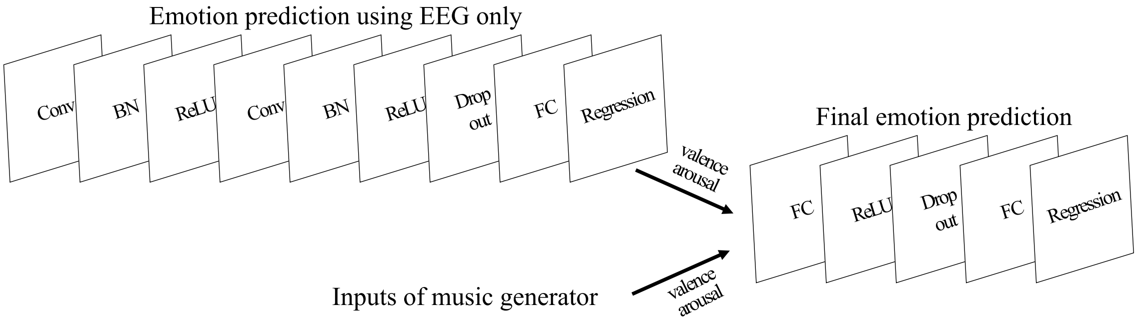


Fig. 4 Structure of neural network using CNN and inputs of music generator

layer, a ReLU layer, a dropout layer (dropout rate of 0.5), a fully connected layer (output dimensionality of 8), a batch normalization layer, 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 first trained the model using the DEAP dataset. After that, we retrained it in the last two layers of the CNN for each participant using our experimental data. We used Adam optimization in the model's pretraining and retraining. The learning rate was 0.001, the batch size was 128, and the learning epoch was 5 when pretraining the model. The learning rate was 0.0005, the batch size was 64, and the learning epoch was 100 when retraining the model.

4.2.4 Neural Network Using CNN and Inputs of Music Generator

If the predicted emotions are significantly different from the emotions felt by the participants, our proposed system might generate music that is unsuitable for emotion induction. Therefore, we propose an emotion prediction method that adds information other than EEG. In this study, we need to use a minimum amount of information and training models because rapid emotion prediction is required. Thus, we used the valence and arousal of the inputs to the music generator. Section 3 showed that the generated music induced emotions similar to the inputs to the music generator. This suggests that these inputs are the predicted emotions to be induced. We aimed to improve the accuracy of the emotion prediction using the emotion predicted from EEG and the inputs to the music generator.

We predicted the valence and the arousal from EEG using the CNN of Sect. 4.2.2. Then the final valence and arousal were predicted using a neural network from four inputs: the valence and arousal predicted from the CNN, and the valence and arousal that were input to the music generator.

The neural network is comprised 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. The model's structure is shown in Fig. 4.

4.3 Result

EEG characteristics vary among participants [52]. To account for such individuality, we trained models for each participant. To compare the models, we first investigated the cross-validation method. We compared the RMSE of the training data and the test data with and without samples from the same music. Next we examined the number of EEG channels used for training and compared the model trained from all 29 channels with the model trained from the selected 14 channels. Finally, we used these results to determine how the models were compared and examined the model with the lowest RMSE among the four trained models. The models were trained using MATLAB (2019b).

4.3.1 Investigation of Cross-Validation Method

We acquired 20 samples from each piece of music in this experiment because the length of the music was 20 sec. A previous study [43] also acquired multiple samples from one piece of music and adopted cross-validation without considering the type of music to train models. However, emotions are predicted using EEG when listening to unknown music in the emotion induction system. Cross-validation without considering the type of music is a different type of verification from the situation where we use the emotion induction system in real life. In addition, the training and test data contain samples obtained from EEG while listening to the same music with similar features. The RMSE might be lower when these data are used than when the system will be used. Therefore, we propose a leave-one-music-out cross-validation that takes into account the type of music. It uses 20 samples of a piece of music as the test data and the rest of the samples as the training data. We compared the leave-one-music-out cross-validation with k-fold cross-validation using 20 samples as the test data without considering the type of music and investigated how much the RMSE differs.

The models used for validation were linear regression and CNN. We trained them using the 14 channels of features used in a previous study [43]: AF3, AF4, F3, F4, F7, F8, FC5, FC6, T7, T8, P7, P8, O1, and O2. The RMSE of the emotions predicted by the trained model and the measured subjective evaluations obtained in the experiment were

Table 2 RMSE and R^2 of felt and predicted emotions applying leave-one-music-out (LOMO) cross-validation and k-fold cross-validation

Par.	RMSE							
	Linear regression				CNN			
	LOMO		k-fold		LOMO		k-fold	
	val	aro	val	aro	val	aro	val	aro
1	0.310	0.379	0.222	0.255	0.290	0.321	0.242	0.270
2	0.324	0.262	0.239	0.193	0.308	0.247	0.257	0.212
3	0.302	0.310	0.221	0.228	0.286	0.267	0.248	0.236
4	0.130	0.179	0.097	0.130	0.119	0.159	0.107	0.141
5	0.354	0.316	0.261	0.234	0.342	0.289	0.288	0.243
6	0.268	0.328	0.196	0.239	0.253	0.310	0.213	0.253
7	0.431	0.445	0.309	0.321	0.381	0.401	0.333	0.342
8	0.288	0.328	0.206	0.239	0.243	0.300	0.211	0.253
9	0.243	0.319	0.173	0.230	0.222	0.287	0.190	0.239
10	0.080	0.188	0.058	0.137	0.081	0.174	0.069	0.151
11	0.179	0.252	0.131	0.184	0.172	0.240	0.142	0.203
12	0.239	0.309	0.174	0.226	0.217	0.277	0.189	0.241
13	0.245	0.341	0.175	0.253	0.237	0.326	0.190	0.272
14	0.210	0.254	0.154	0.187	0.193	0.226	0.168	0.196
15	0.279	0.269	0.200	0.199	0.250	0.246	0.210	0.211
16	0.100	0.185	0.076	0.139	0.102	0.182	0.090	0.156
17	0.492	0.410	0.360	0.299	0.434	0.398	0.376	0.323
18	0.192	0.209	0.134	0.139	0.174	0.169	0.146	0.145
19	0.093	0.237	0.067	0.174	0.081	0.211	0.073	0.184
20	0.419	0.400	0.310	0.289	0.397	0.338	0.330	0.296
Mean	0.259	0.296	0.188	0.215	0.239	0.268	0.204	0.228
Std	0.113	0.077	0.083	0.055	0.102	0.070	0.087	0.057

