Music Generation and Emotion Estimation from EEG Signals for Inducing Affective States

Kana Miyamoto
Division of Information Science, Nara Institute of Science and Technology
Ikoma-shi, Nara, Japan
miyamoto.kana.mk4@is.naist.jp

Hiroki Tanaka
Division of Information Science, Nara Institute of Science and Technology
Ikoma-shi, Nara, Japan
hiroki-tan@is.naist.jp

Satoshi Nakamura
Division of Information Science, Nara Institute of Science and Technology
Ikoma-shi, Nara, Japan
s-nakamura@is.naist.jp

ABSTRACT

Although emotion induction using music has been studied, the emotions felt by listening to it vary among individuals. In order to provide personalized emotion induction, it is necessary to predict an individual's emotions and select appropriate music. Therefore, we propose a feedback system that generates music from the continuous value of emotion estimated from electroencephalogram (EEG). In this paper, we describe a music generator and a method of emotion estimation from EEG to construct a feedback system. First, we generated music by calculating parameters from the valence and arousal values of the desired emotion. Our generated music was evaluated by crowdworkers. The median of the correlation coefficients between the input of the music generator and the emotions felt by the crowdworkers were valence r=0.60 and arousal r=0.76. Next, we recorded EEG when listening to music and estimated emotions from them. We compared three regression models: linear regression and convolutional neural network (with/without transfer learning). We obtained the lowest RMSE (valence: 0.1807, arousal: 0.1945) between the actual and estimated emotional values with a convolutional neural network with transfer learning.

CCS CONCEPTS

• **Human-centered computing** → **Interaction design**.

KEYWORDS

Electroencephalogram; emotion induction; music generation; brain computer interface; human computer interaction

ACM Reference Format:
Kana Miyamoto, Hiroki Tanaka, and Satoshi Nakamura. 2020. Music Generation and Emotion Estimation from EEG Signals for Inducing Affective States. In *Companion Publication of the 2020 International Conference on Multimodal Interaction (ICMI '20 Companion), October 25–29, 2020, Virtual event, Netherlands.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3395035.3425225

ICMI '20 Companion, October 25–29, 2020, Virtual event, Netherlands

ACM ISBN 978-1-4503-8002-7/20/10…$15.00
https://doi.org/10.1145/3395035.3425225

1 INTRODUCTION

Music is known to induce emotions [6, 18]. Music has been used to improve health such as elderly people's mental health [11]. Currently, music therapy is done by music therapists who carefully observe participant behaviors and decide what kind of music to use. Since musical possibilities include a variety of existing kinds as well as musical improvisation, therapists are burdened by needing to select music. Therefore, Sourina et al. proposed a system that automatically plays music that is designed to induce emotions [9, 17].

The system suffers from two problems. The first is how music is chosen. The emotions that are promoted when listening to music vary among individuals. The same person may have different emotional levels depending on the situation. The system needs to choose music that is appropriate for the individual and the situation. The second problem is a method that predicts the emotional state of participants. The system needs to observe emotion as diligently as a music therapist. Hence, Sourina et al. proposed emotion prediction using EEG of biological signals, and music selection by predicted emotion to solve these problems. EEG records the brain's electrical activity and measures time-series data in multiple channels. There are various researches using EEG, such as the classification of the dynamic representation of speaker stance [5] and the detection of syntactic violations [10]. EEG is also effective for emotional prediction. The method using EEG has been actively researched for the system to recognize emotions such as the classification of emotions [2, 16]. After emotion estimation from EEG, Sourina et al. planned to compare the desired emotion with the predicted emotion and select music from music databases. Although it is thought to be possible to induce emotions using this method, it is difficult to express emotions in detail and to change the music flexibly. Therefore, we propose a system that expresses emotions as continuous values and generates music from EEG.

