

MUSIC-BASED EMOTION PROFILING SYSTEM**CROSS-REFERENCE TO RELATED APPLICATION**

[0001] This application claims the benefit of priority to the Singapore application no. 10202260089U filed November 14, 2022, the contents of which are hereby incorporated by reference in their entirety for all purposes.

TECHNICAL FIELD

[0002] This application relates generally to the field of biomedical analytics, and more particularly, to a system and method of generating emotional reaction profiles of users.

BACKGROUND

[0003] Emotion-related mental disorders can negatively impact the daily lives of patients and people around them. Some emotion regulation methods can help relieve some emotion-related mental disorders. However, there remains a need for tools to help healthcare providers and patients implement effective emotion regulation programs.

SUMMARY

[0004] According to an aspect, disclosed herein a system. The system comprises: memory storing instructions; and a processor coupled to the memory and configured to process the stored instructions to implement: a module configured to perform a method of generating a personalized emotion profile for a user, the method including: generating a predicted emotional response in a first stage, the first stage including: providing an EEG input acquired from the user to a feature learner, the feature learner being pre-trained on previous EEG inputs to output a hidden embedding; and using a self-assessed emotion rating as a prediction target in a first

class predictor to finetune an output of the hidden embedding; and in a second stage, generating a final predicted emotional response by applying a refine function to both the predicted emotional response and a personalized emotional response of the user to a music clip, the personalized emotional response being generated using a second class predictor on the output of the hidden embedding, wherein the EEG input is acquired concurrent with a time period in which the music clip is played to the user, and wherein the self-assessed emotion rating was received from the user responsive to hearing the music clip, and wherein the music clip is associated with an emotional intensity.

[0005] According to another aspect, disclosed herein a method to generate a personalized emotion profile for a user. The method comprises: generating a predicted emotional response in a first stage, the first stage including: providing an EEG input acquired from the user to a feature learner, the feature learner being pre-trained on previous EEG inputs to output a hidden embedding; and using a self-assessed emotion rating as a prediction target in a first class predictor to finetune an output of the hidden embedding; and in a second stage, generating a final predicted emotional response by applying a refine function to both the predicted emotional response and a personalized emotional response of the user to a music clip, the personalized emotional response being generated using a second class predictor on the output of the hidden embedding, wherein the EEG input is acquired concurrent with a time period in which the music clip is played to a user, and wherein the self-assessed emotion rating was received from the user responsive to hearing the music clip, and wherein the music clip is associated with an emotional intensity.

[0006] According to yet another aspect, disclosed herein a system. The system comprises: a media player, the media player being configured to play a music clip audible to a user for a time period, the music clip being associated with an emotional intensity; electroencephalogram (EEG) electrodes, the EEG electrodes being attachable to the user's head to detect activity in a

plurality of functional areas of the user's brain, the EEG electrodes being configured to acquire EEG input concurrent with the time period; a user input device, the user input device operable by the user to input a self-assessed emotion rating responsive to hearing the music clip; memory storing instructions; and a processor coupled to the memory and configured to process the stored instructions to implement: a module being configured to perform a method of generating a personalized emotion profile for the user, the method including: generating a predicted emotional response in a first stage, the first stage including: providing the EEG input to a feature learner, the feature learner being pre-trained on previous EEG inputs to output a hidden embedding; and using the self-assessed emotion rating as a prediction target in a first class predictor to finetune an output of the hidden embedding; and in a second stage, generating a final predicted emotional response by applying a refine function to both the predicted emotional response and a personalized emotional response of the user to the music clip, the personalized emotional response being generated using a second class predictor on the output of the hidden embedding, wherein the module is further configured to select one or more selected music clips to compile an emotion regulation playlist for use in emotion regulation of the user, and wherein each of the one or more selected music clips is selected according to the personalized emotion profile of the user.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] Various embodiments of the present disclosure are described below with reference to the following drawings:

[0008] FIG. 1 is a schematic diagram showing a system according to embodiments of the present disclosure;

[0009] FIG. 2 is a schematic diagram showing the structure of a deep learning model of the present system;

- [0010] FIG. 3 is a schematic diagram illustrating the space-aware temporal convolutional layer of the deep learning model of FIG. 2;
- [0011] FIG. 4A schematically illustrates the training stage in the proposed method of generating a personalized emotion profile;
- [0012] FIG. 4B schematically illustrates the evaluation stage in the proposed method of generating a personalized emotion profile;
- [0013] FIG. 5 shows an example of a music-based emotion profile of one user;
- [0014] FIG. 6 shows the polynomial regression correlating user response with the emotion intensity of the stimuli;
- [0015] FIG. 7 schematically illustrates a system to develop a stimuli program based on the emotion profile of the present disclosure; and
- [0016] FIG. 8 is a schematic diagram of a processor system.

DETAILED DESCRIPTION

- [0017] The following detailed description is made with reference to the accompanying drawings, showing details and embodiments of the present disclosure for the purposes of illustration. Features that are described in the context of an embodiment may correspondingly be applicable to the same or similar features in the other embodiments, even if not explicitly described in these other embodiments. Additions and/or combinations and/or alternatives as described for a feature in the context of an embodiment may correspondingly be applicable to the same or similar feature in the other embodiments.
- [0018] In the context of various embodiments, the articles “a”, “an” and “the” as used with regard to a feature or element include a reference to one or more of the features or elements.

[0019] In the context of various embodiments, the term “about” or “approximately” as applied to a numeric value encompasses the exact value and a reasonable variance as generally understood in the relevant technical field, e.g., within 10% of the specified value.

[0020] As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items.

[0021] As used herein, terms “concurrently”, “simultaneously”, “at the same time”, or the like, may refer to events or actions that coincide or overlap within a period of time, regardless of whether the events start at the same time instant, and regardless of whether the events end at the same time instant.

[0022] As used herein, emotion-related mental disorders refer generically to medical conditions such as generalized anxiety disorders (GAD) and depression. Persons suffering from GAD may experience an excessive amount of fear (in situations that generally do not elicit such strong reactions in persons without GAD). Depression is known to cause biological aging, increase the risk of obesity, frailty, diabetes, cognitive impairment, and mortality.

[0023] As used herein, the term “emotion regulation” refers to a process by which people try to control or regulate their emotional mental states in terms of which emotion they should have, when they should have the desired emotions, how they should express their emotions, and even how they should experience the desired emotions. Some therapists try to train their patients in emotion regulation based on cognitive reappraisal and/or expressive suppression. Cognitive reappraisal involves having the patient interpret or re-evaluate the reasons that lead to the negative emotions, to reduce the impact of the negative emotions. Expressive suppression focuses on modulating the patient’s response to the negative emotions, to try to break the self-reinforcing cycle of negative emotion and expressive behavior caused by the negative emotions.