Par.	R^2							
	Linear regression				CNN			
	LOMO		k-fold		LOMO		k-fold	
	val	aro	val	aro	val	aro	val	aro
1	-0.208	-0.671	0.380	0.243	-0.060	-0.199	0.263	0.151
2	-0.271	-0.365	0.308	0.263	-0.147	-0.207	0.199	0.108
3	-0.337	-0.498	0.285	0.190	-0.201	-0.113	0.099	0.134
4	-0.368	-0.466	0.242	0.225	-0.139	-0.163	0.078	0.093
5	-0.230	-0.346	0.331	0.263	-0.147	-0.130	0.189	0.204
6	-0.275	-0.183	0.318	0.371	-0.135	-0.056	0.196	0.295
7	-0.376	-0.334	0.291	0.306	-0.080	-0.084	0.177	0.212
8	-0.636	-0.345	0.165	0.283	-0.171	-0.127	0.119	0.198
9	-0.382	-0.441	0.298	0.255	-0.156	-0.163	0.154	0.194
10	-0.348	-0.322	0.295	0.299	-0.369	-0.130	0.013	0.144
11	-0.281	-0.294	0.321	0.311	-0.186	-0.181	0.201	0.158
12	-0.488	-0.482	0.216	0.208	-0.226	-0.197	0.067	0.098
13	-0.217	-0.262	0.380	0.307	-0.141	-0.155	0.266	0.196
14	-0.448	-0.331	0.221	0.282	-0.218	-0.052	0.076	0.213
15	-0.419	-0.410	0.269	0.229	-0.136	-0.178	0.194	0.134
16	-0.121	-0.222	0.355	0.317	-0.165	-0.172	0.101	0.139
17	-0.391	-0.336	0.256	0.290	-0.085	-0.259	0.188	0.172
18	-0.391	-0.537	0.320	0.315	-0.139	-0.010	0.196	0.261
19	-0.615	-0.272	0.168	0.310	-0.240	-0.012	0.007	0.236
20	-0.299	-0.567	0.292	0.182	-0.162	-0.118	0.194	0.142
Mean	-0.355	-0.384	0.285	0.273	-0.165	-0.135	0.149	0.174
Std	0.129	0.123	0.060	0.048	0.067	0.066	0.075	0.054

Table 3 RMSE and R^2 of felt and predicted emotions using models trained on different channels. Bold indicates lowest RMSE.

Par.	RMSE							
	Linear regression				CNN			
	14 channels		29 channels		14 channels		29 channels	
	val	aro	val	aro	val	aro	val	aro
1	0.310	0.379	0.314	0.382	0.290	0.321	0.288	0.317
2	0.324	0.262	0.347	0.271	0.308	0.247	0.313	0.258
3	0.302	0.310	0.318	0.299	0.286	0.267	0.289	0.275
4	0.130	0.179	0.146	0.210	0.119	0.159	0.129	0.169
5	0.354	0.316	0.369	0.349	0.342	0.289	0.350	0.293
6	0.268	0.328	0.286	0.346	0.253	0.310	0.268	0.316
7	0.431	0.445	0.455	0.453	0.381	0.401	0.390	0.409
8	0.288	0.328	0.313	0.350	0.243	0.300	0.256	0.315
9	0.243	0.319	0.248	0.327	0.222	0.287	0.228	0.295
10	0.080	0.188	0.092	0.216	0.081	0.174	0.082	0.176
11	0.179	0.252	0.186	0.257	0.172	0.240	0.180	0.248
12	0.239	0.309	0.227	0.323	0.217	0.277	0.218	0.274
13	0.245	0.341	0.247	0.356	0.237	0.326	0.246	0.336
14	0.210	0.254	0.210	0.279	0.193	0.226	0.193	0.231
15	0.279	0.269	0.276	0.259	0.250	0.246	0.253	0.239
16	0.100	0.185	0.110	0.204	0.102	0.182	0.118	0.192
17	0.492	0.410	0.527	0.436	0.434	0.398	0.446	0.404
18	0.192	0.209	0.194	0.210	0.174	0.169	0.175	0.177
19	0.093	0.237	0.095	0.250	0.081	0.211	0.089	0.219
20	0.419	0.400	0.456	0.418	0.397	0.338	0.395	0.362
Mean	**0.259**	**0.296**	0.271	0.310	**0.239**	**0.268**	0.245	0.275
Std	0.113	0.077	0.121	0.077	0.102	0.070	0.102	0.071