Emotions are expressed in two-dimensional space [12]. The horizontal dimension is valence, representing a range from pleasant to unpleasant. The vertical dimension is arousal, representing a range from activation to deactivation. In our research, these are expressed as continuous values from 0 to 1 and express emotions in detail. Ehrlich et al. proposed a music generation method to make participants perceive the continuous values of emotions intended by the music, and they estimated the emotions using EEG while listening to music created by the method [3]. Our research was based on their research, and we proposed new methods of music generation for emotional elicitation and emotion estimation with three regression models by EEG.

2 FRAMEWORK OF AN EMOTION-INDUCING SYSTEM

We propose a feedback system that generates music from the continuous value of emotion estimated from EEG. The music generation feedback system's outline, shown in Figure 1, consists of a music generator that creates appropriate music to emotion elicitation and an emotion estimator that estimates emotions by EEG. First, the desired emotion that specified in the experiment is input to the music generator. Next, the participant's current emotion is estimated from the EEG while they are listening to music. The difference value between the estimated emotion and the desired emotion is added to the previous input of the music generator. The new input to the music generator is calculated and music is generated again. The system induces their emotions by repeating this cycle.

In this paper, we describe how to create the music generator and the emotion estimator to construct the system based on the research of Ehrlich et al. [3]. We examined a new method of emotion elicitation using the music generator, and new models such as CNN for emotion estimation by EEG.

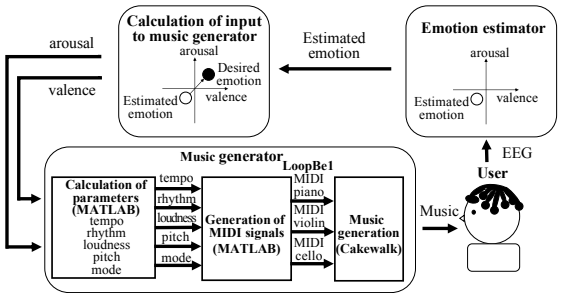

Figure 1: System that generates music from EEG

3 MUSIC GENERATOR

We created a music generator using the method of Ehrlich et al. whose music generator was designed to help the participants perceive the intended emotion of the music [3]. However, we design to change the emotions being experienced by the participants through music. Based on previous research, the felt and perceived emotions are different, but it is thought that these emotions are the same or the felt emotion often appears weaker than the perceived emotion [4, 14]. Therefore, we thought the music generator of Ehrlich et al. might also effectively induce emotions and created it. In addition, we evaluated the music generator and recreated the music generator to produce music more suitable for emotional elicitation.

3.1 Parameter-based music generation

Previous research generated music by five parameters: tempo, rhythm, loudness, pitch, and mode [3]. We generated music using the following formula.

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

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

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

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

$$\text{mode}: \quad 7 - (6 * \text{val}) \in 1, ..., 7 \subset \mathbb{N}.$$

Tempo represents a note's length in seconds. Rhythm represents the probability that a note appears. However, in this expression, music isn't generated when the arousal is 0. Therefore, when the arousal was 0.03 or less, it was set to 0.03 and rhythm was calculated. Loudness represents the note's volume. Pitch represents which scale is used. Mode determines which method to use: 1. Lydian (4th mode), 2. Ionian (1st mode), 3. Mixolydian (5th mode), 4. Dorian (2nd mode), 5. Aeolian (6th mode), 6. Phrygian (3rd mode), and 7. Locrian (7th mode) [13]. We used a C major chord. The chord symbol changes as I-IV-V-I for each measure. To generate music, we calculated these music parameters by MATLAB from valence and arousal inputs between 0 and 1 and generated musical instrument digital interface (MIDI) signals with MATLAB from the music parameters. MIDI signals were sent to DAW software called Cakewalk using LoopBe1 of virtual MIDI cable software, and music was generated by a piano, a violin, and a cello.

3.2 Evaluation of created music

We evaluated the music produced by the music generator. We investigated whether the music generator designed by Ehrlich et al. was effective in inducing as well as perceiving emotions.

3.2.1 Assessment methods. Sample pieces of music (15 s) are {val, aro}={0,0}; {0,1}; {0.5,0.5}; {1,0}; {1,1}. The evaluated music (30 s) are a combination of 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. Crowdworkers of CrowdWorks, Inc. listened to the music and evaluated its valence and arousal using self assessment mannequin (SAM) [1] in nine steps between 0 to 1 every time.