[0024] As used herein, electroencephalogram (EEG) refers to the electrical signals (also referred to herein as EEG data) corresponding to brain activity. EEG data includes a temporal dimension and a spatial dimension.

[0025] The temporal dimension reflects the variations in brain activities over time. In the temporal dimension, EEG data may fall within different frequency bands, e.g., delta band (1 Hz - 4 Hz), theta band (4 Hz - 7 Hz), alpha band (8 Hz - 13 Hz), beta band (13 Hz - 30 Hz), and gamma band (greater than 30 Hz). Emotion is associated with the alpha band, the beta band, and the gamma band.

[0026] The spatial dimension, or the channel dimension, can reflect the cognitive processes in the different brain functional areas of the brain. The spatial domain is related to the physical placement of EEG sensors or electrodes on the surface of the user's head at various locations corresponding to the different functional areas. Functional areas in the context of the human brain refers to the frontal lobe, the parietal lobe, the temporal lobe, and the occipital lobe. In terms of the spatial dimension, frontal asymmetric responses can be used as an indicator of emotional processes in the brain. Asymmetry patterns in emotion-related EEG signals can additionally be observed in the prefrontal, parietal, and/or temporal lobes of the brain.

[0027] Some conventional methods apply single-scale 1D (one-dimensional) convolutional kernels or flattened feature vectors. For example, some conventional methods may use flattened relative power spectral density (rPSD) features as direct input into a temporal convolutional neural network, essentially ignoring spatial patterns in the EEG signals. Some other conventional methods use features averaged across different frequency bands. In the conventional methods, the spectral patterns across the EEG channel, as well as the spatial patterns of the EEG channels, could not be effectively extracted.

[0028] In reality, the human brain is a complicated system with hierarchical spatial and functional combinations at various levels of neurons, local circuits, and functional areas, with

different functional areas being involved in certain brain functions while not working independently of other functional areas. In addition, emotion is a high-order cognitive process that is a challenge to model accurately without sacrificing efficiency. It has been difficult to extract emotion-related EEG patterns in a sufficiently efficient manner for useful applications. The nature of EEG data contributes to the difficulties involved. For example, EEG signals may be characterized by a relatively low signal-to-noise ratio (SNR). External noises (e.g., from movements as small as blinking of an eye, etc.) and internal noises (e.g., from unrelated mental activities, etc.) can also add to the difficulties involved. EEG data is relatively high-dimensional data. For example, based on an acquisition of EEG signals using 32 channels at a sampling rate of 128 Hz, one second of data would contain $32 \times 128 = 4096$ data points.

[0029] Referring to FIG. 1, in one aspect, the system 200 according to some embodiments of the present disclosure is configurable as a tool to train a user 300 in emotion regulation. In another aspect, the system 200 is configured to generate an emotion profile 250. In particular, the system 200 is configurable to generate a personalized emotion profile or to perform an emotion-based profiling for individuals. For example, using the system 200, the outcome of performing emotion-based profiling is an emotion profile 250 that is personalized or specific to a user 300, in contrast to a general profile that assumes that demographically similar persons will respond in the same manner to similar stimuli. The emotion profile 250 can form the basis for predicting how the user 300 may react (is likely to react) in terms of emotions described or perceived by the user himself/herself. This does not prevent a healthcare provider from generating a first emotion profile of a first user, and subsequently using the first emotion profile for reference in the course of assisting a patient (a second user) in emotion regulation.

[0030] Referring to the functional block diagram shown in FIG. 1, in some embodiments, the system 200 includes a stimulus generation block (SGB) 210. The stimulus generation block 210 may be configured to generate a stimulus 213 by selecting one or more stimuli from a

library of stored stimuli. The stimulus generation block 210 may be configured to present a selected stimulus 213 to the user 300, e.g., by retrieving and playing a previously stored stimulus. In some examples, the stimulus generation block 210 may be configured to digitally generate a stimulus 213 and playing the generated stimulus before the user 300. The stimulus generation block 213 may be configured to provide at least two stimuli, in which each stimulus 213 may be associated with a different emotional intensity 214.

[0031] The system 200 is configured such that the stimulus 213 includes a music clip. As used herein, the term “music clip” is used in a general sense to refer to a short recording of music. A music clip may be a combination of sounds or notes over a relatively short period of time (e.g., a few minutes, or less than a minute). One or more of the music clips used may be made by recording the playing of one or more musical instruments. One or more of the music clips used may be digitally created. The music clips of the present disclosure may not fall under the conventional definition of “music”, e.g., the music clip may not be melodious, harmonious, or composed according to conventional musical genres. The term “musical” as used herein refers more generally to an arrangement of notes or concurrent combinations of notes (e.g., chords) played or generated over defined periods of time.

[0032] The music clips selected or created for use in the present system includes music clips that are known to trigger a change in the emotional intensity of a listener. Emotional intensity may be described with reference to the user’s reaction upon hearing a music clip.

[0033] The emotional response to the same music clip may differ among different users, e.g., a first user may have a negative response to a music clip and a second user may have a neutral response to the same music clip. Different music clips may trigger similar emotional responses with different emotional intensity in the same user, e.g., a user may associate a first music clip with a sad emotional response and the same user may associate a second music clip with a very sad emotional response.

[0034] Some music clips may trigger a change of the types of emotions experienced by the listener. In some examples, a selected music clip may cause a listener to change from feeling happy to feeling sad. Some music clips may trigger a change in the emotional intensity experienced by the listener. In one example, all the music clips used in the system may be associated with the same type of emotion (e.g., excitement) to different emotional intensities, e.g., a little excited, extremely excited, etc.

[0035] Still referring to FIG. 1, in some embodiments, the system 200 includes an electroencephalogram (EEG) acquisition block (EAB) 220. For example, the system 200 may be configured to receive input in the form of EEG signals or data 302 via EEG sensors attached to the user's head. The EEG block 220 is configured to output processed EEG data 224 based on the EEG 302 acquired from the user 300.

[0036] The system 200 includes a self-assessment block (SAB) 230 configured to output emotion ratings 234 based on user input 303 provided by the same user 300. Concurrently with the acquisition of the user input 303, EEG data 302 is collected by the EEG acquisition block 220.

[0037] The system 200 includes a dual-branch emotion profiling block (DPB) 240. The emotion profiling block 240 is configured to output an emotion profile (output block) 250 that is personalized to the user 300, based on the emotional intensity 214 of the music clips (known emotion intensity), the processed EEG data 224, and the emotion ratings 234 (user's self-ratings). The emotion profiling block 240 is configured to perform a unified deep learning-based emotion recognition method and a dynamic profiling refinement method, in accordance with embodiments of the present disclosure.