Par.	R^2							
	Linear regression				CNN			
	14 channels		29 channels		14 channels		29 channels	
	val	aro	val	aro	val	aro	val	aro
1	-0.208	-0.671	-0.239	-0.694	-0.060	-0.199	-0.042	-0.167
2	-0.271	-0.365	-0.461	-0.450	-0.147	-0.207	-0.186	-0.316
3	-0.337	-0.498	-0.478	-0.390	-0.201	-0.113	-0.221	-0.175
4	-0.368	-0.466	-0.707	-1.017	-0.139	-0.163	-0.346	-0.309
5	-0.230	-0.346	-0.337	-0.641	-0.147	-0.130	-0.204	-0.156
6	-0.275	-0.183	-0.454	-0.311	-0.135	-0.056	-0.271	-0.098
7	-0.376	-0.334	-0.537	-0.384	-0.080	-0.084	-0.127	-0.130
8	-0.636	-0.345	-0.942	-0.530	-0.171	-0.127	-0.299	-0.244
9	-0.382	-0.441	-0.442	-0.512	-0.156	-0.163	-0.221	-0.227
10	-0.348	-0.322	-0.770	-0.745	-0.369	-0.130	-0.423	-0.152
11	-0.281	-0.294	-0.374	-0.349	-0.186	-0.181	-0.298	-0.253
12	-0.488	-0.482	-0.346	-0.622	-0.226	-0.197	-0.241	-0.167
13	-0.217	-0.262	-0.240	-0.377	-0.141	-0.155	-0.228	-0.222
14	-0.448	-0.331	-0.442	-0.600	-0.218	-0.052	-0.217	-0.102
15	-0.419	-0.410	-0.389	-0.305	-0.136	-0.178	-0.165	-0.118
16	-0.121	-0.222	-0.348	-0.474	-0.165	-0.172	-0.552	-0.307
17	-0.391	-0.336	-0.597	-0.514	-0.085	-0.259	-0.147	-0.300
18	-0.391	-0.537	-0.420	-0.550	-0.139	-0.010	-0.161	-0.104
19	-0.615	-0.272	-0.679	-0.421	-0.240	-0.012	-0.495	-0.084
20	-0.299	-0.567	-0.535	-0.706	-0.162	-0.118	-0.153	-0.284
Mean	-0.355	-0.384	-0.487	-0.530	-0.165	-0.135	-0.250	-0.196
Std	0.129	0.123	0.179	0.175	0.067	0.066	0.125	0.080

calculated, as shown in Table 2. The average RMSE of the participants was lower for k-fold cross-validation in both linear regression and CNN. There was also a significant difference between k-fold and leave-one-music-out cross-validations for both linear regression and CNN from the Wilcoxon signed rank test ($p < 0.05$). We also obtained R^2. The average R^2 of the participants was higher for k-fold cross-validation in both linear regression and CNN.

We found that the k-fold cross-validation significantly lowered the RMSE more than the leave-one-music-out cross-validation. In the following experiments, we used leave-one-music-out cross-validation to investigate whether the model can predict emotions from EEG when listening to unknown music not included in training data.

4.3.2 Effect of the Number of Channels Used for Training

A previous study [43] used an EPOC of an electroencephalograph manufactured by Emotiv that can measure 14 channels of EEG: AF3, AF4, F3, F4, F7, F8, FC5, FC6, T7, T8, P7, P8, O1, and O2. We use a Quick-30 manufactured by CGX. This electroencephalograph can measure 29 EEG channels: 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 investigated whether models trained with 14 or 29 channels would lower the RMSE. We used linear regression and CNN.

The RMSE values of the emotions predicted by the trained model and the measured subjective evaluations obtained in the experiment were calculated, as shown in Table 3. The average RMSE of the participants was lower for the model trained with 14 channels in both the linear regression and CNN. We found a significant difference between the models with 14 and 29 channels for both linear regression and CNN from the Wilcoxon signed rank test ($p < 0.05$). We also obtained R^2. The average R^2 of the participants was higher for the model trained with 14 channels in both the linear regression and CNN.

Table 4 RMSE and R^2 of felt and predicted emotions using four models. Bold indicates lowest RMSE. TL is transfer learning and Music gen. is inputs of music generator.

Par.	RMSE									
	Linear		CNN		CNN+TL		CNN+Music gen.		Music gen.	
	val	aro	val	aro	val	aro	val	aro	val	aro
1	0.310	0.379	0.290	0.321	0.281	0.308	0.155	0.178	0.168	0.180
2	0.324	0.262	0.308	0.247	0.301	0.234	0.212	0.170	0.223	0.292
3	0.302	0.310	0.286	0.267	0.275	0.258	0.163	0.159	0.185	0.199
4	0.130	0.179	0.119	0.159	0.115	0.152	0.117	0.150	0.325	0.360
5	0.354	0.316	0.342	0.289	0.341	0.283	0.233	0.178	0.240	0.219
6	0.268	0.328	0.253	0.310	0.252	0.297	0.253	0.197	0.377	0.172
7	0.431	0.445	0.381	0.401	0.379	0.394	0.168	0.222	0.161	0.204
8	0.288	0.328	0.243	0.300	0.237	0.293	0.256	0.212	0.466	0.301
9	0.243	0.319	0.222	0.287	0.220	0.278	0.190	0.151	0.296	0.168
10	0.080	0.188	0.081	0.174	0.071	0.171	0.064	0.148	0.314	0.422
11	0.179	0.252	0.172	0.240	0.165	0.233	0.134	0.176	0.282	0.225
12	0.239	0.309	0.217	0.277	0.205	0.275	0.191	0.245	0.369	0.339
13	0.245	0.341	0.237	0.326	0.231	0.318	0.154	0.245	0.210	0.258
14	0.210	0.254	0.193	0.226	0.185	0.223	0.189	0.155	0.338	0.304
15	0.279	0.269	0.250	0.246	0.245	0.238	0.171	0.211	0.245	0.386
16	0.100	0.185	0.102	0.182	0.097	0.177	0.100	0.162	0.338	0.290
17	0.492	0.410	0.434	0.398	0.432	0.381	0.187	0.209	0.207	0.190
18	0.192	0.209	0.174	0.169	0.170	0.173	0.125	0.163	0.284	0.318
19	0.093	0.237	0.081	0.211	0.077	0.206	0.076	0.150	0.295	0.256
20	0.419	0.400	0.397	0.338	0.397	0.333	0.402	0.311	0.436	0.398
Mean	0.259	0.296	0.239	0.268	0.234	0.261	**0.177**	**0.189**	0.288	0.274
Std	0.113	0.077	0.102	0.070	0.104	0.068	0.075	0.042	0.085	0.079