3.2.2 Perceived emotion by the music generator. 101 crowdworkers listened to the music and evaluated whether they could perceive the intended emotions in the music. Min-max normalization was used to rescale the valence and arousal for each crowdworker. We investigated the Pearson's linear correlation coefficients between the input to the music generator and the evaluation by each crowdworker. We excluded data of participants whose correlation coefficients were negative. The median correlation coefficients were valence r=0.76 and arousal r=0.86. The evaluation of the music generator by Ehrlich et al. investigated the median correlation coefficients were valence r=0.52 and arousal r=0.74.

Figure 2 shows the average evaluations of crowdworkers for the valence and arousal input to the music generator. The horizontal axis is the input of the valence, and the vertical axis is the input of the arousal to the music generator. The colors represent the average evaluations of crowdworkers. Figure 2 (a) of valence evaluation tends to have difficulty perceiving a high valence when the arousal input to the music generator is low because the colors are not aligned vertically. Figure 2 (b) of arousal evaluation does not seem to be affected by the valence input to the music generator because the colors are aligned horizontally. The figures were similar to those

of Ehrlich et al. [3]. These results suggest that we could create the music generator that perceives the emotion intended by music, based on previous research.

3.2.3 Felt emotion by the music generator. 108 crowdworkers gathered separately from perceive emotion evaluation listened to the music, and evaluated the emotions they felt. We investigated the correlation coefficients between the input to the music generator and the evaluations of each crowdworker using the same method as in the previous section. The median correlation coefficients were valence r=0.59 and arousal r=0.83.

Figure 2 (c) of valence evaluation and Figure 2 (d) of arousal evaluation show the same trend as Figure 2 (a) and (b). The change in the color map is more gradual than the valence and arousal evaluations, which were perceived as the music's intention. This result indicates that the emotion felt when listening to music is weaker than the perceived emotion [4, 14].

3.3 Recreating music generator

In the previous section, the evaluation of felt valence was influenced by arousal input to the music generator. The created music generator is likely to have a difference between the input valence and the valence felt by the participants. Therefore, models that input appropriate valence and arousal values to the music generator were trained by support vector regression and immediately connected before the music generator. For the valence model which input appropriate valence, the valence input was predicted from the evaluated valence and arousal. For the arousal model which input appropriate arousal, the arousal input was predicted from the evaluated valence and arousal. The support vector regression models were verified by 3-fold cross-validation. We obtained the RMSE (valence: 0.0231, arousal: 0.0240). We performed training by support vector regression using all the data and recreated the music generator.

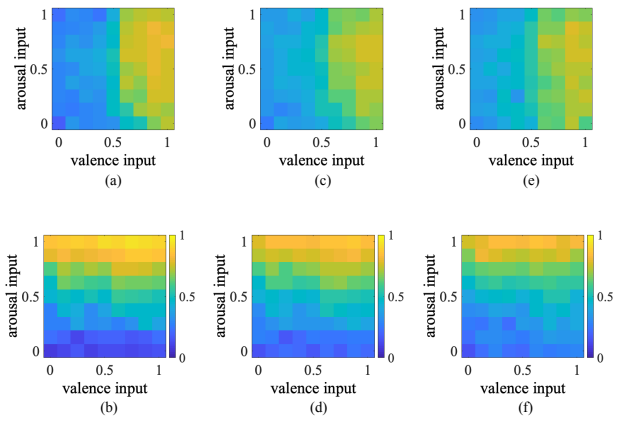

Figure 2: Color map of perceived and felt emotions. (a) The perceived valence evaluation. (b) The perceived arousal evaluation. (c) The felt valence evaluation. (d) The felt arousal evaluation. (e) The felt valence evaluation using recreated the music generator. (f) The felt arousal evaluation using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

104 crowdworkers gathered separately from the previous evaluations listened to the music generated by recreated the music generator, and evaluated the emotions they felt. We investigated the correlation coefficients between the input to the music generator and the evaluations of each crowdworker using the same method as in the previous section. The median correlation coefficients were valence r=0.60 and arousal r=0.76.