[0038] To aid understanding, the following describes one of many possible use scenarios in terms of an experiment conducted and the generation of a personalized emotion profile using data acquired from the experiment.

[0039] In the experiment, the user (the person undergoing the profiling) was seated in a comfortable and relaxing place, e.g., on a soft-cushioned sofa. A plurality of music clips with respective known emotional intensity (selected by the therapist for the desired emotional intensity) was presented to the user as the stimulus. In one trial, the stimulus may last for a period of time ranging from about one minute to several minutes. Concurrently, EEG signals of the user were recorded.

[0040] After each music clip was played, the user provided his/her assessment of the emotion he/she felt while listening to the music clip.

[0041] The emotion profiling block generated a personalized emotion profile for the user. The emotion profiling block took the EEG signals, the changes of emotional intensity of the music clips, and the self-assessments as the inputs. A deep learning model pre-trained on a large-scale emotional dataset or previously acquired EEG signals was utilized as the base learner. The base learner used included two parts, e.g., a feature learner (FL) and a class predictor (CP). The base learner was pre-trained using shorter segments of each trial. Another personalized class predictor (PCP) was used to decode the self-ratings using the hidden embedding from the feature learner of the base predictor. Then PCP was then trained, and the FL was finetuned using the newly collected EEG data from the particular user. An n-fold cross-validation was applied for the training and the evaluation. During the evaluation, both the CP and the PCP were utilized as two branches to generate the final prediction of the emotion responses of each music clip. The final prediction and the emotional intensity of the music clips were combined to generate the personalized emotion profile.

[0042] The SBG block 210 is configured to provide a series of multiple generated music clips according to a pre-defined program of changes of emotional intensities. The changes of emotional intensities are denoted as $Y_{EI} \in \mathbb{R}^{1 \times L_{EI}}$, where L_{EI} is the sample number of the emotional intensities.

[0043] The EAB block 220 is configured to receive or to record EEG signals from the user, e.g., the full head EEG signal is collected. A band-pass filter, e.g., from about 0.3 Hz (Hertz) to about 40 Hz, is applied to remove low-frequency noise and high-frequency noise. This pre-processing step is repeated for EEG signals from all electrodes (EEG sensors) to obtain filtered EEG data $X \in \mathbb{R}^{C \times L}$, where C is the number of electrodes (channels), L is the length of one EEG chunk. To train the PCP and to finetune the FL, the pre-processed EEG signal array is split into S short time segments, using a sliding window of length L_w and overlap L_o . One segment $\bar{X} \in \mathbb{R}^{C \times L_w}$ is used as one training sample.

[0044] The SAB block 230 is configured to receive a self-assessment or user input from the user. After each trial (playing one music clip to the user), the user is required to provide his/her assessment on how he/she feels when he/she was listening to the music clip. The self-assessment is denoted as $Y_{SA} \in \mathbb{R}$. For each trial, the self-assessment provides one scalar as a label.

[0045] The DPB block 240 is configured to generate a personalized emotion profile using the EEG, deep learning, self-assessment, and the changes of emotion intensities of the music clips. The DPB block 240 includes a deep learning model, also referred to herein as multi-anchor space-aware temporal convolutional neural networks (MASA-TCN). In this example, the EEG data includes a sequence of five four-EEG-channel sub-segments. In other examples, different numbers and/or differently defined lengths of the channels and/or sub-segments may be used without going beyond the claimed scope.

[0046] The deep learning model is pre-trained using a large-scale emotion dataset. In the present system, the deep learning model serves as a base learner, denoted by $\mathcal{H}_{base}(\cdot)$. The deep learning model includes two main parts: (i) a feature learner (FL) denoted by $\Psi_{FL}(\cdot)$, and a class predictor (CP) denoted by $\mathcal{F}_{CP}(\cdot)$. The emotion response can be predicted by the base learner as follows:

$$p_{base} = \mathcal{H}_{base}(\bar{X}) = \mathcal{F}_{CP}(\Psi_{FL}(\bar{X})). \quad (1)$$

[0047] As schematically illustrated in FIG. 2, the structure of the proposed deep learning model includes the following: (i) a feature extraction block, (ii) a multi-anchor attentive fusion (MAAF) block, (iii) a temporal convolutional neural networks (TCN) block, and (iv) a regression/classification block.

[0048] A space-aware temporal convolutional layer (SAT) is proposed for a feature extraction block, to extract spatial-spectral patterns of EEG using TCN. To better learn the temporal dynamics underlying the emotional cognitive processes that might appear in different time scales, a MAAF block is proposed. The MAAF block may include multiple parallel SATs each with different lengths of 1D causal convolution kernels. In some examples, the MAAF includes three parallel SATs with different lengths of 1D causal convolution kernels. The outputs of these parallel SATs are attentively fused as the input to several TCN layers which learns the higher-level temporal patterns and generates the final hidden embedding. For continuous emotion regression (CER) tasks, a linear layer is utilized as a regressor to map the hidden embedding to the continuous labels. For discrete emotional state classification (DEC) tasks, these segments in time order share one label of each trial. A sum fusion layer is utilized to generate the final output instead of using a linear layer to get a single output. The sum fusion provided by the present system can improve the generalization ability of the neural networks, as shown by the results which will be described below.

[0049] The construction of the network input will now be described to aid understanding.

[0050] The EEG data of each trial is cut into shorter segments, $\bar{X} \in \mathbb{R}^{C \times L_w}$. The segments are further segmented into sub-segments, denoted by $\bar{X}_{sub} \in \mathbb{R}^{C \times L_{w_{sub}}}$ along the temporal dimension using a sliding window with a certain overlap to learn the long-term temporal dependencies. For each sub-segment, it is still a 2D (two-dimensional) matrix with a spatial dimension and a temporal dimension. The sub-segments are in time order (chronological order) and can be regarded as sequential frames in a video. For each sub-segment, the averaged rPSDs

in each of the following six frequency bands are calculated: (0.3-5Hz), (5-8Hz), (8-12Hz), (12-18Hz), (18-30Hz), and (30-45Hz) 6 frequency bands are calculated. The rPSDs are flattened along the EEG channel dimension, resulting in a feature vector as follows:

$$v = \left[[psd_{ch_0, f_0}, \dots, psd_{ch_0, f_{F-1}}], \dots, [psd_{ch_{C-1}, f_0}, \dots, psd_{ch_{C-1}, f_{F-1}}] \right], \quad (2)$$

where psd is the averaged rPSD, C is the total number of EEG channels, F is the total number of the frequency bands, and $[\cdot]$ is the concatenation. One input to the neural networks would be:

$$v = [v_0, \dots, v_{t-1}]^T, \quad (3)$$

where t is the total number of the rPSD vectors within one segment.