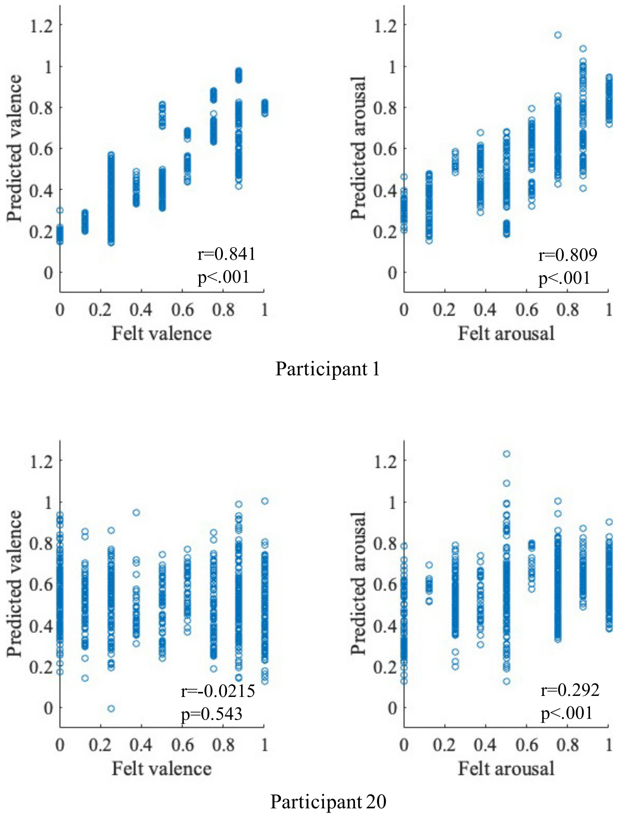
Par.	R^2									
	Linear		CNN		CNN+TL		CNN+Music gen.		Music gen.	
	val	aro	val	aro	val	aro	val	aro	val	aro
1	-0.208	-0.671	-0.060	-0.199	0.005	-0.101	0.697	0.632	0.645	0.624
2	-0.271	-0.365	-0.147	-0.207	-0.096	-0.089	0.457	0.429	0.395	-0.684
3	-0.337	-0.498	-0.201	-0.113	-0.109	-0.037	0.610	0.609	0.498	0.383
4	-0.368	-0.466	-0.139	-0.163	-0.069	-0.056	-0.106	-0.034	-7.491	-4.960
5	-0.230	-0.346	-0.147	-0.130	-0.141	-0.081	0.467	0.571	0.436	0.352
6	-0.275	-0.183	-0.135	-0.056	-0.121	0.033	-0.134	0.574	-1.512	0.674
7	-0.376	-0.334	-0.080	-0.084	-0.063	-0.049	0.789	0.667	0.808	0.720
8	-0.636	-0.345	-0.171	-0.127	-0.106	-0.075	-0.297	0.436	-3.293	-0.134
9	-0.382	-0.441	-0.156	-0.163	-0.136	-0.091	0.151	0.679	-1.054	0.602
10	-0.348	-0.322	-0.369	-0.130	-0.049	-0.089	0.129	0.185	-19.660	-5.667
11	-0.281	-0.294	-0.186	-0.181	-0.087	-0.109	0.285	0.370	-2.161	-0.036
12	-0.488	-0.482	-0.226	-0.197	-0.090	-0.178	0.053	0.070	-2.549	-0.783
13	-0.217	-0.262	-0.141	-0.155	-0.084	-0.095	0.522	0.350	0.105	0.281
14	-0.448	-0.331	-0.218	-0.052	-0.125	-0.024	-0.167	0.509	-2.726	-0.897
15	-0.419	-0.410	-0.136	-0.178	-0.094	-0.103	0.465	0.133	-0.097	-1.914
16	-0.121	-0.222	-0.165	-0.172	-0.038	-0.118	-0.103	0.069	-11.707	-1.983
17	-0.391	-0.336	-0.085	-0.259	-0.076	-0.158	0.799	0.652	0.754	0.712
18	-0.391	-0.537	-0.139	-0.010	-0.094	-0.052	0.411	0.063	-2.046	-2.571
19	-0.615	-0.272	-0.240	-0.012	-0.112	0.034	-0.093	0.489	-15.300	-0.485
20	-0.299	-0.567	-0.162	-0.118	-0.166	-0.086	-0.193	0.055	-0.404	-0.547
Mean	-0.355	-0.384	-0.165	-0.135	-0.093	-0.076	0.237	0.375	-3.318	-0.816
Std	0.129	0.123	0.067	0.066	0.039	0.052	0.356	0.245	5.772	1.816

The model trained with 14 channels significantly lowered the RMSE more than the model trained with 29 channels. In the following experiments, we trained with the selected 14 EEG channels.

4.3.3 Comparison of the Four Models

From our investigation of the cross-validation method and the number of EEG channels, four different models were trained using 14 channels and verified by leave-one-music-out cross-validation.