Figure 2 (e) of valence evaluation and Figure 2 (f) of arousal evaluation does not seem to be affected by the arousal or valence input to the music generator. Although the correlation coefficients of arousal were lower, the evaluation of valence was improved from Figure 2 (c). These suggest that we could create the music generator to induce emotions.

4 EMOTION ESTIMATOR BY EEG

In our proposed music generation feedback system based on participant emotions, valence and arousal between 0 and 1 must be estimated using EEG. Ehrlich et al. estimated emotions based on linear discriminant analysis (LDA) and the sigmoid function using the logarithm of variances of EEG signals while listening to music created by the music generator [3]. We used the same features and linear regression, which resembles Ehrlich et al.'s research. In addition, we proposed CNN and a CNN with transfer learning.

4.1 Participants

20 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.

4.2 Experimental design

The participants sat in front of a desk on which a monitor was placed and wore earphones. Before putting on the electroencephalograph, they listened to sample music (15 s) {val, aro}={0,0}; {0,1}; {0.5,0.5}; {1,0}; {1,1}.

The experimental procedure is as follows. First, they silently gazed at a cross in the center of the monitor for 5 s. Next, they listened to music for 20 s while continuing to gaze at the cross. The screen changed after they listened to the music, and the participants evaluated the valence and the arousal of the emotion they felt by listening to music. SAM was used as the evaluation method, and each response was given in nine steps. The participant did this practice for two pieces of music (20 s) {val, aro}={0.125,0.25}; {0.875,0.75}. The participants put on a CGX Quick-30 electroencephalograph after the practice. We recorded their EEG and the subjective evaluations of 41 pieces of music (20 s) with the same method as in the practice.

4.3 EEG preprocessing

We did preprocessing for each participant based on the following procedure.
(1) We removed the data that caused problems, including music that wasn't played. (2) We selected the data of the following 14 channels: AF3, AF4, F3, F4, F7, F8, FC5, FC6, T7, T8, P7, P8, O1, O2. (3) The EEG signals were downsampled from 1000Hz to 200Hz. (4) The silent state of 3 to 5 s was divided into two epochs of 1 s data. We also divided the 0 to 20 s of music listening state into 20 epochs of 1 s data. (5) We designed 2nd order zero phase Chebyshev

IIR bandpass filters that pass the following values: theta (4-7 Hz), alpha (8-13 Hz), low beta (14-21 Hz), high beta (22-29 Hz), gamma (30-45 Hz). (6) We divided the EEG signals into five frequency bands by the designed filters and calculated the $f = log(var(EEGdata))$, which is the logarithm of the waveform variance for each bit of data. (7) We subtracted the average of the logarithm of the silent waveform variance from the logarithmic waveform variance while they listened to the music.

4.4 Regression

We validated three models such as linear regression, CNN, and CNN with transfer learning using 70 features (14 channels×5 frequency bands) obtained in the preprocessing by MATLAB. We used a hold-out method this time. In the case of linear regression, the train and test data are 9:1. In the case of CNN and CNN with transfer learning, the train, validation, and test data are 8:1:1. The test data is identical to compare the three models.

4.4.1 Linear regression. Linear regression is the baseline model selected by previous research [3]. The features were input in vector format and the data was normalized.

4.4.2 CNN. The features calculated from the logarithm of the waveform variance were treated as 6×6×5 images (Figure 3). This sequence was adapted from Jinpeng et al. [8]. As in Figure 3, the CNN consists of a convolution layer (2×2 size, 1 stride), a batch normalization layer, a ReLU layer, a dropout layer, a fully connected layer (output dimensionality of 100), a dropout layer, a fully connected layer, and a regression output layer. We decided the parameters: initial learning rate of 0.001 and batch size of 20. Using Bayesian optimization [15], the parameters were selected from dropout (0 to 0.2), the number of filters (16 to 256).