[0051] According to embodiments of the present disclosure, the input to TCN is treated as a 2D matrix, whose dimensions are feature and time. This contrasts with the conventional method of treating the feature vector dimension as the channel dimension of 1D CNN, with the TCN utilizing 1D CNN along the temporal dimension.

[0052] The proposed SAT convolutional layer has two types of convolutional kernels: (i) context kernels that extract the spectral patterns channel by channel, and (ii) spatial fusion kernels that learn spatial patterns across all the channels.

[0053] Given the input, $v = [v_0, \dots, v_{t-1}]^T$, the first type of the CNN kernels in SAT is the 2D causal convolutional kernel whose size, step, and dilation are (f, k) , $(f, 1)$, and $(1, 2)$, where f is the number of frequency bands used to calculate rPSDs and k is the length of the CNN kernel in temporal dimension. (The default dilation step is 1 instead of 0 in the PyTorch library, which means there is no dilation in that dimension if the dilation step is set as 1). Because the step in the feature dimension is the same as the height of the kernel, spectral contextual patterns can be learnt across EEG channels. Hence, the kernels are also referred to as context kernels. The context kernel can learn spectral patterns as well as temporal dynamics at the same time due to its 2D shape.

[0054] The first layer of MASA-TCN has a dilation of 2 in the temporal dimension. Due to the causal convolution, the temporal dimension of the input and output are the same. The output $H_{context} \in \mathbb{R}^{s \times C \times t}$, can be calculated as follows:

$$H_{context} = \text{Conv2D}(x, \text{kernel}_{size} = (f, k), \text{strides} = (f, 1), \text{dilation} = (1, 2)), \quad (4)$$

where s is the number of context kernels, C is the number of EEG channels, and t is the total number of the rPSD vectors within one segment, and where Conv2D is the 2D convolution with the input being x , kernel_{size} , strides , and dilation are the parameters for the CNN operation. Note that the parameters are set as the default value in the PyTorch library unless otherwise specified.

[0055] The structure of SAT is shown schematically in FIG. 3. In this example, the context kernel has a size of $(4, 3)$, with the four-EEG channel sample having four spectral features in each EEG channel. Zero padding is added to make the context kernel a causal kernel along the temporal dimension. The final output (output from different kernels) includes the outputs from a plurality of kernels. Solely for illustrative purposes and to avoid obfuscation, in this example, the number of kernels of each type is four, and only one kernel for each type of the CNN kernels is shown in FIG. 3.

[0056] The output of the context kernels is spatially fused by spatial fusion kernels to learn the spatial patterns of EEG channels. The size, stride, and dilation of the spatial kernels are $(C, 1)$, $(1, 1)$, $(1, 1)$ respectively. It can be treated as an attentive fusion of all the EEG channels with the weights of the 1D CNN kernel being the attention scores. After s spatial fusion kernels, the size of the hidden embedding H_{SF} becomes $(s \times 1 \times t)$. This process can be described as:

$$H_{SF} = \text{Conv2D}(x, \text{kernel}_{size} = (C, 1)),$$

where the default value of strides $(1, 1)$ and dilation $(1, 1)$ are utilized.

[0057] There are two steps in the MAAF: (i) parallel SATs with different temporal kernel lengths, and (ii) attentive fusion of the output from these SATs. Referring again to FIG. 2 for a schematic representation of the architecture of MAAF, in the example, three parallel SATs with

different temporal kernel sizes are utilized to capture temporal dynamics in different time scales, and the temporal lengths of the context kernels are $k = [3, 5, 15]$ respectively. The longer the temporal length, the larger the temporal receptive field. Because the weights of the context kernels are distributed along the time dimension with the help of dilation steps, each weight is like an anchor on the time axis. These parallel SATs are also referred to as multi-anchor SATs. From the causal dependence perspective, different temporal kernel sizes can involve different previous results to decide the next output. The multi-anchor SATs can be described as follows:

$$H_{MA}^i = SAT(x, kernel_{size} = (f, k_i)), i \in [0, 1, 2], \quad (5)$$

where SAT contains the sequential operations of Equation (3) and Equation (4).

[0058] First, the three outputs are concatenated along the kernel dimension (channel dimension of CNNs). A one-by-one convolutional layer serves as an attentive fusion layer as well as a dimension reducer that can reduce the concatenated dimensions back to the previous size. In this example, the kernel dimension is reduced from $3 * s$ to s . Hence, the output of the attentive fusion layer can be described as follows:

$$H_{AF} = Conv2D([H_{MA}^0, H_{MA}^1, H_{MA}^2], kernel_{size} = (1, 1)). \quad (6)$$

TCNs are further stacked to learn the temporal dependencies on top of the learned space-aware temporal patterns from MAAF. TCNs learn from temporal sequences by stacking causal convolution layers with the help of dilated 1D CNN kernels and residual connections. A TCN can be described as follows:

$$H^m = TCN(H^{m-1}) = \Phi(H^{m-1} + \sum_{i=0}^{k-1} f(i) \cdot H_{strd-d \cdot i}^{m-1}), \quad (7)$$

where m is the layer index, $f(\cdot)$ is the filter, k is the kernel size, $strd$ is stride, d is the dilation, and $strd - d \cdot i$ is the direction of the past.

[0059] Hence, the $\Psi_{FL}(\cdot)$ is achieved by equation (1) to equation (7). The model is pre-trained using DEC task for this system.

[0060] With the FL, PCP, and CP, the personalized emotion profile can be generated in two stages (730, 740): (i) a training stage, and (ii) an evaluation stage. An n-fold cross-validation may be utilized, with n-1 folds selected as the training data, and with the remaining one fold used as the test data. This step is repeated n times. There are n folds of refined prediction, $\hat{y}_R$, and the change of emotion intensities of the music clips, Y_{EI} , on the test data.

[0061] FIG. 4A and 4B illustrates the respective models used for a training stage and an evaluation stage respectively. Referring to FIG. 4A, the model or base learner may include a feature learner $\Psi_{FL}(\cdot)$ connected to a hidden embedding, the hidden embedding further connected to a first class predictor $\mathcal{F}_{PCP}(\cdot)$. Referring to FIG. 4B, the model or base learner may include a feature learner $\Psi_{FL}(\cdot)$ connected to a hidden embedding, the hidden embedding further connected to a first class predictor $\mathcal{F}_{PCP}(\cdot)$ and a second class predictor $\mathcal{F}_{CP}(\cdot)$. Further, output from the first class predictor $\mathcal{F}_{PCP}(\cdot)$ and the second class predictor $\mathcal{F}_{CP}(\cdot)$ may be provided to a refine module or refine layer.