The RMSE values of the emotions predicted by the trained model and the measured subjective evaluations obtained in the experiment were calculated, as shown in Table 4. The RMSE values of the inputs of the music generator and the measured subjective evaluations obtained were calculated, as shown in Music gen. of Table 4. The average RMSE of the participants was lowest for the neural network using CNN and the inputs of the music generator. There was also a significant difference between the neural network and the linear regression of the baseline model from the Wilcoxon signed rank test ($p < 0.05$). We also found significant differences between CNN and linear regression, CNN using transfer learning and linear regression,



Participant 1

Participant 20

Fig. 5 Plots and Pearson's correlation coefficient of felt and predicted emotions by CNN+Music.gen of participant 1 and 20

CNN using transfer learning and CNN, the neural network and CNN, and the neural network and CNN using transfer learning ($p < 0.05$). We also obtained R^2. The average R^2 of the participants was higher for the neural network with CNN and inputs of music generator.

In summary, among the four models (linear regression, CNN, CNN with transfer learning, and neural network with CNN and inputs of music generator), the model with the lowest RMSE was the neural network with CNN and the music generator's inputs. The three proposed models had significantly lower RMSE values than linear regression. The plots of felt and predicted emotions by the neural network were as shown in Fig. 5. The most of participants' plots were similar to participant 1, but there were some participants with large residuals, such as participant 20.

5. Control of Music Generator

In this section, we describe the proposal and validation of the formulas for controlling the music generator using emotions predicted from EEG. We describe the data collection to verify the system and discuss the validation of its usefulness.

5.1 Data Collection

In this experiment, we constructed an emotion induction system by combining a music generator using support vector regression (Sect. 3) and a neural network-based emotion prediction using CNN and the music generator's inputs (Sect. 4). This system automatically modulates the music based on the emotions predicted from the EEG. We investigated whether the music that changes according to the participants' emotions or the music that does not change is more effective for inducing emotions.

5.1.1 Participants

Six healthy participants (age: 27.2 years; 3 males, 3 females) joined this experiment. They also participated in the experiments in Sect. 4. The study was approved by the ethics committee of the Nara Institute of Science and Technology.

5.1.2 Models for Emotion Prediction

The model for emotion prediction used in the emotion induction system was trained on the neural network using CNN and the music generator's input in Sect. 4. It was trained for each participant. However, there was no test data as in Sect. 4, and the model was trained using all the available data.

5.1.3 Stimuli

The music generator with support vector regression created music of 20 bars. The target emotions are the nine types of {val, aro} = {0.125,0.125}; {0.125,0.5}; {0.125,0.875}; {0.5,0.125}; {0.5,0.5}; {0.5,0.875}; {0.875,0.125}; {0.875,0.5}; {0.875,0.875}.

We prepared two methods for generating music to induce emotions. The music generation methods are outlined in Fig. 6. The baseline method generated constant music without using the predicted emotions by inputting the target emotion only once to the music generator. This is the same generation music as the stimulus used in Sects. 3 and 4. Although the method doesn't use the emotion predicted from EEG, a 1-sec EEG data is measured every four bars for predicting emotions.

The second method dynamically generated music using the predicted emotions. An emotion is predicted from the EEG while the participants listen to the generated music, and the inputs to the music generator are changed to get closer to the target emotion. A 1-sec EEG data is measured every four bars and used to predict the emotion. Using the predicted and target emotions, the inputs are changed every four bars. The following equations updated the inputs. In the equations, s represents the number of times the inputs are updated, and $s = 0$ denotes when the target emotion is input directly to the music generator. Updates were made up to $s = 4$. When s is 1 or more, we added half of the value of the difference between the target and the predicted values to the music generator's input of the previous loop for each valence and arousal. In Sect. 3, it was shown that

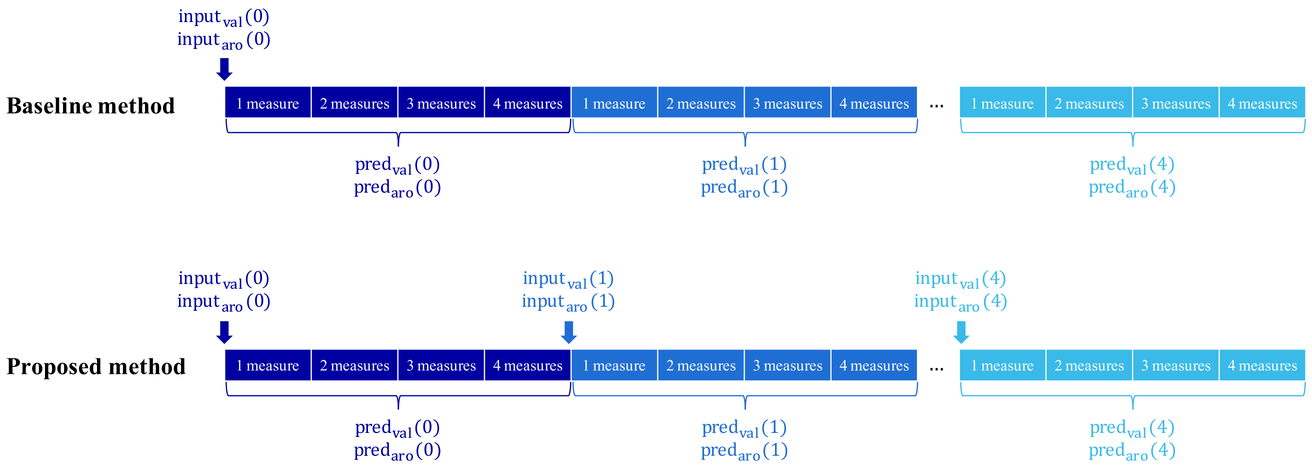


Fig. 6 Music-generation method

there were correlations between the inputs of the music generator and the participants' emotions. Therefore, we tried to increase or decrease the music generator's inputs according to the participants' emotions. Due to its input limitations, the input of 0 or less is set to 0, and the input of 1 or more is set to 1. We used half of the difference between the target and the predicted values to avoid the input from exceeding the upper or lower limit and promote change in the input value. Because this method uses predicted emotions, it can generate music that is personalized to each participant.