4.4.3 CNN with transfer learning. We applied the same reprocessing method to the DEAP dataset [7]. The DEAP dataset recorded the EEG and subjective emotions while the participants watched music videos. Although we didn't use video in our experiment, we performed transfer learning using the DEAP dataset because we used music. First, our CNN was trained by the DEAP dataset. We decided the parameters: initial learning rate of 0.001 and batch size of 2000. Using Bayesian optimization, the parameters were selected from dropout (0 to 0.2), the number of filters (16 to 256). Next, the last three layers (Figure 3 dotted box) were trained with transfer learning. We decided the parameters: initial learning rate of 0.0005, batch size of 20. Using Bayesian optimization, the parameters were selected from learning rate factor for weight and biases of the fully connected layer (2 to 40).

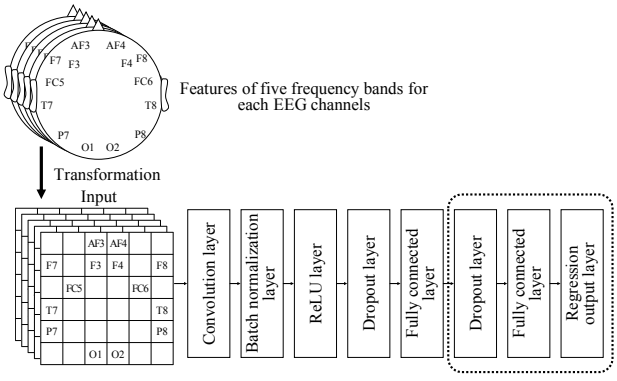

Figure 3: Proposed CNN (with/without transfer learning)

4.4.4 Result. The linear regression, CNN, and CNN with transfer learning results are shown in Table 1. The result of the CNN using DEAP dataset was RMSE (valence: 0.2498, arousal: 0.2383). We obtained the lowest RMSE (valence: 0.1807, arousal: 0.1945) between the felt and estimated emotional values when we used the CNN with transfer learning for both the valence and the arousal. For the Wilcoxon signed-rank test result, we found a significant difference between the linear regression and CNN with transfer learning for both valence and arousal ($p<0.05$).

Table 1: RMSE of felt and estimated emotions. Bold fonts indicate the lowest RMSE of the three models.

Par.	Linear		CNN		CNN TL	
	val	aro	val	aro	val	aro
1	0.1887	0.2254	0.1764	0.2319	0.1636	0.2171
2	0.2517	0.1872	0.2299	0.1837	0.2352	0.1745
3	0.2033	0.2240	0.2088	0.2119	0.1980	0.1904
4	0.1008	0.1253	0.0939	0.1322	0.0839	0.1206
5	0.2506	0.2221	0.2354	0.2087	0.2332	0.2017
6	0.2199	0.2548	0.2206	0.2400	0.2081	0.2442
7	0.2896	0.3108	0.2942	0.2684	0.2779	0.2546
8	0.1952	0.2505	0.1957	0.2547	0.2028	0.2566
9	0.1980	0.2239	0.2105	0.2194	0.2073	0.2021
10	0.0617	0.1341	0.0633	0.1292	0.0581	0.1195
11	0.1265	0.1660	0.1299	0.1798	0.1209	0.1575
12	0.1849	0.2253	0.2028	0.2218	0.1834	0.2115
13	0.1785	0.2356	0.1731	0.2298	0.1601	0.2116
14	0.1507	0.1699	0.1635	0.1640	0.1691	0.1702
15	0.2211	0.1675	0.2286	0.2084	0.2148	0.1758
16	0.0866	0.1426	0.0923	0.1423	0.0766	0.1350
17	0.3437	0.2912	0.3149	0.2865	0.3155	0.2869
18	0.1221	0.1358	0.1214	0.1220	0.1088	0.1032
19	0.0664	0.1935	0.0736	0.1711	0.0712	0.1684
20	0.3503	0.3094	0.3407	0.3140	0.3261	0.2891
mean	0.1895	0.2098	0.1885	0.2060	**0.1807**	**0.1945**
std	0.0826	0.0567	0.0775	0.0533	0.0777	0.0538

5 CONCLUSION

We proposed the system that expresses emotions as continuous values and generates music from EEG. We created the music generator of the system to induce emotion using support vector regression. We also confirmed CNN with transfer learning was more effective than linear regression for emotion estimation from EEG while listening to generated music.