[0062] In the training stage, as schematically illustrated in FIG. 4A, the pre-trained model $\mathcal{H}_{base}(\cdot) = \mathcal{F}_{CP}(\Psi_{FL}(\cdot))$ is used as a base learner. $\Psi_{FL}(\cdot)$ is finetuned with a learning rate of $lr_{finetune}$. A $\mathcal{F}_{PCP}(\cdot)$ is added and trained with a learning rate of lr_{train} . Note that $lr_{finetune} \leq lr_{train}$. The self-rating as acquired from the user is utilized as the prediction target.

[0063] In the evaluation stage, as schematically illustrated in FIG. 4B, the finetuned $\Psi_{FL}(\cdot)$ project the input into hidden embedding space. Two branches of the first and second predictors are then used to generate the emotion probabilities of each segment. The trained $\mathcal{F}_{PCP}(\cdot)$ generates $\hat{y}_{PCP}$ from the hidden embeddings. The classifier from base learner, $\mathcal{F}_{CP}(\cdot)$ generates $\hat{y}_{CP}$ using the same hidden embeddings. The refine module combines the information from $\hat{y}_{PCP}$ and $\hat{y}_{CP}$ and generates a final refined prediction $\hat{y}_R$. The refine function can be sum(), mean(), weighted sum(), selection(), and/or any other one or more functions that can combine

the two predictions from the two branches together. For example, a refine function that includes a selection() may be as follows:

$$\hat{y}_R = \begin{cases} \hat{y}_{PCP} & \text{if } |\hat{y}_{PCP} - \hat{y}_{CP}| \leq T \\ \text{None} & \text{if } |\hat{y}_{PCP} - \hat{y}_{CP}| > T \end{cases} \quad (8)$$

[0064] where $|\cdot|$ is the absolute function, T is a adjustable threshold. By using the selection function, only the $(\hat{y}_R^i, Y_{EI}^i)$ pairs with an absolute difference between $\hat{y}_{CP}^i$ and $\hat{y}_{PCP}^i$ smaller than a certain threshold T will be retained.

[0065] The emotion profile of the user may be expressed in terms of pairs of the refined prediction (predicted emotional response), $\hat{y}_R$, and the corresponding change of emotion intensity associated with the music clip, Y_{EI} . In some examples, based on the $(\hat{y}_R^i, Y_{EI}^i)$ pairs obtained, regression modelling may be utilized to model the relationship between $\hat{y}_R$ and Y_{EI} .

[0066] In some embodiments, for the implementation of the system, the model or base learner as obtained in the training stage may be modified and re-used in the evaluation stage with the addition of another class predictor (the second class predictor) and the refine module. In other embodiments, a training model for the training stage and an evaluation model for the evaluation stage may be independently provided. Thereafter, upon completion of the training stage, the parameters of the training model (inclusive of the feature learner and the hidden embedding) may be replicated in or copied to the evaluation model as respective parameters in the feature learner and the hidden embedding of the evaluation model. Therefore, in such embodiments, the evaluation model may include modules corresponding to the training model with the additional of the second class predictor and the refine module.

[0067] FIG. 5 shows a scatter diagram plotted based on the $(\hat{y}_R^i, Y_{EI}^i)$ pairs obtained as described above. It can be seen that the scatter diagram can be one way of representing an emotion profile of a user.

[0068] In some examples, a polynomial regression with a degree of two may be utilized to generate a regression curve. The curve in FIG. 6 shows the results of regression. The related model coefficients are listed in Table 1 below:

Table 1. Model coefficients	
Regression Polynomial:	$y = -0.1215x^2 + 0.0784x + 0.3082$
R-squared:	$r^2 = 0.0008$
Adjusted R-square:	$r_{adj}^2 = -0.0103$
Residual Standard Error:	0.1453 on 89 degrees of freedom

[0069] The coefficients in the regression polynomial can provide more information about the relationship among the user response and the emotion intensity of the stimuli.

[0070] Experiments & Results

[0071] Emotion recognition experiments were conducted using a publicly available dataset DEAP for emotion recognition task. DEAP is a multi-modal human affective states dataset, including EEG, facial expressions, and galvanic skin response (GSR). 40 emotional music videos were carefully selected as stimuli to induce different emotions to the subject (the user). Each video lasted for one minute. Before each trial, there was a three-second baseline recording stage. An online self-assessment tool was provided to the user to collect the feedback of users on arousal, valence, dominance, and liking. For each dimension, a continuous 9-point scale was adopted to measure the level of those dimensions. The valence dimension was utilized for the emotion classification task. 32 subjects participated in the data collection experiments. During the experiment, EEG, GSR, and facial expressions were recorded. A 32-channel Biosemi ActiveTwo system was used with the sampling rate being 512 Hz.

[0072] Firstly, the three-second pre-trial baseline was acquired from each trial. After that, the data were downsampled to 128 Hz. EOG was removed. A band-pass filter was applied to remove the signals which were lower than 4 Hz and higher than 45 Hz. An average reference was conducted on the filtered data to get the final pre-processed data. The label for each

dimension was a continuous 9-point scale. To divide each dimension into high classes and low classes, five were chosen as the threshold to project the nine discrete values into low classes and high classes in each dimension. Each trial was further split into shorter segments of four seconds without overlap to train the neural network.

[0073] The MASA-TCN for binary classification of valence (e.g., classifying an emotion as a positive emotion or as a negative emotion) was compared with the conventional methods of DEC tasks. For a fair comparison, all the methods in the experiment used the same data preprocessing steps, the same segment length (eight seconds) with an overlap of 50%, and the same training strategies as the ones of MASA-TCN. The results are shown in Table 2.

[0074] As seen in Table 2, MASA-TCN achieves higher ACC and F1 Scores than the compared methods. It is observed that the differences in ACC among deep learning methods are not significant, while they all get a higher ACC compared with SVM. MASA-TCN has larger improvements over the compared methods in term of F1 scores. MASA-TCN has a 4.51% higher F1 Score than DeepConvNet, and the improvement over the one of EEGNet is 6.44%. The F1 score of TSception is 7.49% lower than the one of MASA-TCN. MASA-TCN improves the F1 score from 58.07% to 64.58% compared with SVM.

Table 2. Classification results of EEG emotion recognition on DEAP		
Method	Valence	
	ACC	F1
SVM	55.17	58.07
EEGNet	58.45	58.14
DeepConvNet	59.30	60.07
TSception	58.71	57.09
The present system	60.20	64.58

[0075] As described above, the system may be referred to as a personalized music-based emotion profiling system. The system 200 is configured to automatically generate an emotion

profile specific to a certain user, in which the emotion profile corresponds to a personalized response pattern of the user to music-based stimuli. In some embodiments, the system 200 includes developing a stimulus program based on the emotion profile. To aid understanding, an example will be described with reference to FIG. 7.