$$input_{\mathrm{val}}(s+1) = input_{\mathrm{val}}(s) + 0.5 * (target_{\mathrm{val}} - pred_{\mathrm{val}}(s)) \quad (6)$$

$$input_{\mathrm{aro}}(s+1) = input_{\mathrm{aro}}(s) + 0.5 * (target_{\mathrm{aro}} - pred_{\mathrm{aro}}(s)) \quad (7)$$

5.1.4 Experimental Protocol

Our experimental protocol was based on the same method shown in Sect. 4. However, the participants evaluated the emotions on a 17-point scale from 0 to 1 before and after listening to music with SAM to observe small changes in emotions. After listening to the music every time, they took a 10-sec break. For recording, MATLAB (R2019b) and Lab Streaming Layer were used for EEG analysis and music generation.

5.2 Result

We verified our emotion induction system based on a previous study [43] by comparing target and induced emotions. The induced emotion is the last emotion predicted while listening to music. The distance between the target and induced emotions was calculated by the following formula:

$$distance = \sqrt{(target_{\mathrm{val}} - pred_{\mathrm{val}})^2 + (target_{\mathrm{aro}} - pred_{\mathrm{aro}})^2} \quad (8)$$

Table 5 shows the RMSE and R^2 of emotion prediction and average distance in nine types of emotional induction. The RMSE and R^2 were calculated using the subjective evaluation of the emotion after listening to music and the last predicted emotion while listening to music. The average distance of the participants was shorter in the proposed method than in the baseline method. There was a significant difference between both methods in the distance from the Wilcoxon signed rank test ($p < 0.05$).

Table 5 RMSE and R^2 of emotion prediction and distance between target and induced emotions. Bold indicates lowest distance.

Par.	Distance		Errors of emotion prediction			
	Baseline	Proposed	RMSE		R^2	
			val	aro	val	aro
1	0.434	0.422	0.245	0.287	-0.238	-0.211
2	0.154	0.087	0.206	0.160	0.234	0.691
3	0.227	0.207	0.175	0.171	-0.314	0.369
4	0.388	0.348	0.326	0.182	-1.216	0.076
5	0.335	0.283	0.105	0.100	0.302	0.468
6	0.240	0.141	0.147	0.179	-1.329	-1.084
Mean	0.296	**0.248**	0.201	0.180	-0.427	0.051
Std	0.107	0.127	0.078	0.061	0.700	0.639

6. Conclusion and Future Work

This paper provided personalized emotion induction by combining a music generator, emotion predictions using EEG, and control of the music generator. We conducted three experiments: creation of a music generator, emotion prediction using EEG while listening to generated music, and evaluation of the emotion induction system to control the inputs to the music generator.

In the first experiment, we reproduced the music generator for perceiving the intended emotions proposed in a previous study. Since the felt emotions resemble the perceived emotions from listening to music, we investigated whether the music generator effectively induced emotions. We found that the input of the arousal to the music generator hindered the induction of valence in our experiment. We then trained a support vector regression model that adjusted the input of the arousal to the music generator. After connecting the trained model to the front of the music generator, it became capable of inducing both valence and arousal. The parameters of the music generation method were arbitrarily determined in the experiment. Their verification will be cited in future work.

In the second experiment, emotions were predicted from EEG while listening to music made by our music generator. After training four types of regression models using 14 EEG channels and verifying them by leave-one-music-out cross-validation, the model using CNN had significantly lower RMSE than the linear regression of the baseline model. Training the positional relationship of the electrodes is effective for emotion prediction. Of the four models, the neural network using CNN and music generator inputs had the lowest RMSE. Even if the emotions predicted from the CNN or the emotions of inputs to the music generator are significantly different from the emotions that are actually felt, it was possible to predict emotions that are close to the emotions actually felt using either emotion. The emotion prediction model for each participant was trained in the experiment. However, since this has a high learning cost, we believe that participants will need an independently trained emotion prediction model in the future.

In the third experiment, we proposed formulas for controlling the music generator and constructed an emotion induction system that combines the music generator and emotion prediction from EEG. As a result of the distance between the target and predicted emotions after emotion induction, there was a significant difference between the baseline method of generating music without considering the predicted emotions and the proposed method of generating music based on them. This suggests that a feedback loop that generates personalized music can effectively induce emotions. However, since our experiment only had six participants, we must increase the number of participants in the future. The performance of emotion prediction models influences the music generation and calculation of the distance between the target and predicted emotions. We obtained the RMSE of emotion prediction (valence: 0.201, arousal: 0.180). The system might have generated music that was harder to get close to the target emotions by the RMSE. It also affects the reliability of the distance between the target and the predicted emotions. We believe that we can provide more effective emotion induction when RMSE is closer to zero, in such a case the music generation and calculation of the distance can be much more accurate. Finding models with smaller RMSE is one of our future works. We trained the emotion prediction models from the EEG data recorded on a different day than the day the system was used. The position of channels and the characteristics of the EEG might be different between these days. For high performance of emotion prediction, the EEG features during training and using the model should be similar. Recording the EEG data and training the models just before using the system is one of the solutions for this problem. Although the lack of EEG data affects the training of the models, it is a burden for participants to record EEG data for a long time before using the system. Therefore, we are working on emotion prediction with high performance using a small amount of EEG data [53]. Furthermore, we arbitrarily determined the method and interval for updating the music parameters used in the proposed method. In the future, we will investigate the changes in emotions when using the emotion induction system and verify its parameters. The analysis of emotions before and after listening to music is also expected to deepen research on emotion induction.