In order to construct the feedback system, we will improve models for emotion estimation and validate them using the leave-one-out whether they can respond to music participants have never listened to. Furthermore, we will construct a feedback system that combines the music generator and the emotion estimator and compare the emotion estimated from the EEG when using the system with the desired emotion.

6 ACKNOWLEDGMENTS

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

REFERENCES

[1] Margaret M Bradley and Peter J Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* 25, 1 (1994), 49–59.

[2] Harsh Dabas, Chaitanya Sethi, Chirag Dua, Mohit Dalawat, and Divyashikha Sethia. 2018. Emotion Classification Using EEG Signals. In *Proceedings of the 2018 2nd International Conference on Computer Science and Artificial Intelligence*. 380–384.

[3] Stefan K Ehrlich, Kat R Agres, Cuntai Guan, and Gordon Cheng. 2019. A closed-loop, music-based brain-computer interface for emotion mediation. *PloS one* 14, 3 (2019), 1–24.

[4] Alf Gabrielsson. 2001. Emotion perceived and emotion felt: Same or different? *Musicae scientiae* 5, 1_suppl (2001), 123–147.

[5] Xiaoming Jiang. 2019. Single-trial Based EEG Classification of the Dynamic Representation of Speaker Stance: A Preliminary Study with Representational Similarity Analysis. In *NeuroManagement and Intelligent Computing Method on Multimodal Interaction*. 1–4.

[6] Stefan Koelsch. 2009. A neuroscientific perspective on music therapy. *Ann. NY Acad. Sci.* 1169 (2009), 374–384.

[7] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2011. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing* 3, 1 (2011), 18–31.

[8] Jinpeng Li, Zhaoxiang Zhang, and Huiguang He. 2018. Hierarchical convolutional neural networks for EEG-based emotion recognition. *Cognitive Computation* 10, 2 (2018), 368–380.

[9] Yisi Liu, Olga Sourina, and Minh Khoa Nguyen. 2011. Real-time EEG-based emotion recognition and its applications. In *Transactions on computational science XII*. Springer, 256–277.

[10] Shunnosuke Motomura, Hiroki Tanaka, and Satoshi Nakamura. 2019. Detecting Syntactic Violations from Single-trial EEG using Recurrent Neural Networks. In *Adjunct of the 2019 International Conference on Multimodal Interaction*. 1–5.

[11] Rafael Ramirez, Manel Palencia-Lefler, Sergio Giraldo, and Zacharias Vamvakousis. 2015. Musical neurofeedback for treating depression in elderly people. *Frontiers in neuroscience* 9 (2015), 354.

[12] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[13] Mark A Schmuckler. 1989. Expectation in music: Investigation of melodic and harmonic processes. *Music Perception: An Interdisciplinary Journal* 7, 2 (1989), 109–149.

[14] Emery Schubert. 2013. Emotion felt by the listener and expressed by the music: literature review and theoretical perspectives. *Frontiers in psychology* 4 (2013), 837.

[15] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. 2012. Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*. 2951–2959.

[16] Morteza Zangeneh Soroush, Keivan Maghooli, Seyed Kamaledin Setarehdan, and Ali Motie Nasrabadi. 2017. A review on EEG signals based emotion recognition. *International Clinical Neuroscience Journal* 4, 4 (2017), 118.

[17] Olga Sourina, Yisi Liu, and Minh Khoa Nguyen. 2012. Real-time EEG-based emotion recognition for music therapy. *Journal on Multimodal User Interfaces* 5, 1-2 (2012), 27–35.

[18] Isaac Wallis, Todd Ingalls, and Ellen Campana. 2008. Computer-generating emotional music: The design of an affective music algorithm. *DAFx-08, Espoo, Finland* 712 (2008), 7–12.