[0076] FIG. 7 is a schematic diagram to illustrate the system 200 configured to perform a method 700 of developing a music-based stimuli program in accordance with various embodiments of the present disclosure. According to various embodiments, the method 700 includes a stage of collecting inputs 710. For example, the system according to various embodiments of the present disclosure may include a media player. The media player may be configured to play a music clip audible to a user for a time period. In various examples, the music clip may be selected from a library of music clips. Concurrent with the playing of the music clip in the time period, EEG input is collected from the user. For example, EEG electrodes may be attached to the user's head to detect activity in a plurality of functional areas of the user's brain. The system may include a user input device that is operable by the user to input a self-assessed emotion rating responsive to hearing the music clip.

[0077] According to various embodiments, the method 700 includes a stage 720 of generating a personalized emotion profile of a user based on EEG signals, in which the EEG signals are acquired in real-time responsive to the user being exposed to a stimulus, and the stimulus being at least audible to the user. The audible stimulus may be in the form of a music clip or in the form of a media file or recording that includes at least a music clip. For example, the system may include a module configured to generate a personalized emotion profile for the user. For example, the personalized emotion profile of a first user is a prediction of the emotional response of the first user, and not a prediction of the emotional response of a group of users, neither is it a prediction of the emotional response of another user.

[0078] According to various embodiments, the method 700 includes generating a predicted emotional response in two stages: a first stage and a second stage. The first stage includes: providing the EEG input to a feature learner, the feature learner being pre-trained on previous EEG inputs to output a hidden embedding; and using a self-assessed emotion rating as a prediction target in a first class predictor to finetune an output of the hidden embedding. The second stage includes: generating a final predicted emotional response by applying a refine function to both the predicted emotional response and a personalized emotional response of the user to the music clip, the personalized emotional response being generated using a second class predictor on the output of the hidden embedding.

[0079] According to various embodiments, in a subsequent stage (740), the module may be configured to select one or more selected music clips to compile an emotion regulation playlist for use in emotion regulation of the user, in which each of the one or more selected music clips is selected according to the personalized emotion profile of the user. In various examples, the module may be further configured to select one or more selected music clips according to a target emotional intensity to be triggered in the user. For example, in an emotion regulation training session for a user (a patient) who needs to learn how to regulate feelings of fear, the system 200 may be used to select and compile a playlist of selected music clips that are likely to evoke feelings of fear in the user. The selected music clips may not evoke similar emotional responses in other users. However, based on the emotion profile, such music clips can be correctly selected and used in an emotion regulation training session. The user can thus benefit from a more systematic program of emotion regulation training. Additionally, the user can undergo emotion regulation training under more controlled conditions. Music clips can be recorded and/or generated using a music synthesizing software, and stored in a memory for use with the system.

[0080] The foregoing description makes apparent various practical benefits of the present system and method. For example, with an emotion profile generated by the present system, a therapist will be better able to select a stimulus appropriate to the therapy or intervention. For example, with pre-intervention knowledge of how a patient is likely to respond, the therapist will be able to design a more effective therapeutic program. For example, as a part of a diagnostic toolkit, the therapist may be better able to ascertain from a comparison of emotion profiles whether a specific patient's emotion profile shows features characteristic or associated with certain mental disorders.

[0081] The personalized music-based emotion profiling system may be implemented by a processor system 900 as illustrated in the schematic block diagram of FIG. 8. Components of the processing system 900 may be provided within one or more computing device to carry out the functions of the modules or any other modules. One skilled in the art will recognize that the exact configuration or arrangement illustrated in FIG. 8 is provided by way of example only, e.g., each processing system provided may be different and the exact configuration of processing system 900 may vary.

[0082] In embodiments of the present disclosure, the processing system 900 may include a controller 901 and user interface 902. User interface 902 is configured to enable manual interactions between a user and the computing module as required. For this purpose, the processing system 900 includes the input/output components required for the user to enter instructions to provide updates to each of the modules. A person skilled in the art will recognize that components of user interface 902 may vary from embodiment to embodiment but may typically include one or more input devices 935 such as but not limited to a touchscreen, a keyboard, a joystick, a mouse, a microphone, etc. The user interface 902 also includes EEG sensors 933 that can be attached to the user's head to sense the user's brain activity. The user

interface 902 can also include a media player 940, which can be in the form of one or more playback devices, including but not limited to a display, a speaker, earphones, headsets, etc.

[0083] The controller 901 is configured to be in data communication with the user interface 902 via bus 915. The controller 901 includes memory 920 and processor 905 mounted on a circuit board to process instructions and data, e.g., to perform the method of the present disclosure. The controller 901 includes an operating system 906, an input/output (I/O) interface 930 for communicating with user interface 902, and a communications interface, e.g., a network card 950. The network card 950 may, for example, be configured to send data from the controller 901 via a wired or wireless network to other processing devices or to receive data via the wired or wireless network. Wireless networks that may be utilized by the network card 950 include, but are not limited to, Wireless-Fidelity (Wi-Fi), Bluetooth, Near Field Communication (NFC), cellular networks, satellite networks, telecommunication networks, Wide Area Networks (WAN), and etc.

[0084] Memory 920 and operating system 906 are in data communication with central processing unit (CPU) 905 via bus 910. The memory 920 may include both volatile and non-volatile memory. The memory 920 may include more than one of each type of memory, e.g., Random Access Memory (RAM) 923, Read Only Memory (ROM) 925, and a mass storage device 945. The mass storage device 945 may include one or more solid-state drives (SSDs). One skilled in the art will recognize that the memory described above includes non-transitory computer-readable media and shall be taken to include all computer-readable media except for a transitory, propagating signal. Typically, instructions are stored as program code in the memory but can also be hardwired. Memory 920 may include a kernel and/or programming modules such as a software application that may be stored in either volatile or non-volatile memory.

[0085] Herein, the term “processor” is used to refer generically to any device or component that can process computer-readable instructions, including for example, a microprocessor, microcontroller, programmable logic device, or other computational device. That is, processor 905 may be provided by any suitable logic circuitry for receiving inputs, processing them in accordance with instructions stored in memory, and generating outputs (for example to the memory components or media player 940). In the present disclosure, processor 905 may be a single core or multi-core processor with memory addressable space. In one example, processor 905 may be multi-core, comprising—for example—an 8 core CPU. In another example, it could be a cluster of CPU cores operating in parallel to accelerate computations.