Acknowledgments

This work was supported by JST CREST Grant Number JPMJCR19A5, Japan, and RIKEN Junior Research Associate Program.

References

[1] P. Ekman, "An argument for basic emotions," Cognition & emotion, vol.6, no.3-4, pp.169–200, 1992.

[2] A.M. Kring and A.H. Gordon, "Sex differences in emotion: expression, experience, and physiology.," Journal of personality and social psychology, vol.74, no.3, pp.686–703, 1998.

[3] S.D. Kreibig, "Autonomic nervous system activity in emotion: A review," Biological psychology, vol.84, no.3, pp.394–421, 2010.

[4] B.L. Fredrickson and C. Branigan, "Positive emotions broaden the scope of attention and thought-action repertoires," Cognition & emotion, vol.19, no.3, pp.313–332, 2005.

[5] V. Santos, F. Paes, V. Pereira, O. Arias-Carrión, A.C. Silva, M.G. Carta, A.E. Nardi, and S. Machado, "The role of positive emotion and contributions of positive psychology in depression treatment: systematic review.," Clinical practice and epidemiology in mental health, vol.9, pp.221–237, 2013.

[6] P.N. Juslin and D. Vastfjall, "Emotional responses to music: The need to consider underlying mechanisms," Behavioral and brain sciences, vol.31, no.5, pp.559–575, 2008.

[7] G. Kreutz, U. Ott, D. Teichmann, P. Osawa, and D. Vaitl, "Using music to induce emotions: Influences of musical preference and absorption," Psychology of music, vol.36, no.1, pp.101–126, 2008.

[8] P. Gomez and B. Danuser, "Relationships between musical structure and psychophysiological measures of emotion.," Emotion, vol.7, no.2, pp.377–387, 2007.

[9] T. Eerola, A. Friberg, and R. Bresin, "Emotional expression in music: contribution, linearity, and additivity of primary musical cues," Frontiers in psychology, vol.4, p.487, 2013.

[10] 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.

[11] A. Gabrielsson, "Emotion perceived and emotion felt: Same or different?," Musicae scientiae, vol.5, no.1_suppl, pp.123–147, 2001.

[12] J.A. Russell, "A circumplex model of affect.," Journal of personality and social psychology, vol.39, no.6, pp.1161–1178, 1980.

[13] 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.

[14] I. Wallis, T. Ingalls, E. Campana, and J. Goodman, "A rule-based generative music system controlled by desired valence and arousal," Proc. 8th international sound and music computing conference (SMC), pp.156–157, 2011.

[15] R.J. Larsen and T. Ketelaar, "Personality and susceptibility to positive and negative emotional states.," Journal of personality and social psychology, vol.61, no.1, pp.132–140, 1991.

[16] 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.

[17] T. Song, W. Zheng, P. Song, and Z. Cui, "Eeg emotion recognition using dynamical graph convolutional neural networks," IEEE Transactions on Affective Computing, vol.11, no.3, pp.532–541, 2018.

[18] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, "Analysis of eeg signals and facial expressions for continuous emotion detection," IEEE Transactions on Affective Computing, vol.7, no.1, pp.17–28, 2015.

[19] A. Singhal, P. Kumar, R. Saini, P.P. Roy, D.P. Dogra, and B.-G. Kim, "Summarization of videos by analyzing affective state of the user through crowdsource," Cognitive Systems Research, vol.52, pp.917–930, 2018.

[20] J.A. Miranda-Correa, M.K. Abadi, N. Sebe, and I. Patras, "Amigos: A dataset for affect, personality and mood research on individuals and groups," IEEE Transactions on Affective Computing, vol.12, no.2, pp.479–493, 2021.

[21] 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 transactions on affective computing, vol.3, no.1, pp.18–31, 2011.

[22] J. Hu, C. Wang, Q. Jia, Q. Bu, R. Sutcliffe, and J. Feng, "Scalingnet: extracting features from raw eeg data for emotion recognition," Neurocomputing, vol.463, pp.177–184, 2021.

[23] S. Makeig, A.J. Bell, T.P. Jung, T.J. Sejnowski, et al., "Independent component analysis of electroencephalographic data," Advances in neural information processing systems, pp.145–151, 1996.

[24] L. Pion-Tonachini, K. Kreutz-Delgado, and S. Makeig, "Iclabel: An automated electroencephalographic independent component classifier, dataset, and website," NeuroImage, vol.198, pp.181–197, 2019.

[25] W. Sun, Y. Su, X. Wu, and X. Wu, "A novel end-to-end 1d-rescnn model to remove artifact from eeg signals," Neurocomputing, vol.404, pp.108–121, 2020.

[26] F. Fahimi, Z. Zhang, W.B. Goh, T.S. Lee, K.K. Ang, and C. Guan, "Inter-subject transfer learning with an end-to-end deep convolutional neural network for eeg-based bci," Journal of neural engineering, vol.16, no.2, p.026007, 2019.

[27] V.J. Lawhern, A.J. Solon, N.R. Waytowich, S.M. Gordon, C.P. Hung, and B.J. Lance, "Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces," Journal of neural engineering, vol.15, no.5, p.056013, 2018.

[28] N. Thammasan, K. Moriyama, K.-i. Fukui, and M. Numao, "Familiarity effects in eeg-based emotion recognition," Brain informatics, vol.4, no.1, pp.39–50, 2017.

[29] M.V. Yeo, X. Li, K. Shen, and E.P. Wilder-Smith, "Can svm be used for automatic eeg detection of drowsiness during car driving?," Safety Science, vol.47, no.1, pp.115–124, 2009.

[30] M. Murugappan, R. Nagarajan, and S. Yaacob, "Comparison of different wavelet features from eeg signals for classifying human emotions," 2009 IEEE symposium on industrial electronics & applications, pp.836–841, IEEE, 2009.

[31] S. Tripathi, S. Acharya, R.D. Sharma, S. Mittal, and S. Bhattacharya, "Using deep and convolutional neural networks for accurate emotion classification on deap dataset," Proc. Thirty-First AAAI Conference on Artificial Intelligence, pp.4746–4752, 2017.

[32] F. Wang, S. Wu, W. Zhang, Z. Xu, Y. Zhang, C. Wu, and S. Coleman, "Emotion recognition with convolutional neural network and eeg-based efdms," Neuropsychologia, vol.146, p.107506, 2020.

[33] 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.112767–112776, 2019.

[34] J. Li, S. Qiu, Y.-Y. Shen, C.-L. Liu, and H. He, "Multisource transfer learning for cross-subject eeg emotion recognition," IEEE Trans. Cybern., vol.50, no.7, pp.3281–3293, 2019.

[35] C.-S. Wei, Y.-P. Lin, Y.-T. Wang, C.-T. Lin, and T.-P. Jung, "A subject-transfer framework for obviating inter- and intra-subject variability in eeg-based drowsiness detection," NeuroImage, vol.174, pp.407–419, 2018.

[36] 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.

[37] A. Frydenlund and F. Rudzicz, "Emotional affect estimation using video and eeg data in deep neural networks," Canadian Conference on Artificial Intelligence, pp.273–280, Springer, 2015.

[38] 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.

[39] A.S. Widge, D.D. Dougherty, and C.T. Moritz, "Affective brain-computer interfaces as enabling technology for responsive psychiatric stimulation," Brain-Computer Interfaces, vol.1, no.2, pp.126–136, 2014.

[40] 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.

[41] Y. Liu, O. Sourina, and M.K. Nguyen, "Real-time eeg-based emotion recognition and its applications," in Transactions on computational science XII, pp.256–277, Springer, 2011.

[42] O. Sourina, Y. Liu, and M.K. Nguyen, "Real-time eeg-based emotion recognition for music therapy," Journal on Multimodal User Interfaces, vol.5, no.1-2, pp.27–35, 2012.

[43] 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.

[44] K. Miyamoto, H. Tanaka, and S. Nakamura, "Music generation and emotion estimation from eeg signals for inducing affective states," Companion Publication of the 2020 International Conference on Multimodal Interaction, pp.487–491, 2020.

[45] K. Miyamoto, H. Tanaka, and S. Nakamura, "Emotion estimation from eeg signals and expected subjective evaluation," 2021 9th International Winter Conference on Brain-Computer Interface (BCI), pp.1–6, IEEE, 2021.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] Y. Yang, Q. Wu, Y. Fu, and X. Chen, "Continuous convolutional neural network with 3d input for eeg-based emotion recognition," International Conference on Neural Information Processing, pp.433–443, Springer, 2018.

[51] S.J. Pan and Q. Yang, "A survey on transfer learning," IEEE Trans. Knowl. Data Eng., vol.22, no.10, pp.1345–1359, 2009.

[52] Y.-P. Lin and T.-P. Jung, "Improving eeg-based emotion classification using conditional transfer learning," Frontiers in human neuroscience, vol.11, p.334, 2017.

[53] K. Miyamoto, H. Tanaka, and S. Nakamura, "Meta-learning for emotion prediction from eeg while listening to music," Companion Publication of the 2021 International Conference on Multimodal Interaction, pp.324–328, 2021.

Kana Miyamoto received a B.S. degree from the Osaka Institute of Technology in 2019 and an M. Eng. degree from the Nara Institute of Science and Technology in 2021. She is currently a Ph.D. student at the Nara Institute of Science and Technology. Her research interests include brain-computer interfaces.

Hiroki Tanaka received the master's and Ph.D. degrees from the Nara Institute of Science and Technology, Japan, in 2012 and 2015, respectively. He is an Assistant Professor with the Graduate School of Information Science, Nara Institute of Science and Technology. His research interest is assisting people with disabilities through human–computer interaction.

Satoshi Nakamura is a professor at the Nara Institute of Science and Technology, Team Leader of the Tourism Information Analytics Team, AIP Center, RIKEN, and Honorary professor at the Karlsruhe Institute of Technology, Germany. He received his B.S. from the Kyoto Institute of Technology in 1981 and a Ph.D. from the Kyoto University in 1992. He was the Director of ATR Spoken Language Communication Research Laboratories in the period 2000-2008 and Vice President of ATR in the period 2007-2008. He was the Director General of Keihanna Research Laboratories, National Institute of Information and Communications Technology in 2009-2010. He is currently a professor of the Augmented Human Communication laboratory, Graduate School of Science and Technology, and the director at the Data Science Center, Nara Institute of Science and Technology, Japan. He is working on modeling and systems of spoken language processing including speech-to-speech translation. He is one of the leaders of speech-to-speech translation research and has been serving for various speech-to-speech translation research projects in the world. He received Yamashita Research Award, Kiyasu Award from the Information Processing Society of Japan, Telecom System Award, AAMT Nagao Award, Docomo Mobile Science Award in 2007, ASJ Award for Distinguished Achievements in Acoustics. He received the Commendation for Science and Technology by the Minister of Education, Science and Technology, and the Commendation for Science and Technology by the Minister of Internal Affair and Communications. He also received LREC Antonio Zampolli Award 2012. He was an Elected Board Member of the International Speech Communication Association, ISCA, in the period June 2011-2019, and IEEE Signal Processing Magazine Editorial Board member in 2012–2015, IEEE SPS Speech and Language Technical Committee Member in 2013–2015. He is an IEEE Fellow, ISCA Fellow, IPSJ Fellow, and ATR Fellow.