[0086] Further, one skilled in the art will recognize that certain functional units in this description have been labelled as modules throughout the specification. The person skilled in the art will also recognize that a module may be implemented as circuits, logic chips or any sort of discrete component. Still further, one skilled in the art will also recognize that a module may be implemented in software which may then be executed by a variety of processor architectures. In embodiments of the disclosure, a module may also comprise computer instructions or executable code that may instruct a computer processor to carry out a sequence of events based on instructions received. In further embodiments, the module may comprise a combination of different types of modules or sub-modules. The choice of the implementation of the modules may be determined by a person skilled in the art and does not limit the scope of the claimed subject matter in any way.

[0087] According to various embodiments of the disclosure, a system is disclosed. The system includes: memory storing instructions; and a processor coupled to the memory and configured to process the stored instructions to implement: a module configured to perform a method of generating a personalized emotion profile for a user. The method includes: generating a predicted emotional response in a first stage, the first stage including: providing

an EEG input acquired from the user to a feature learner, the feature learner being pre-trained on previous EEG inputs to output a hidden embedding; and using a self-assessed emotion rating as a prediction target in a first class predictor to finetune an output of the hidden embedding; and in a second stage, generating a final predicted emotional response by applying a refine function to both the predicted emotional response and a personalized emotional response of the user to a music clip, the personalized emotional response being generated using a second class predictor on the output of the hidden embedding. In some embodiments, the EEG input is acquired concurrent with a time period in which the music clip is played to the user, and wherein the self-assessed emotion rating was received from the user responsive to hearing the music clip, and wherein the music clip is associated with an emotional intensity.

[0088] In some embodiments, the personalized emotion profile comprises a correlation between a range of selected emotion intensities and a respective final predicted emotional response. In some embodiments, the module is configured to select one or more selected music clips for an emotion regulation playlist for use in emotion regulation of the user, and wherein each of the one or more selected music clips is selected according to the personalized emotion profile of the user. In some embodiments, feature learner is trained to learn spatial features and temporal features of the EEG input. In some embodiments, the feature learner is configured to: apply spectral context kernels to segmented EEG data to extract spectral patterns within each of multiple channels corresponding to functional areas of the user's brain, wherein the segmented EEG are segmented according to a sliding time window. In some embodiments, the feature learner is configured to: apply spatial fusion kernels to an output of the spectral context kernels to extract space-aware temporal features from the spectral patterns across all of the multiple channels, the spatial fusion kernels being of different temporal kernel lengths; and attentively fuse the space-aware temporal features.

[0089] In some embodiments, the refine function comprises a selection function, and wherein the personalized predicted emotional response is selected to be the final predicted emotional response only if an absolute difference between the personalized predicted emotional response and the predicted emotional response is smaller than a predetermined threshold. In some embodiments, the refine function comprises any one of the following functions: a sum, a mean, and a weighted sum.

[0090] In some embodiments, the system further includes a media player, the media player being configured to play the music clip audible to a user over the time period. In some embodiments, the system further includes electroencephalogram (EEG) electrodes, the EEG electrodes being attachable to the user's head to detect activity in a plurality of functional areas of the user's brain, the EEG electrodes being configured to acquire the EEG input concurrent with the time period. In some embodiments, the system further includes a user input device, the user input device operable by the user to input the self-assessed emotion rating responsive to hearing the music clip.

[0091] According to various embodiments, a method to generate a personalized emotion profile for a user is disclosed. The method includes: generating a predicted emotional response in a first stage, the first stage including: providing an EEG input acquired from the user to a feature learner, the feature learner being pre-trained on previous EEG inputs to output a hidden embedding; and using a self-assessed emotion rating as a prediction target in a first class predictor to finetune an output of the hidden embedding; and in a second stage, generating a final predicted emotional response by applying a refine function to both the predicted emotional response and a personalized emotional response of the user to a music clip, the personalized emotional response being generated using a second class predictor on the output of the hidden embedding. In some embodiments, the EEG input is acquired concurrent with a time period in which the music clip is played to a user, and wherein the self-assessed emotion rating was

received from the user responsive to hearing the music clip, and wherein the music clip is associated with an emotional intensity.

[0092] In some embodiments, the personalized emotion profile comprises a correlation between a range of selected emotion intensities and a respective final predicted emotional response. In some embodiments, the method further includes selecting one or more selected music clips to compile an emotion regulation playlist for use in emotion regulation of the user, wherein each of the one or more selected music clips is selected according to the personalized emotion profile of the user. In some embodiments, the feature learner is trained to learn spatial features and temporal features of the EEG input. In some embodiments, the feature learner is configured to: apply spectral context kernels to segmented EEG data to extract spectral patterns within each of multiple channels corresponding to functional areas of the user's brain, wherein the segmented EEG are segmented according to a sliding time window. In some embodiments, the feature learner is configured to: apply spatial fusion kernels to an output of the spectral context kernels to extract space-aware temporal features from the spectral patterns across all of the multiple channels, the spatial fusion kernels being of different temporal kernel lengths; and attentively fuse the space-aware temporal features. In some embodiments, the refine function comprises a selection function, and wherein the personalized predicted emotional response is selected to be the final predicted emotional response only if an absolute difference between the personalized predicted emotional response and the predicted emotional response is smaller than a predetermined threshold. In some embodiments, the refine function comprises any one of the following functions: a sum, a mean, and a weighted sum.

[0093] In various embodiments, a system is disclosed. The system includes: a media player, the media player being configured to play a music clip audible to a user for a time period, the music clip being associated with an emotional intensity; electroencephalogram (EEG) electrodes, the EEG electrodes being attachable to the user's head to detect activity in a plurality

of functional areas of the user's brain, the EEG electrodes being configured to acquire EEG input concurrent with the time period; a user input device, the user input device operable by the user to input a self-assessed emotion rating responsive to hearing the music clip; memory storing instructions; and a processor coupled to the memory and configured to process the stored instructions to implement: a module being configured to perform a method of generating a personalized emotion profile for the user. In some embodiments, the method includes: generating a predicted emotional response in a first stage, the first stage including: providing the EEG input to a feature learner, the feature learner being pre-trained on previous EEG inputs to output a hidden embedding; and using the self-assessed emotion rating as a prediction target in a first class predictor to finetune an output of the hidden embedding; and in a second stage, generating a final predicted emotional response by applying a refine function to both the predicted emotional response and a personalized emotional response of the user to the music clip, the personalized emotional response being generated using a second class predictor on the output of the hidden embedding. In some embodiments, the module is further configured to select one or more selected music clips to compile an emotion regulation playlist for use in emotion regulation of the user, and wherein each of the one or more selected music clips is selected according to the personalized emotion profile of the user. In some embodiments, the module is further configured to select the one or more selected music clips according to a target emotional intensity to be triggered in the user.

[0094] All examples described herein, whether of apparatus, methods, materials, or products, are presented for the purpose of illustration and to aid understanding, and are not intended to be limiting or exhaustive. Modifications may be made by one of ordinary skill in the art without departing from the scope of the invention as claimed.

CLAIMS

1. A system, comprising:
memory storing instructions; and
a processor coupled to the memory and configured to process the stored instructions to implement:
a module configured to perform a method of generating a personalized emotion profile for a user, the method including:
generating a predicted emotional response in a first stage, the first stage including:
providing an EEG input acquired from the user to a feature learner, the feature learner being pre-trained on previous EEG inputs to output a hidden embedding; and
using a self-assessed emotion rating as a prediction target in a first class predictor to finetune an output of the hidden embedding; and
in a second stage, generating a final predicted emotional response by applying a refine function to both the predicted emotional response and a personalized emotional response of the user to a music clip, the personalized emotional response being generated using a second class predictor on the output of the hidden embedding,
wherein the EEG input is acquired concurrent with a time period in which the music clip is played to the user, and wherein the self-assessed emotion rating was received from the user responsive to hearing the music clip, and
wherein the music clip is associated with an emotional intensity.

2. The system as recited in claim 1, wherein the personalized emotion profile comprises a correlation between a range of selected emotion intensities and a respective final predicted emotional response.

3. The system as recited in claim 2, wherein the module is configured to select one or more selected music clips for an emotion regulation playlist for use in emotion regulation of the user, and wherein each of the one or more selected music clips is selected according to the personalized emotion profile of the user.

4. The system as recited in any one of claims 1 to 3, wherein the feature learner is trained to learn spatial features and temporal features of the EEG input.

5. The system as recited in any one of claims 1 to 4, wherein the feature learner is configured to:

apply spectral context kernels to segmented EEG data to extract spectral patterns within each of multiple channels corresponding to functional areas of the user's brain, wherein the segmented EEG are segmented according to a sliding time window.

6. The system as recited in claim 5, wherein the feature learner is configured to:
apply spatial fusion kernels to an output of the spectral context kernels to extract space-aware temporal features from the spectral patterns across all of the multiple channels, the spatial fusion kernels being of different temporal kernel lengths; and
attentively fuse the space-aware temporal features.

7. The system as recited in any one of claims 1 to 6, wherein the refine function comprises a selection function, and wherein the personalized predicted emotional response is selected to be the final predicted emotional response only if an absolute difference between the personalized predicted emotional response and the predicted emotional response is smaller than a predetermined threshold.

8. The system as recited in any one of claims 1 to 6, wherein the refine function comprises any one of the following functions: a sum, a mean, and a weighted sum.

9. The system as recited in any ones of claims 1 to 8, further comprising a media player, the media player being configured to play the music clip audible to a user over the time period.

10. The system as recited in any ones of claims 1 to 9, further comprising electroencephalogram (EEG) electrodes, the EEG electrodes being attachable to the user's head to detect activity in a plurality of functional areas of the user's brain, the EEG electrodes being configured to acquire the EEG input concurrent with the time period.

11. The system as recited in any ones of claim 1 to 10, further comprising a user input device, the user input device operable by the user to input the self-assessed emotion rating responsive to hearing the music clip.

12. A method to generate a personalized emotion profile for a user, comprising:
generating a predicted emotional response in a first stage, the first stage including:
providing an EEG input acquired from the user to a feature learner, the feature learner being pre-trained on previous EEG inputs to output a hidden embedding; and

using a self-assessed emotion rating as a prediction target in a first class predictor to finetune an output of the hidden embedding; and
in a second stage, generating a final predicted emotional response by applying a refine function to both the predicted emotional response and a personalized emotional response of the user to a music clip, the personalized emotional response being generated using a second class predictor on the output of the hidden embedding,
wherein the EEG input is acquired concurrent with a time period in which the music clip is played to a user, and wherein the self-assessed emotion rating was received from the user responsive to hearing the music clip, and wherein the music clip is associated with an emotional intensity.

13. The method as recited in claim 12, wherein the personalized emotion profile comprises a correlation between a range of selected emotion intensities and a respective final predicted emotional response.

14. The method as recited in claim 13, further comprising selecting one or more selected music clips to compile an emotion regulation playlist for use in emotion regulation of the user, wherein each of the one or more selected music clips is selected according to the personalized emotion profile of the user.

15. The method as recited in any one of claims 12 to 13, wherein the feature learner is trained to learn spatial features and temporal features of the EEG input.

16. The method as recited in any one of claims 12 to 14, wherein the feature learner is configured to:

apply spectral context kernels to segmented EEG data to extract spectral patterns within each of multiple channels corresponding to functional areas of the user's brain, wherein the segmented EEG are segmented according to a sliding time window.

17. The method as recited in claim 15, wherein the feature learner is configured to:
apply spatial fusion kernels to an output of the spectral context kernels to extract space-aware temporal features from the spectral patterns across all of the multiple channels, the spatial fusion kernels being of different temporal kernel lengths; and
attentively fuse the space-aware temporal features.

18. The method as recited in any one of claims 12 to 16, wherein the refine function comprises a selection function, and wherein the personalized predicted emotional response is selected to be the final predicted emotional response only if an absolute difference between the personalized predicted emotional response and the predicted emotional response is smaller than a predetermined threshold.

19. The method as recited in any one of claims 12 to 16, wherein the refine function comprises any one of the following functions: a sum, a mean, and a weighted sum.

20. A system, comprising:
a media player, the media player being configured to play a music clip audible to a user for a time period, the music clip being associated with an emotional intensity;
electroencephalogram (EEG) electrodes, the EEG electrodes being attachable to the user's head to detect activity in a plurality of functional areas of the user's brain, the EEG electrodes being configured to acquire EEG input concurrent with the time period;

a user input device, the user input device operable by the user to input a self-assessed emotion rating responsive to hearing the music clip;

memory storing instructions; and

a processor coupled to the memory and configured to process the stored instructions to implement: a module being configured to perform a method of generating a personalized emotion profile for the user, the method including:

- generating a predicted emotional response in a first stage, the first stage including:
 - providing the EEG input to a feature learner, the feature learner being pre-trained on previous EEG inputs to output a hidden embedding;
 - and
 - using the self-assessed emotion rating as a prediction target in a first class predictor to finetune an output of the hidden embedding; and
- in a second stage, generating a final predicted emotional response by applying a refine function to both the predicted emotional response and a personalized emotional response of the user to the music clip, the personalized emotional response being generated using a second class predictor on the output of the hidden embedding,

wherein the module is further configured to select one or more selected music clips to compile an emotion regulation playlist for use in emotion regulation of the user, and

wherein each of the one or more selected music clips is selected according to the personalized emotion profile of the user.

21. The system as recited in claim 20, wherein the module is further configured to select the one or more selected music clips according to a target emotional intensity to be triggered in the user.